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). 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.

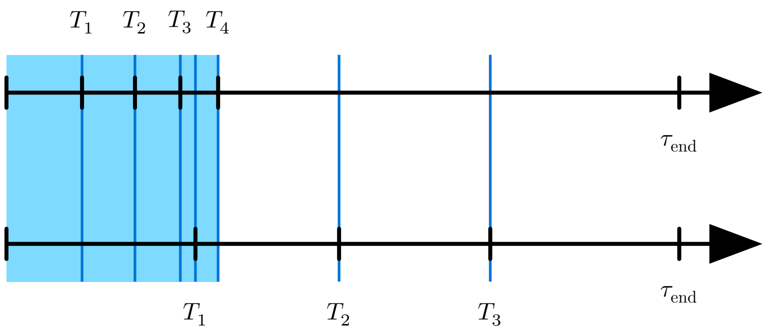

Figure 1: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let $t_1 < ... < t_m$ be all the event times in the sample. Let P^{G^*} denote the interventional probability measure. Then, given $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{t_r}\right]$,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

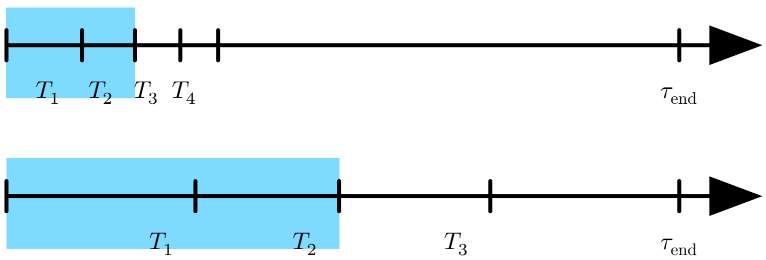

Figure 2: The figure illustrates the sequential regression approach proposed in this article. For each event number k in the sample, we regress back on the history $\mathcal{F}_{T_{(k-1)}}$. Let P^{G^*} denote the interventional probability measure. That is, given $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{T_{(k)}}\right]$,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

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 statistical experiment on which all processes and random variables are defined.

At baseline, we record the values of the treatment $A(0)$ and the time-varying covariates $L(0)$ and let $\mathcal{F}_0 = \sigma(A(0), L(0))$ be the σ-algebra corresponding to the baseline information. It is not a loss of generality to assume that we have two treatment options over time so that $A(t) \in \{0, 1\}$ (e.g., placebo and active treatment), where $A(t)$ denotes the treatment at time $t \geq 0$. The time-varying confounders $L(t)$ at time $t > 0$ are assumed to take values in a finite subset $\mathcal{L} \subset \mathbb{R}^m$, so that $L(t) \in \mathcal{L}$ for all $t \geq 0$. We assume that the stochastic processes $(L(t))_{t\geq 0}$ and $(A(t))_{t\geq 0}$ are càdlàg (right-continuous with left limits), jump processes. Furthermore, we require that the times at which the treatment and covariate values may change are dictated entirely by the counting processes $(N^a(t))_{t\geq 0}$ and $(N^\ell(t))_{t\geq 0}$, respectively in the sense that $\Delta A(t) \neq 0$ only if $\Delta N^a(t) \neq 0$ and $\Delta L(t) \neq 0$ only if $\Delta N^\ell(t) \neq 0$. We *emphasize* the importance of this assumption: Random changes of covariate values L and treatment A may only happen at a possibly random discrete set of time points. For technical reasons and ease of notation, we shall assume that the number of jumps $K(t)$ at time t for the processes L and A satisfies $K(\tau_{\text{end}}) \leq K - 1$ P-a.s. for some finite $K \geq 1$.

We also have counting processes representing the event of interest $(N^y(t))_{t\geq 0}$ and the competing event $(N^d(t))_{t\geq 0}$. Let N^c be the censoring process. Initially, we shall allow only administrative censoring, i.e., $N^c(t) = \mathbb{1}\{t > \tau_{\text{end}}\}$ for all $t \geq 0$. For all counting processes involved, we assume for simplicity that the jump times differ with probability 1 (i.e., if $x \neq y$, we have $\Delta N^x \Delta N^y \equiv 0$.). This is not a limiting assumption for real data as events are never truly simultaneous. Thus, we have observations from a jump process $\alpha(t) = (N^a(t), A(t), N^\ell(t), L(t), N^y(t), N^d(t))$, and the natural filtration $(\mathcal{F}_t)_{t\geq 0}$ is given by $\mathcal{F}_t = \sigma(\alpha(s) \mid s \leq t) \vee \mathcal{F}_0$. Let $T_{(k)}$ be the k'th ordered jump time of α, that is $T_0 = 0$ and $T_{(k)} = \inf\{t > T_{(k-1)} \mid \alpha(t) \neq \alpha(T_{(k-1)})\} \in [0, \infty]$ be the time of the k'th event and let $\Delta_{(k)} \in \{c, y, d, a, \ell\}$ be the status of the k'th event, i.e., $\Delta_{(k)} = x$ if $\Delta N^x(T_{(k)}) = 1$. We let $T_{(k+1)} = \infty$ if $T_{(k)} = \infty$ or $\Delta_{(k-1)} \in \{y, d, c\}$. As is common in the point process literature, we define $\Delta_{(k)} = \emptyset$ if $T_{(k)} = \infty$ or $\Delta_{(k-1)} \in \{y, d, c\}$ for the empty mark.

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

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

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

We observe $O = (T_{(K)}, \Delta_{(K)}, A(T_{(K-1)}), L(T_{(K-1)}), T_{(K-1)}, \Delta_{(K-1)}, ..., A(0), L(0)) \sim P \in \mathcal{M}$ where $\mathcal{M}$ is the statistical model, i.e., a set of probability measures and obtain a sample $O = (O_1, ..., O_n)$ of size n. For a single individual, we might observe $A(0) = 0$ and $L(0) = 2$, $A(T_{(1)}) = 1$, $L(T_{(1)}) = 2$, $T_{(1)} = 0.5$, and $\Delta_{(1)} = a$, $A(T_{(2)}) = 1$, $L(T_{(2)}) = 2$, $T_{(2)} = 1.5$, and $\Delta_{(2)} = y$, and $T_{(3)} = \infty$, $\Delta_{(3)} = \emptyset$, so $K(t) = 2$ for that individual. Another person might have $\Delta_{(1)} = d$ and so $K(t) = 0$ for that individual. For the confused reader, we refer to Table 1, which gives the long format of a hypothetical longitudinal dataset with time-varying covariates and treatment registered at irregular time points, and its conversion to wide format in Table 2, representing the data set in the form of O.

id	time	event	L	A
1	0	baseline	2	1
1	0.5	visitation time; stay on treatment	2	1
1	8	primary event	Ø	Ø
2	0	baseline	1	0
2	10	primary event	Ø	Ø
3	0	baseline	3	1
3	2	side effect (L)	4	1
3	2.1	visitation time; discontinue treatment	4	0
3	5	primary event	Ø	Ø

Table 1: An example of a longitudinal dataset from electronic health records or a clinical trial with $\tau_{\text{end}} = 15$ with $K = 2$ for $n = 3$ (3 observations). Here, the time-varying covariates only have dimension 1. Events are registered at irregular/subject-specific time points and are presented in a long format. Technically, though, events at baseline are not to be considered events, but we include them here for completeness.

id	$L(0)$	$A(0)$	$L(T_{(1)})$	$A(T_{(1)})$	$T_{(1)}$	$\Delta_{(1)}$	$L(T_{(2)})$	$A(T_{(2)})$	$T_{(2)}$	$\Delta_{(2)}$	$T_{(3)}$	$\Delta_{(3)}$
1	2	1	2	1	0.5	a	Ø	Ø	8	y	∞	Ø
2	1	0	Ø	Ø	10	y	Ø	Ø	∞	Ø	∞	Ø
3	3	1	4	1	2	ℓ	4	0	2.1	a	5	y

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

We will also work within the so-called canonical setting for technical reasons (Last & Brandt (1995), Section 2.2). Intuitively, this means that we assume that P defines only the distribution for the sequence of random variables given by O and that we work with the natural filtration $(\mathcal{F}_t)_{t\geq 0}$ generated by the random measure N^α. This is needed to ensure the existence of compensators can be explicitly written via by the regular conditional distributions of the jump times and marks, but also to ensure that $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k-1)}\right), L\left(T_{(k-1)}\right)\right) \vee \mathcal{F}_{T_{(k-1)}}$ and $\mathcal{F}_0 = \sigma(A(0), L(0))$, where $\mathcal{F}_{T_{(k)}}$ stopping time σ-algebra $\mathcal{F}_{T_{(k)}}$ – representing the information up to and including the k'th event – associated with stopping time $T_{(k)}$. We will interpret $\mathcal{F}_{T_{(k)}}$ as a random variable instead of a 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 $\Delta_{(k)} \in \{a, \ell\}$.

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 given $\Delta_{(k)} = \ell, T_{(k)} = t$, 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 (see e.g., Appendix A5.3 of Last & Brandt (1995)). Note that in many places, we will not distinguish between $\Lambda_k^x\left((0,t], \mathcal{F}_{T_{(k-1)}}\right)$ and $\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right)$. At baseline, we let $\pi_0(L(0))$ be the probability of being treated given $L(0)$ and $\mu_0(\cdot)$ be the probability measure for the covariate value.

3 Estimand of interest and iterative representation

We are interested in the causal effect of a treatment regime g on the cumulative incidence function of the event of interest y at time τ. We consider regimes which naturally act upon the treatment decisions at each visitation time but not the times at which the individuals visit the doctor. The treatment regime g specifies for each event $k = 1, \dots, K-1$ with $\Delta_{(k)} = a$ (visitation time) the probability that a patient will remain treated until the next visitation time via $\pi_k^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)$ and at $k = 0$ the initial treatment probability $\pi_0^*(L(0))$.

We first define a *version* of the likelihood ratio process,

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

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

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

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

$$\Psi_\tau^g(P) = \mathbb{E}_{P^{G^*}}[N^y(\tau)] = \mathbb{E}_P[N^y(\tau)W^g(\tau)], \tag{2}$$

where $N^y(t) = \sum_{k=1}^K \mathbb{1}\{T_{(k)} \leq t, \Delta_{(k)} = y\}$. In our application, π_k^* may be chosen arbitrarily, so that, in principle, *stochastic*, *dynamic*, and *static* treatment regimes can be considered. However, for simplicity of presentation, we use the static observation plan $\pi_0^*(L(0)) = 1$ and $\pi_k^*\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) = 1$ for all $k = 1, ..., K-1$, and the methods we present can easily be extended to more complex treatment regimes and contrasts. In this paper, we will assume that Equation 2 causally identifies the estimand of interest.

We now present a simple iterated representation of the data target parameter $\Psi_\tau^g(P)$ 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)).

Our idea builds upon the works of Bang & Robins (2005) (in discrete time) and Rytgaard et al. (2022) (in continuous time), who suggest the use of so-called *iterated regressions* of the target parameter. We discuss more thoroughly the implications for inference of this representation, the algorithm for estimation and examples in Section 4 where we also deal with right-censoring.

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

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

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

$$\Psi_\tau^g(P) = \mathbb{E}_P\left[\bar{Q}_{0,\tau}^g(1, L(0))\right]. \tag{4}$$

Furthermore,

$$\bar{Q}_{k-1,\tau}^g\left(\mathcal{F}_{T_{(k-1)}}\right) = p_{k-1,a}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k-1,\ell}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k-1,y}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) \tag{5}$$

where,

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

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

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

for $k = 0, ..., K$ satisfies the desired property of Equation 3. 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\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\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\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)}}\Bigg) \middle| \mathcal{F}_{T_{(k)}}\right] \\
&= \mathbb{E}_P\Bigg(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} \qquad (6) \\
&\qquad\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\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] \middle| \mathcal{F}_{T_{(k)}}\Bigg) \\
&= \mathbb{E}_P\Big(\mathbb{1}\left\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\right\} \\
&\qquad\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\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\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\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 3 follows. A similar calculation shows that $\Psi^g_\tau(P) = \mathbb{E}_P[\bar{Q}^g_{0,\tau}(1, L(0))]$ and so Equation 4 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))

$$P\left(\left(T_{(k)},\Delta_{(k)}\right)\in d(t,m)\mid \mathcal{F}_{T_{(k-1)}}\right)=\sum_{x=a,\ell,d,y}\delta_x(\mathrm{d}m)\prod_{s\in\left(T_{(k-1)},t\right)}\left(1-\sum_{x=\ell,a,d,y}\Lambda_k^x\left(\mathrm{d}s,\mathcal{F}_{T_{(k-1)}}\right)\right)\Lambda_k^z\left(\mathrm{d}t,\mathcal{F}_{T_{(k-1)}}\right),\tag{7}$$

whenever $T_{(k-1)}<\infty$ and $\Delta_{(k-1)}\in\{a,\ell\}$, so we get Equation 5 by plugging in Equation 7 to the second last equality of Equation 6. □

Two approaches are suggested by Theorem 1. The representation in Theorem 1 has a natural interpretation: $\bar{Q}^g_{k,\tau}$ is the counterfactual probability of the primary event occuring at or before time τ given the history up to and including the k'th event (among the people who are at risk of the event before time τ after k events). Equation 3 then suggests that we can estimate $\bar{Q}^g_{k-1,\tau}$ via $\bar{Q}^g_{k,\tau}$ by considering what has happened as the k'th event: For each individual in the sample, we calculate the integrand in Equation 3 depending on their value of $T_{(k)}$ and $\Delta_{(k)}$, and apply the treatment regime as specified by π^*_k if the individuals event is a treatment event. Then, we regress these values directly on $\left(A\left(T_{(k-1)}\right),L\left(T_{(k-1)}\right),T_{(k-1)},\Delta_{(k-1)},...,A(0),L(0)\right)$ to obtain an estimator of $\bar{Q}^g_{k-1,\tau}$. On the other hand, we do not recommend using Equation 5 directly, which involves iterative integration, as this method becomes computationally infeasible even for small values of K.

A benefit of this representation compared to full inverse probability weighting, where we also weight with treatment propensities, is that the estimator may be more stable in the case of small estimated treatment propensities.

4 Censoring

In this section, we allow for right-censoring. That is, we introduce a right-censoring time $C>0$ at which we stop observing the multivariate jump process α. We will introduce the notation necessary to discuss the algorithm for the the ICE-IPCW estimator in Section 5 and later discuss the assumptions necessary for consistency of the ICE-IPCW estimator in Section 6. In the remainder of the paper, we will assume that $C\neq T_{(k)}$ for all k with probability 1. As before, we let $\left(T_{(k)},\Delta_{(k)},A\left(T_{(k)}\right),L\left(T_{(k)}\right)\right)$ be the event times and marks for the N^α process.

Let $\left(\bar{T}_{(k)},\bar{\Delta}_{(k)},A(\bar{T}_k),L(\bar{T}_k)\right)$ for $k=1,...,K$ be the observed data given by

$$\begin{aligned}
\bar{T}_{(k)}&=C\wedge T_{(k)}\\
\bar{\Delta}_{(k)}&=\begin{cases}\Delta_{(k)} \text{ if } C>T_{(k)}\\ \text{c if } C\leq T_{(k)} \text{ and } \bar{\Delta}_{(k-1)}\neq c\\ \text{Ø otherwise}\end{cases}\\
A(\bar{T}_k)&=\begin{cases}A\left(T_{(k)}\right)\text{if } C>T_{(k)}\\ \text{Ø otherwise}\end{cases}\\
L(\bar{T}_k)&=\begin{cases}L\left(T_{(k)}\right) \text{ if } C>T_{(k)}\\ \text{Ø otherwise}\end{cases}
\end{aligned}\tag{8}$$

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

$$\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}=\sigma\left(\bar{T}_{(k)},\bar{\Delta}_{(k)},A(\bar{T}_k),L(\bar{T}_k),...,\bar{T}_{(1)},\bar{\Delta}_{(1)},A(\bar{T}_1),L(\bar{T}_1),A(0),L(0)\right),\tag{9}$$

defining the observed history up to and including the k'th event. Thus $O=\left(\bar{T}_{(1)},\bar{\Delta}_{(1)},A(\bar{T}_1),L(\bar{T}_1),...,\bar{T}_{(K)},\bar{\Delta}_{(K)},A(\bar{T}_K),L(\bar{T}_K)\right)$ is the observed data and a sample consists of $O=(O_1,...,O_n)$ for n independent and identically distributed observations with $O_i\sim P$. We will formally show Equation 9 later.

Define $\tilde{\Lambda}^c_k\left(\mathrm{d}t,\mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)$ as the cause-specific cumulative hazard measure of the k'th event and that the event was a censoring event at time t given the observed history $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}$ and define the corresponding censoring survival function $\tilde{S}^c\left(t\mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)=\prod_{s\in\left(T_{(k-1)},t\right]}\left(1-\tilde{\Lambda}^c_k\left(\mathrm{d}s,\mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\right)$. This determines the probability of being observed at time t given the observed history up to $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}$.

5 Algorithm for ICE-IPCW Estimator

In this section, we present an algorithm for the ICE-IPCW estimator and consider its use in a simple data example.

It requires as inputs the dataset $\mathcal{D}_n$, a time point τ of interest, and a cause-specific cumulative hazard model $\mathcal{L}_h$ for the censoring process. This model takes as input the event times, the cause of interest, and covariates from some covariate space $\mathbb{X}$, and outputs an estimate of the cumulative cause-specific hazard function $\hat{\Lambda}^c : (0, \tau) \times \mathbb{X} \to \mathbb{R}_+$ for the censoring process. It is technically allowed for this procedure to only give estimates of $\frac{1}{P\left(\bar{T}_{(k)} \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} P\left(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$, which is always estimable from observed data, and not $\frac{1}{P\left(C \geq t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} P\left(C \in \mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$. For all practical purposes, we will assume that these are the same.

The algorithm also takes a model $\mathcal{L}_o$ for the iterative regressions, which returns a prediction function $\hat{\nu} : \mathbb{X} \to \mathbb{R}_+$ for the pseudo-outcome. Ideally, both models should be chosen flexibly, since even with full knowledge of the data-generating mechanism, the true functional form of the regression model cannot typically be derived in closed form. Also, the model should be chosen such that the predictions are $[0, 1]$-valued.

For use of the algorithm in practice, we shall choose K such that $K = \max_{i \in \{1,...,n\}} K_i(\tau)$, where $K_i(\tau)$ is the number of non-terminal events for individual i in the sample. However, we note that this may not always be possible as there might be few people with many events. Therefore, one may have to prespecify K instead and define a composite outcome. Specifically, we let $k^* = \inf\left\{k \in \{K+1, ..., \max_i K_i(\tau)\} \mid \bar{\Delta}_{(k)} \in \{y, d, c\}\right\}$, and $\bar{T}^*_{(K+1)} = \bar{T}_{(k^*)}$ and $\bar{D}^*_{(K+1)} = \bar{\Delta}_{(k^*)}$ and use the data set where $\bar{T}_{(k)}, \bar{\Delta}_{(k)}, A(\bar{T}_k), L(\bar{T}_k)$ for $k > K+1, ..., k^*$ are removed from the data and instead $\bar{T}^*_{(K+1)}$ and $\bar{D}^*_{(K+1)}$ are used as the event time and status for the $(K+1)$'th event. Strictly speaking, we are not estimating the interventional cumulative incidence function at time τ as we set out to do originally because the intervention has changed. In this situation, the doctor will only have to prescribe treatment to patients who visit the doctor as part of their k^* first events. However, this estimand is likely to be close to the original estimand of interest. The algorithm can then be stated as follows:

- For each event point $k = K+1, K, ..., 1$ (starting with $k = K$):
  1. Regress $\bar{S}_{(k)} = \bar{T}_{(k)} - \bar{T}_{(k-1)}$, known as the k'th *interarrival* time, with the censoring as the cause of interest on $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}$ (among the people who are still at risk after $k-1$ events, that is $R_k = \left\{i \in \{1, ..., n\} \mid \bar{\Delta}_{k-1,i} \in \{a, \ell\}\right\}$ if $k > 1$ and otherwise $R_1 = \{1, ..., n\}$) using $\mathcal{L}_h$ to obtain an estimate of the cause-specific cumulative hazard function $\hat{\Lambda}^c_k$. For $k = 1$, note that we take $\bar{T}_0 = 0$.
  2. Obtain estimates $\hat{S}^c\left(\bar{T}_{(k)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) = \prod_{s \in (0, \bar{T}_{k+1} - \bar{T}_k)} \left(1 - \hat{\Lambda}^c_k\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)$ from step 1.
  3. Calculate the subject-specific pseudo-outcome:

$$\widehat{m}_k = \frac{\mathbb{1}\left\{\bar{T}_{(k)} \leq \tau, \bar{\Delta}_{(k)} = y\right\}}{\hat{S}^c\left(\bar{T}_{(k)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}$$

  Then,

  - If $k \leq K$:

  Let $\mathcal{F}^g_{\bar{T}_{(k)}}$ denote the history with $A(\bar{T}_k) = 1$ if $\bar{\Delta}_{(k)} = a$. Otherwise, $\mathcal{F}^g_{\bar{T}_{(k)}} = \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}$. Then calculate,

$$\hat{Z}^a_k = \frac{\mathbb{1}\left\{\bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} \in \{a, \ell\}\right\} \hat{\nu}_k\left(\mathcal{F}^g_{\bar{T}_{(k)}}\right)}{\hat{S}^c\left(\bar{T}_{(k)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} + \widehat{m}_k.$$

  - If $k = K+1$, put

$$\hat{Z}^a_k = \widehat{m}_k.$$

4. Regress $\hat{Z}_k^a$ on $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}$ with model $\mathcal{L}_o$ on the data with $\bar{T}_{(k-1)} < \tau$ and $\bar{\Delta}_{(k-1)} \in \{a, \ell\}$ to obtain a prediction function $\hat{\nu}_{k-1} : \mathcal{H}_{k-1} \to \mathbb{R}^+$. Here we denote by $\mathcal{H}_{k-1}$ the set of possible histories of the process up to and including the $k-1$'th event.

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

Note that the third step, we can alternatively in $\mathcal{F}^g_{\bar{T}_{(k)}}$ replace all prior treatment values with 1, that is $A(0) = ... = A(\bar{T}_{k-1}) = 1$. This is certainly closer to the iterative conditional expectation estimator as proposed by Bang & Robins (2005), but is mathematically equivalent as we, in the next iterations, condition on the treatment being set to 1. This follows from standard properties of the conditional expectation (see e.g., Theorem A3.13 of Last & Brandt (1995)).

In the first step, the modeler wish to alter the history from an intuitive point of view, so that, in the history $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}$, we use the variables $\bar{T}_{k-1} - \bar{T}_j$ for $j \leq k-1$ instead of the variables $\bar{T}_j$, altering the event times in the history to "time since last event" instead of the "event times" (note that we should then remove $\bar{T}_{k-1}$ from the history as it is identically zero). This makes our regression procedure in step 1 intuitively look like a simple regression procedure at time zero.

We also need to discuss what models should be used for $\bar{Q}^g_{k,\tau}$. Note that

$$\mathbb{1}\left\{T_{(k)} < \tau \wedge \Delta_{(k)} \in \{a, \ell\}\right\} \bar{Q}^g_{k,\tau} = \mathbb{E}_{P^{G^*}}\left[N^y(\tau) \mid \mathcal{F}_{T_{(k)}}\right] \mathbb{1}\left\{T_{(k)} < \tau \wedge \Delta_{(k)} \in \{a, \ell\}\right\}$$

We see thus see that we should use a model for $\bar{Q}^g_{k,\tau}$ that is able to capture the counterfactual probability of the primary event occuring at or before time τ given the history up to and including the k'th event (among the people who are at risk of the event before time τ after k events).

5.1 Example usage of the Algorithm

To help illustrate the algorithm, we present a simple example in Table 3 in the case where $\tau = 5$. Since $K = 2$ in Table 3, we start at $k = 3$.

Iteration $k = 3$

1. First, we fit a cause-specific hazard model for people at risk of the k'th event, R_3. We find that $R_3 = \{3, 4, 7\}$. Here, the interarrival times are given by $\bar{S}_{(3),3} = 5 - 2.1 = 2.9, \bar{S}_{(3),4} = 8 - 6.7 = 1.3, \bar{S}_{(3),7} = 4.9 - 4.7 = 0.2$, and status indicators $\bar{\Delta}^c_{3,3} = 0$, $\bar{\Delta}^c_{3,4} = 0$, $\bar{\Delta}^c_{3,7} = 1$ respectively. Let $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{2,i}} = \left(L_i(0), A_i(0), L_i(\bar{T}_{1,i}), A_i(\bar{T}_{1,i}), \bar{T}_{1,i}, \bar{\Delta}_{1,i}, L_i(\bar{T}_{2,i}), A_i(\bar{T}_{2,i}), \bar{T}_{2,i}, \bar{\Delta}_{2,i}\right)$. To obtain $\hat{\Lambda}^c$, we regress the event times $\bar{S}_3$ on $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_2}$ with status indicator $\bar{\Delta}^c_3$.

2/3. From $\hat{\Lambda}^c$ obtain estimates $\widehat{S}^c\left(\bar{T}_{3,i} - \mid \mathcal{F}^{\beta}_{\bar{T}_{2,i}}\right)$ for $i \in R_3 \cap \{i : \bar{T}_{2,i} < \tau\} = \{3, 7\}$. We assume that these are given by $\widehat{S}^c\left(\bar{T}_{3,3} - \mid \mathcal{F}^{\beta}_{\bar{T}_{2,3}}\right) = 0.8$, $\widehat{S}^c\left(\bar{T}_{3,7} - \mid \mathcal{F}^{\beta}_{\bar{T}_{2,7}}\right) = 0.9$. Then, we calculate $\hat{Z}^a_{3,3} = \frac{\mathbb{1}\{\bar{T}_{3,3} \leq \tau, \bar{\Delta}_{3,3} = y\}}{\widehat{S}^c\left(\bar{T}_{3,3} - \mid \mathcal{F}^{\beta}_{\bar{T}_{2,3}}\right)} = \frac{1}{0.8} = 1.25$, $\hat{Z}^a_{3,7} = \frac{\mathbb{1}\{\bar{T}_{3,7} \leq \tau, \bar{\Delta}_{3,7} = y\}}{\widehat{S}^c\left(\bar{T}_{3,7} - \mid \mathcal{F}^{\beta}_{\bar{T}_{2,7}}\right)} = \frac{0}{0.9} = 0$.

4. Regress the predicted values $\hat{Z}^a_3$ on $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(2)}}$ to obtain a prediction function $\hat{\nu}_2$.

Iteration $k = 2$

1. As in the case $k = 3$, we fit a cause-specific hazard model for people at risk of the k'th event, R_2. We find that $R_2 = \{1, 3, 4, 6, 7\}$. Here, the interarrival times are given by $\bar{S}_{(2),1} = 8 - 0.5 = 7.5$, $\bar{S}_{(2),3} = 2.1 - 2 = 0.1$, $\bar{S}_{(2),4} = 6.7 - 6 = 0.7$, $\bar{S}_{(2),6} = 5 - 1 = 4$, $\bar{S}_{(2),7} = 4.7 - 4 = 0.7$, and $\bar{\Delta}^c_{2,i} = 0$ for $i = 1, 3, 4, 6, 7$. Regress the event times $\bar{S}_2$ on $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_1}$ with status indicators $\bar{\Delta}^c_2$ to obtain $\hat{\Lambda}^c$.

2/3. From $\hat{\Lambda}^c$ obtain estimates $\widehat{S}^c\left(\bar{T}_{2,i}- \mid \mathcal{F}^{\beta}_{\bar{T}_{1,i}}\right)$ for $i \in R_2 \cap \{i : \bar{T}_{1,i} < \tau\} = \{1,3,6,7\}$. We assume that these are given by $\widehat{S}^c\left(\bar{T}_{2,1}- \mid \mathcal{F}^{\beta}_{\bar{T}_{1,1}}\right) = 0.9$, $\widehat{S}^c\left(\bar{T}_{2,3}- \mid \mathcal{F}^{\beta}_{\bar{T}_{1,3}}\right) = 0.8$, $\widehat{S}^c\left(\bar{T}_{2,6}- \mid \mathcal{F}^{\beta}_{\bar{T}_{1,6}}\right) = 0.7$, $\widehat{S}^c\left(\bar{T}_{2,7}- \mid \mathcal{F}^{\beta}_{\bar{T}_{1,7}}\right) = 0.6$. Then, we calculate $\hat{Z}^a_{2,1} = \frac{0}{0.9} = 0, \hat{Z}_{2,6} = \frac{0}{0.7} = 0$. We now apply the prediction functions from step $k = 3$.

- For $i = 3$, we produce the altered history, where $\mathcal{F}^{g}_{\bar{T}_{2,i}} = (3,1,4,1,2,\ell,4,1,2.1,a)$ and apply $\hat{\nu}_2\left(\mathcal{F}^{g}_{\bar{T}_{2,3}}\right) = 0.3$, so we get $\hat{Z}^a_{2,i} = \frac{\mathbb{1}\{\bar{T}_{2,i}<\tau,\bar{\Delta}_{2,i}\in\{a,\ell\}\}\hat{\nu}_2\left(\mathcal{F}^{g}_{\bar{T}_{2,i}}\right)}{\widehat{S}^c\left(\bar{T}_{2,i}- \mid \mathcal{F}^{\beta}_{\bar{T}_{1,i}}\right)} = \frac{0.3}{0.8} = 0.375$.
- For $i = 7$, we keep the history and get $\hat{\nu}_2\left(\mathcal{F}^{g}_{\bar{T}_{2,7}}\right) = 0.25$, so we get $\hat{Z}^a_{2,3} = \frac{\mathbb{1}\{\bar{T}_{2,3}<\tau,\bar{\Delta}_{2,3}=a\}}{\widehat{S}^c\left(\bar{T}_{2,3}- \mid \mathcal{F}^{\beta}_{\bar{T}_{1,3}}\right)}\hat{\nu}_2\left(\mathcal{F}^{g}_{\bar{T}_{1,3}}\right) = \frac{0.25}{0.8} = 0.3125$.

4. Regress the predicted values $\hat{Z}^a_2$ on $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(1)}}$ to obtain a prediction function $\hat{\nu}_1$.

Iteration $k = 1$

Same procedure as $k = 2$. Note that the interarrival times are simply given by the event times here.

Iteration $k = 0$

We get the estimate $\hat{\Psi}_n = \frac{1}{7}\sum_{i=1}^{7} \hat{\nu}_0(1, L_i(0))$ for $n = 7$, where we obtained $\hat{\nu}_0$ from $k = 1$.

id	$L(0)$	$A(0)$	$L(\bar{T}_1)$	$A(\bar{T}_1)$	$\bar{T}_{(1)}$	$\bar{\Delta}_{(1)}$	$L(\bar{T}_2)$	$A(\bar{T}_2)$	$\bar{T}_{(2)}$	$\bar{\Delta}_{(2)}$	$\bar{T}_{(3)}$	$\bar{\Delta}_{(3)}$
1	2	1	2	1	0.5	a	Ø	Ø	8	y	∞	Ø
2	1	0	Ø	Ø	10	y	Ø	Ø	∞	Ø	∞	Ø
3	3	1	4	1	2	ℓ	4	0	2.1	a	5	y
4	3	1	4	1	6	ℓ	4	0	6.7	a	8	y
5	1	1	Ø	Ø	3	d	Ø	Ø	∞	Ø	∞	Ø
6	1	1	0	3	1	ℓ	Ø	Ø	5	d	Ø	Ø
7	3	1	4	1	4	ℓ	5	1	4.7	ℓ	4.9	c

Table 3: Example data for illustration of the ICE-IPCW algorithm.

6 Consistency of the ICE-IPCW Estimator

Now let $N^c(t) = \mathbb{1}\{C \le t\}$ the counting process for the censoring process and let T^e further denote the (uncensored) terminal event time given by

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

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

Then, we observe the trajectories of the process given by $t \mapsto N^\beta(t \wedge C \wedge T^e)$ and the observed filtration is given by

$\mathcal{F}^{\tilde{\beta}}_t = \sigma(\beta(s \wedge C \wedge T^e) \mid s \leq t)$. The observed data is then given by Equation 8. Importantly, we have in fact*.

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

Abusing notation a bit, we see that for observed histories, we have $\mathcal{F}_{T_{(k)}} = \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}$ if $\bar{\Delta}_{(k)} \neq c$. Note that here we also have not shown that $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right), ..., T_{(1)}, \Delta_{(1)}, A\left(T_{(1)}\right), L\left(T_{(1)}\right), A(0), L(0)\right)$. However, our results up to this point only rely on conditioning on the variables representing the history up to and including the k'th event.

In this section, we present the conditions under which the ICE-IPCW estimator is consistent for the target parameter. What we require for the identification via the iterated regressions, is that

$$\begin{aligned}
&P\left(T_{(k)} \in [t, t + \mathrm{d}t), \Delta_{(k)} = x, A\left(T_{(k)}\right) = m, L\left(T_{(k)}\right) = l \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} \geq t\right) \\
&= P\left(\bar{T}_{(k)} \in [t, t + \mathrm{d}t), \bar{\Delta}_{(k)} = x, A(\bar{T}_k) = m, L(\bar{T}_k) = l \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}, \bar{T}_{(k)} \geq t\right), \qquad x \neq c.
\end{aligned} \tag{10}$$

for uncensored histories, i.e., when $\bar{\Delta}_{(k-1)} \neq c$ as well as regularity condition 1 and 2 in Theorem 2.

We posit specific conditions in Theorem 2 similar to those that may be found the literature based on independent censoring (Andersen et al. (1993); Definition III.2.1) or local independence conditions (Røysland et al. (2024); Definition 4). Alternatively, we may assume coarsening at random which will imply Equation 10 (e.g., Gill et al. (1997)).

*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). Moreover, from this we have $\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = \mathcal{F}^{\beta}_{T_{(k)} \wedge C \wedge T^e}$ and we may apply Theorem 2.1.14 to the right-hand side of $\mathcal{F}^{\beta}_{T_{(k)} \wedge C \wedge T^e}$ to get the desired statement.

Theorem 2: 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

$$\begin{aligned}
&\mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\} P\left(\left(\bar{T}_n, \bar{\Delta}_n, A(\bar{T}_n), L(\bar{T}_n)\right) \in d(t,m,a,l) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) \\
&= \mathbb{1}\left\{\bar{T}_{n-1} < t, \bar{\Delta}_{(n-1)} \neq c\right\}\left(\tilde{S}\left(t- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) \sum_{x=a,\ell,d,y} \delta_x(\mathrm{d}m)\psi_{n,x}(t, d(a,l))\Lambda_n^x\left(\mathrm{d}t, \mathcal{F}_{T_{(n-1)}}\right)\right. \\
&\quad \left. + \delta_{\left(c, A\left(T_{(n-1)}\right), L\left(T_{(n-1)}\right)\right)}(\mathrm{d}(m,a,l))\tilde{\Lambda}_n^c\left(\mathrm{d}t, \mathcal{F}^{\beta}_{\bar{T}_{(n-1)}}\right)\right)
\end{aligned} \tag{11}$$

where

$$\begin{aligned}
\psi_{n,x}\left(t, \mathcal{F}_{T_{(n-1)}}, \mathrm{d}(m,a,l)\right) &= \mathbb{1}\{x=a\}\left(\delta_1(\mathrm{d}a)\pi_n\left(t, \mathcal{F}_{T_{(n-1)}}\right) + \delta_0(\mathrm{d}a)\left(1 - \pi_n\left(t, \mathcal{F}_{T_{(n-1)}}\right)\right)\right)\delta_{L\left(T_{(n-1)}\right)}(\mathrm{d}l) \\
&+ \mathbb{1}\{x=\ell\}\mu_n\left(dl, t, \mathcal{F}_{T_{(n-1)}}\right)\delta_{A\left(T_{(n-1)}\right)}(\mathrm{d}a) \\
&+ \mathbb{1}\{x \in \{y,d\}\}\delta_{A\left(T_{(n-1)}\right)}(\mathrm{d}a)\delta_{L\left(T_{(n-1)}\right)}(\mathrm{d}l).
\end{aligned} \tag{12}$$

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

Further suppose that:

1. $\mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\}\tilde{S}\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) = \mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\}\tilde{S}^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right)S\left(t \mid \mathcal{F}_{T_{(n-1)}}\right)$.
2. $\tilde{S}^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) > \eta$ for all $t \in (0,\tau]$ and $n \in \{1, \ldots, K\}$ P-a.s. for some $\eta > 0$.

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

$$\begin{aligned}
\mathbb{1}\left\{\bar{\Delta}_{(k-1)} \neq c\right\}\bar{Q}^g_{k-1,\tau} = \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \neq c\right\}\mathbb{E}_P\Bigg[&\frac{\mathbb{1}\left\{\bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = \ell\right\}}{\tilde{S}^c\left(\bar{T}_{(k-1)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\bar{Q}^g_{k,\tau}\left(A(\bar{T}_{k-1}), L(\bar{T}_k), \bar{T}_{(k)}, \bar{\Delta}_{(k)}, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \\
&+ \frac{\mathbb{1}\left\{\bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = a\right\}}{\tilde{S}^c\left(\bar{T}_{(k-1)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\bar{Q}^g_{k,\tau}\left(1, L(\bar{T}_{k-1}), \bar{T}_{(k)}, \bar{\Delta}_{(k)}, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \\
&+ \frac{\mathbb{1}\left\{\bar{T}_{(k)} \leq \tau, \bar{\Delta}_{(k)} = y\right\}}{\tilde{S}^c\left(\bar{T}_{(k-1)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \Bigg| \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\Bigg]
\end{aligned} \tag{13}$$

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

$$\Psi^g_\tau(P) = \mathbb{E}_P\left[\bar{Q}^g_{0,\tau}(1, L(0))\right]. \tag{14}$$

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

$$\Lambda^\alpha(\mathrm{d}(t,m,a,l)) = K'((L(0),A(0)), N^\alpha, t, \mathrm{d}(m,a,l))V'((A(0),L(0)), N^\alpha, \mathrm{d}t) \tag{15}$$

for some kernel K' from $\{0,1\} \times \mathcal{L} \times N_{\boldsymbol{X}} \times \mathbb{R}_+$ to $\mathbb{R}_+ \times \boldsymbol{X}$ and some predictable kernel V' from $\{0,1\} \times \mathcal{L} \times N_{\boldsymbol{X}} \times \mathbb{R}_+$ to $\mathbb{R}_+ \times \boldsymbol{X}$, because the *canonical* compensator is uniquely determined (so we first find the canonical compensator for the smaller filtration $\mathcal{F}^\alpha_t$ and then conclude that it must also be the canonical compensator for the larger filtration $\mathcal{F}^\beta_t$).

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

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

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

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

Then $\rho((L(0), A(0)), N^\beta, d(t, m, a, l))$ is a compensator, so it is by definition the canonical compensator. In view of Theorem 4.3.8 of Last & Brandt (1995),

$$K''\Big((l_0, a_0), \mathcal{F}_{\bar{T}_{(n-1)}}, t\Big) V'\Big((a_0, l_0), \mathcal{F}_{\bar{T}_{(n-1)}}, (0, t]\Big) = \tilde{\Lambda}_n^c\Big(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\Big).$$

and similarly, we see that

$$K'\Big((l_0, a_0), \mathcal{F}_{\bar{T}_{(n-1)}}, t, \mathrm{d}(m, a, l)\Big) V'\Big((a_0, l_0), \mathcal{F}_{\bar{T}_{(n-1)}}, \mathrm{d}(t, m, a, l)\Big) = \sum_{x=a,\ell,d,y} \psi_{n,x}\Big(t, d(a, l), \mathcal{F}_{T_{(n-1)}}\Big) \Lambda_n^x\Big((0, t] \mid \mathcal{F}_{T_{(n-1)}}\Big)$$

Let $T^*_{(k)}$ denote the ordered event times of the process N^β. With $S := T^e \wedge C \wedge T_{(k)}$, we have $T_{S,0} = T^e \wedge C \wedge T_{(k)} = \bar{T}_{(k)}$. Using Theorem 4.3.8 of Last & Brandt (1995), it therefore holds that

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

Further note that $T^*_k = \bar{T}_{(k)}$ whenever $T_{(k-1)} < C$. By definition, $T_{S,1} = T^*_{k+1}$ if $N^\beta\big(T_{(k)} \wedge C \wedge T^e\big) = k$. If $\bar{\Delta}_{(1)} \notin \{c, y, d\}, ..., \bar{\Delta}_{(k)} \notin \{c, y, d\}$, then $N^\beta\big(T_{(k)} \wedge C \wedge T^e\big) = k$ and furthermore $T^*_{k+1} = \bar{T}_{(k+1)}$, so

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

and we are done. From this, we get Equation 11. Applying this to the right hand side of Equation 13 shows that it is equal to Equation 3. □

Note that Equation 11 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. The stated conditional independence condition in Theorem 3 is likely much stronger than we need for identification, but is overall simple. Theorem 4 also gives a criterion, but is more generally stated. A simple consequence of the second is that if compensator of the (observed) censoring process is absolutely continuous with respect to the Lebesgue measure, then the survival function factorizes.

Theorem 3: 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

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

and the local independence statement given in Equation 11 holds.

Proof: We first show the local independence statement $\tilde{\mathcal{F}}^\beta = \sigma(\alpha(s) \mid s \le t) \vee \mathcal{Z}_0$, 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 (since $\mathbb{E}_P[N_t] \leq K < \infty$, we may work with martingales instead of *local* martingales). It follows that

$$\begin{aligned}
&\mathbb{E}_P\left[M^\alpha((0,t] \times \cdot \times \cdot \times \cdot) \mid \mathcal{F}_s^\beta\right] \\
&\overset{\text{(i)}}{=} \mathbb{E}_P\left[\mathbb{E}_P\left[M^\alpha((0,t] \times \cdot \times \cdot \times \cdot) \mid \widetilde{\mathcal{F}}_s^\beta\right] \mid \mathcal{F}_s^\beta\right] \\
&\overset{\text{(ii)}}{=} \mathbb{E}_P\left[M^\alpha((0,s] \times \cdot \times \cdot \times \cdot) \mid \mathcal{F}_s^\beta\right] \\
&\overset{\text{(iii)}}{=} M^\alpha((0,s] \times \cdot \times \cdot \times \cdot)
\end{aligned}$$

which implies the desired statement. In part (i), we use the law of iterated expectations, in part (ii), we use that the martingale is a martingale for $\widetilde{\mathcal{F}}_t^\beta$. In part (iii), we use that the martingale is $\mathcal{F}_t^\alpha$-adapted. This shows the desired local independence statement.

By the stated independence condition, it follows immediately that

$$\begin{aligned}
\mathbb{1}\left\{\bar{\Delta}_{(k)} \neq c\right\}\tilde{S}\left(t \mid \mathcal{F}_{T_{(k)}}\right) &= \mathbb{1}\left\{\bar{\Delta}_{(k)} \neq c\right\}P\left(\min\left\{T_{(k)}, C\right\} > t \mid \mathcal{F}_{T_{(k)}}\right) \\
&= \mathbb{1}\left\{\bar{\Delta}_{(k)} \neq c\right\}P\left(T_{(k)} > t, C > t \mid \mathcal{F}_{T_{(k)}}\right) \\
&= \mathbb{1}\left\{\bar{\Delta}_{(k)} \neq c\right\}S\left(t \mid \mathcal{F}_{T_{(k)}}\right)S^c\left(t \mid \mathcal{F}_{T_{(k)}}\right).
\end{aligned}$$

By the first part of Theorem 2, we have that

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

and we have that

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

so it follows that we just need to show that

$$\mathbb{1}\left\{\bar{\Delta}_{(k)} \neq c\right\} \prod_{s \in (0,t]} \left(1 - \Lambda_k^c\left(\mathrm{d}s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right) = \mathbb{1}\left\{\bar{\Delta}_{(k)} \neq c\right\} \prod_{s \in (0,t]} \left(1 - \tilde{\Lambda}_k^c\left(\mathrm{d}s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)$$

Because of the independence condition (using, for instance, Tonelli's theorem ???), we have that

$$P\left(\bar{T}_{(k)} \leq t, \bar{\Delta}_{(k)} = c \mid \mathcal{F}_{T_{(k-1)}}\right) = P\left(C \leq t, C > T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\right) = \int_{(0,t]} S\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)S^c\left(s- \mid \mathcal{F}_{T_{(k-1)}}\right)\Lambda_k^c\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right) \quad (16)$$

and thus

$$\tilde{\Lambda}_k^c\left(\mathrm{d}t \mid \mathcal{F}_{T_{(k-1)}}\right) = \frac{P\left(\bar{T}_{(k)} \in \mathrm{d}t, \bar{\Delta}_{(k)} = c \mid \mathcal{F}_{T_{(k-1)}}\right)}{S\left(t- \mid \mathcal{F}_{T_{(k-1)}}\right)S^c\left(t- \mid \mathcal{F}_{T_{(k-1)}}\right)} = \Lambda_k^c\left(\mathrm{d}t \mid \mathcal{F}_{T_{(k-1)}}\right)$$

where

$$S^c\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = \prod_{s \in (0,t]} \left(1 - \Lambda_k^c\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right)\right).$$

□

Theorem 4: Assume independent censoring as in Theorem 2. Then the left limit of the survival function factorizes on $(0, \tau]$, i.e.,

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

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

$$\Delta\tilde{\Lambda}_k^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) + \sum_x \Delta\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.} \Rightarrow \Delta\tilde{\Lambda}_{k+1}^c\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.} \vee \sum_x \Delta\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.}$$

Proof: First, we argue that for every $t \in (0, \tau]$ with $\tilde{S}\left(t- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) > 0$ (so dependent on the history), we have To show this, consider the quadratic covariation process which by the no simultaneous jump condition implies is zero, and thus

$$0 = \left[M^c(\cdot \wedge T^e), \sum_x M^x(\cdot \wedge C)\right]_t = \int_0^t \Delta\tilde{\Lambda}_c \sum_{x=a,\ell,y,d} d\Lambda_x$$

where $\tilde{\Lambda}_c$ and Λ_x are the compensators of the censoring process and the rest of the counting processes, respectively. Using this, we have

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

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

$$\mathbb{1}\left\{T_{(k-1)} \wedge C < t\right\}\tilde{S}\left(t- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \sum_{\bar{T}_{(k)} < s \leq t} \Delta\tilde{\Lambda}^c_k\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\left(\sum_{x=a,\ell,y,d} \Delta\Lambda^x_k\left(s, \mathcal{F}_{T_{(k-1)}}\right)\right) = 0, \tag{17}$$

where we also use that $\Delta\tilde{\Lambda}^c_k\left(s, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right) \neq 0$ for only a countable number of s's. This already means that the continuous part of the Lebesgue-Steltjes integral is zero, and thus the integral is evaluated to the sum in Equation 17. It follows that for every t with $\tilde{S}\left(t- \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_k}\right) > 0$,

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

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

Then, we have shown that

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

since

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

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

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

from which the result follows. (**NOTE**: We already the seen implication of the first part to the second part since $\Delta\gamma(u) + \Delta\zeta(u) \leq 1$; otherwise the survival function given in Theorem 2 would not be well-defined.) □

7 Efficient estimation

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

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

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

The efficient influence function in the nonparametric setting enables the use of machine learning methods to estimate the nuisance parameters under certain regularity conditions to provide inference for the target parameter. To achieve this, we need to debias our initial ICE-IPCW estimator either through double/debiased machine learning (Chernozhukov et al. (2018)) or targeted minimum loss estimation (van der Laan & Rubin (2006)). A method for constructing this estimator is presented in Section 7.1.

We derive the efficient influence function using the iterative representation given in Equation 13, working under the conclusions of Theorem 2, by finding the Gateaux derivative of the target parameter. Note that this does not constitute a rigorous proof that Equation 19 is the efficient influence function, but rather a heuristic argument. To proceed, we introduce additional notation and define

$$\bar{Q}_{k,\tau}^g\left(u \mid \mathcal{F}_{T_{(k)}}\right) = p_{ka}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k\ell}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{ky}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right), u < \tau. \tag{18}$$

A key feature of our approach is that the efficient influence function is expressed in terms of the martingale for the censoring process. This representation is often computationally simpler, as it avoids the need to estimate multiple martingale terms, unlike the approach of Rytgaard et al. (2022). For a detailed comparison, we refer the reader to the appendix, where we show that our efficient influence function is the same as the one derived by Rytgaard et al. (2022) in the setting with no competing events (**NOTE:** The section in the appendix is incomplete).

Theorem 5 (Efficient influence function): Let for each $P \in \mathcal{M}$, $\tilde{\Lambda}_k^c\left(t \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}; P\right)$ be the corresponding cause-specific cumulative hazard function for the observed censoring for the k'th event. Suppose that there is a universal constant $C > 0$ such that $\tilde{\Lambda}_k^c\left(\tau_{\text{end}} \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}; P\right) \leq C$ for all $k = 1, \dots, K$ and every $P \in \mathcal{M}$. The efficient influence function is then given by

$$\begin{aligned}\varphi_\tau^*(P) = {} & \frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))} \sum_{k=1}^K \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{A(\bar{T}_j)=1\}}{\pi_j\left(\bar{T}_{(j)}, \mathcal{F}_{T_{(j-1)}}\right)}\right)^{\mathbb{1}\{\bar{\Delta}_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} \tilde{S}^c\left(\bar{T}_{(j)} - \mid \mathcal{F}_{\bar{T}_{(j-1)}}^{\tilde{\beta}}\right)} \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \in \{\ell, a\}, \bar{T}_{(k-1)} < \tau\right\} \\ & \times \left(\left(\bar{Z}_{k,\tau}^a - \bar{Q}_{k-1,\tau}^g\right) + \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \left(\bar{Q}_{k-1,\tau}^g\left(\tau \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}\right) - \bar{Q}_{k-1,\tau}^g\left(u \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}\right)\right) \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}\right) S\left(u - \mid \mathcal{F}_{\bar{T}_{(k-1)}}^{\tilde{\beta}}\right)} \widetilde{M}^c(\mathrm{d}u)\right) \\ & + \bar{Q}_{0,\tau}^g - \Psi_\tau^g(P),\end{aligned} \tag{19}$$

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

Proof: Define

$$\bar{Z}^a_{k,\tau}(P \mid s, t_k, d_k, l_k, a_k, f_{k-1}) = \frac{\mathbb{1}\{t_k < s, d_k = \ell\}}{\tilde{S}^c_P\left(t_k - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)})\bar{Q}^g_{k,\tau}(P \mid a_{k-1}, l_k, t_k, d_k, f_{k-1}) \\ + \frac{\mathbb{1}\{t_k < s, d_k = a\}}{\tilde{S}^c_P\left(t_k - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}\bar{Q}^g_{k,\tau}(P \mid 1, l_{k-1}, t_k, d_k, f_{k-1}) \\ + \frac{\mathbb{1}\{t_k \leq s, d_k = y\}}{\tilde{S}^c_P\left(t_k - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}), s \leq \tau \tag{20}$$

and let

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

We compute the efficient influence function by calculating the derivative (Gateaux derivative) of $\Psi^g_\tau(P_\varepsilon)$ with $P_\varepsilon = P + \varepsilon(\delta_O - P)$ at $\varepsilon = 0$.

First note that:

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

so that

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \prod_{u \in (s,t)} \left(1 - \tilde{\Lambda}^c_{k,\varepsilon}\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right) \\
&= \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(t, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)} \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_{k,\varepsilon}\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right) \\
&\overset{(*)}{=} -\frac{1}{1 - \Delta\tilde{\Lambda}^c_{k,\varepsilon}\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)} \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right) \int_{(s,t]} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \tilde{\Lambda}^c_k\left(\mathrm{d}u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&\quad + \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right) \frac{1}{\left(1 - \Delta\tilde{\Lambda}^c_k\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)\right)^2} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \Delta\tilde{\Lambda}^c_{k,\varepsilon}\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right) \\
&= -\frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)} \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right) \int_{(s,t)} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \tilde{\Lambda}^c_k\left(\mathrm{d}u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&\quad -\frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)} \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right) \int_{\{t\}} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \tilde{\Lambda}^c_k\left(\mathrm{d}u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) \\
&\quad + \prod_{u \in (s,t]} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right) \frac{1}{\left(1 - \Delta\tilde{\Lambda}^c_k\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)\right)^2} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \Delta\tilde{\Lambda}^c_{k,\varepsilon}\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right) \\
&= -\prod_{u \in (s,t)} \left(1 - \tilde{\Lambda}^c_k\left(\mathrm{d}t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right) \int_{(s,t)} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \tilde{\Lambda}^c_k\left(\mathrm{d}u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right).
\end{aligned}
$$

In $(*)$, we use the product rule of differentiation and a well known result for product integration (Theorem 8 of Gill & Johansen (1990)), which states that the (Hadamard) derivative of the product integral $\mu \mapsto \prod_{u \in (s,t]}(1 + \mu(u))$ in the direction h is given by (for μ is of uniformly bounded variation)

$$
\int_{(s,t]} \prod_{v \in (s,u)} (1 + \mu(\mathrm{d}v)) \prod_{v \in (u,t]} (1 + \mu(\mathrm{d}v)) h(\mathrm{d}u) = \prod_{v \in (s,t]} (1 + \mu(\mathrm{d}v)) \int_{(s,t]} \frac{1}{1 + \Delta\mu(u)} h(\mathrm{d}u)
$$

Furthermore, for $P_\varepsilon = P + \varepsilon\left(\delta_{(X,Y)} - P\right)$, a simple calculation yields the well-known result

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

Now, we are ready to recursively calculate the derivative

$$
\frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} \bar{Q}^a_{k-1,\tau}(P_\varepsilon \mid a_{k-1}, l_{k-1}, t_{t-1}, d_{k-1}, f_{k-2})
$$

which yields,

$$
\begin{aligned}
&\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \bar{Q}^a_{k-1,\tau}(P_\varepsilon \mid a_{k-1}, l_{k-1}, t_{t-1}, d_{k-1}, f_{k-2}) \\
&= \frac{\delta_{\mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}(f_{k-1})}}{P\left(\mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}} = f_{k-1}\right)} \Bigg(\bar Z^a_{k,\tau} - \bar Q^g_{k-1,\tau}\left(\tau, \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) + \\
&+ \int_{\bar T_{(k-1)}}^{\tau} \bar Z^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{(\bar T_{(k-1)}, t_k)} \frac{1}{1 - \Delta\tilde\Lambda^c_k\left(s \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}} = f_{k-1}\right)} \frac{1}{\tilde S\left(s - \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}} = f_{k-1}\right)} \left(\widetilde N^c(\mathrm{d}s) - \tilde\Lambda^c(\mathrm{d}s)\right) \\
&\qquad P_{\left(\bar T_{(k)}, \bar\Delta_{(k)}, L(\bar T_k), A(\bar T_k)\right)}\left(\mathrm{d}(t_k, d_k, l_k, a_k) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}} = f_{k-1}\right)\Bigg) \\
&+ \int_{\bar T_{(k-1)}}^{\tau} \frac{\mathbb{1}\{t_k < \tau, d_k \in \{a, \ell\}\}}{\tilde S^c\left(t_k - \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}} = f_{k-1}\right)} \left(\frac{\mathbb{1}\{a_k = 1\}}{\pi_k\left(t_k, \mathcal{F}_{T_{(k-1)}}\right)}\right)^{\mathbb{1}\{d_k = a\}} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} \bar Q^g_{k,\tau}(P_\varepsilon \mid a_k, l_k, t_k, d_k, f_{k-1}\big) \\
&\qquad P_{\left(\bar T_{(k)}, \bar\Delta_{(k)}, L(\bar T_k), A(\bar T_k)\right)}\left(\mathrm{d}(t_k, d_k, l_k, a_k) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}} = f_{k-1}\right)
\end{aligned}
$$

Now note for the second term, we can write

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

by an exchange of integrals. Here, we apply the result of Theorem 2 to get the last equation. Combining the results iteratively yields the result. □

7.1 One-step ICE-IPCW estimator

In this section, we provide a one step estimator for the target parameter Ψ^g_τ. For a collection of estimators $\eta = \left(\left\{\hat\Lambda^x_k\right\}, \left\{\hat\Lambda^c_k\right\}, \{\hat\pi_k\}, \{\nu_{k,\tau}\}, \{\tilde\nu_{k,\tau}\}, \hat P_{L(0)}\right)$, we consider plug-in estimates of the efficient influence function

$$
\begin{aligned}
\varphi^*_\tau(O;\eta) &= \frac{\mathbb{1}\{A(0)=1\}}{\hat{\pi}_0(L(0))} \sum_{k=1}^{K} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{A(\bar{T}_j)=1\}}{\pi_j\left(\bar{T}_{(j)}, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(j-1)}}\right)} \right)^{\mathbb{1}\{\bar{\Delta}_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} \hat{S}^c\left(\bar{T}_{(j)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(j-1)}}\right)} \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \in \{\ell, a\}, \bar{T}_{(k-1)} < \tau\right\} \\
&\quad \times \Bigg(\left(\bar{Z}^a_{k,\tau}\left(\hat{S}^c_{k-1}, \nu_{k,\tau}\right) - \nu_{k-1,\tau}\right) \\
&\qquad + \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) - \mu_{k-1}\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right) \frac{1}{\hat{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \hat{S}\left(u - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \left(\tilde{N}^c(\mathrm{d}u) - \tilde{\Lambda}^c\left(\mathrm{d}u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right) \Bigg) \\
&\quad + \nu_{0,\tau}(1, L(0)) - \hat{P}_{L(0)}\left[\nu_{0,\tau}(1,\cdot)\right]
\end{aligned}
\tag{21}
$$

where

$$
\begin{aligned}
\mu_k\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) &= \int_{\bar{T}_{(k)}}^{u} \prod_{s \in \left(\bar{T}_{(k)}, u\right)} \left(1 - \sum_{x=a,l,d,y} \hat{\Lambda}^x_k\left(\mathrm{d}s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right)\right) \\
&\quad \times \left[\hat{\Lambda}^y_{k+1}\left(\mathrm{d}s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) + \mathbb{1}\{s < u\}\tilde{\nu}_{k+1,\tau}\left(1, s, a, \mathcal{F}_{T_{(k)}}\right)\hat{\Lambda}^a_{k+1}\left(\mathrm{d}s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) + \mathbb{1}\{s < u\}\tilde{\nu}_{k+1,\tau}\left(A\left(T_{(k-1)}\right), s, \ell, \mathcal{F}_{T_{(k)}}\right)\hat{\Lambda}^\ell_{k+1}\left(\mathrm{d}s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right)\right].
\end{aligned}
\tag{22}
$$

Here, we let $\tilde{\nu}_{k,\tau}(u \mid f_k)$ be an estimate of $\bar{Q}^{g,-L}_{k,\tau}(u \mid f_k) := \mathbb{E}_P\left[\bar{Q}^g_{k,\tau}\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) \mid A(\bar{T}_k) = a_k, \bar{\Delta}_{(k)} = d_k, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right]$, let $\nu_{k,\tau}(f_k)$ be an estimate of $\bar{Q}^g_{k,\tau}(\tau \mid f_k)$, and let $\hat{P}_{L(0)}$ be an estimate of $P_{L(0)}$, the distribution of the covariates at time 0. We use the notation $\bar{Z}^a_{k,\tau}\left(\tilde{S}^c_{k-1}, \nu_{k,\tau}\right)$ to explicitly denote the dependency on $\tilde{S}^c_{k-1}$ and $\nu_{k,\tau}$.

We will now describe how to estimate the efficient influence function in practice. Overall, we consider the same procedure as in Section 5 with additional steps:

1. For $\{\nu_{k,\tau}(f_k)\}$, use the procedure described in Section 5.
2. For $\{\tilde{\nu}_{k,\tau}(f_k)\}$ use a completely similar procedure to the one given in Section 5 using the estimator $\nu_{k+1,\tau}$ to obtain $\tilde{\nu}_{k,\tau}$. Now we do not include the latest time varying covariate $L(\bar{T}_k)$ in the regression, so that $\tilde{\nu}_{k-1,\tau} = \mathbb{E}_{\hat{P}}\left[\bar{Z}^a_{k,\tau}\left(\hat{S}^c_{k-1}, \nu_{k,\tau}\right) \mid A(\bar{T}_k) = a_k, \bar{\Delta}_{(k)} = d_k, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right]$.
3. Find $\left\{\hat{\Lambda}^x_k\right\}$ for $x = a, \ell, d, y$ and $\left\{\hat{\Lambda}^c_k\right\}$ using step 1 in Section 5.
4. Obtain an estimator of the propensity score $\{\pi_k(t, f_{k-1})\}$ by regressing $A(\bar{T}_k)$ on $\left(\bar{T}_{(k)}, \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$ among subjects with $\bar{\Delta}_{(k)} = a$ and $\bar{\Delta}_{(k-1)} \in \{a, \ell\}$ for each k and for $k = 0$ estimate $\pi_0(L(0))$ by regressing $A(0)$ on $L(0)$.
5. Use the estimates $\tilde{\nu}_{k,\tau}(f_k)$ and the estimates of $\hat{\Lambda}^x_k, x = a, \ell, d, y$ to numerically compute μ_{k-1} via Equation 22.
6. Use the estimated survival functions from the cumulative hazards in step 3 to compute the martingale term in Equation 21. See also Section 7.2 for details on how to approximately compute the censoring martingale term.
7. Substitute the rest of the estimates into Equation 21 and obtain the estimate of the efficient influence function.

There are multiple computational aspects of the stated procedure that should be addressed. First note that $\bar{Q}^g_{k,\tau}(\tau)$ is estimated twice. This redundancy is intentional: it ensures both the computability of the terms involved in the censoring martingale and that we can use $\nu_{k,\tau}$ required for subsequent iterations of the algorithm (avoiding the high dimensionality of the integrals as discussed in Section 3).

Our decision to use $\nu_{k,\tau}$ rather than $\mu_{k,\tau}$ as an estimator for the regression terms $(\bar{Z}^a_{k,\tau} - \bar{Q}^g_{k-1,\tau})$ in Equation 21 is motivated by practical considerations. In particular, numerical integration may yield less accurate results due to the need to compute Λ^x_k for $x = a, \ell, d, y$. In practice, the contribution of the censoring martingale to the efficient influence function is typically small. As such, a simplified procedure that excludes the censoring martingale term or one that computes the censoring martingale term only at a sparse grid of time points may offer substantial computational efficacy with minimal bias.

It is also more efficient computationally to use $\nu_{k,\tau}$ rather than $\mu_{k,\tau}$. To see this for $k=1$, note that we would not only need to compute $\mu_{0,\tau}(1, L_i(0))$ for $i \in \{1,...,n\}$ with $A_i(0)=1$, but for all $i=1,...,n$ to estimate the term in the efficient influence function given by $\mu_{0,\tau}(1, L_i(0))$.

We have also elected not to estimate Equation 18 using the procedure described in the algorithm in Section 5 (ICE-IPCW), as it may be prohibitively expensive to do so even along a sparse grid of time points. Moreover, the resulting estimators are not guaranteed to be monotone in u which $\bar{Q}^g_{k,\tau}\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\tilde{T}_{(k)}}\right)$ is.

Another alternative is to use parametric/semi-parametric models for the estimation of the cumulative cause-specific hazard functions for the censoring. In that case, we may not actually need to debias the censoring martingale, but can still apply machine learning methods to iterated regressions.

Now, we turn to the resulting one-step procedure. For the ICE-IPCW estimator $\hat{\Psi}^0_n$, we let $\hat{\eta} = \left(\left\{\hat{\Lambda}^x_k\right\}_{k,x}, \left\{\hat{\Lambda}^c_k\right\}, \{\hat{\pi}_k\}_k, \{\hat{\nu}_{k,\tau}\}_k, \{\tilde{\nu}_{k,\tau}\}_k, \mathbb{P}_n\right)$ be a given estimator of the nuisance parameters occuring in $\varphi^*_\tau(O;\eta)$, where $\mathbb{P}_n$ denotes the empirical distribution, and consider the decomposition

$$\begin{aligned}\hat{\Psi}^0_n - \Psi^g_\tau(P) &= \mathbb{P}_n \varphi^*_\tau(\cdot;\eta) \\ &-\mathbb{P}_n \varphi^*_\tau(\cdot;\hat{\eta}) \\ &+(\mathbb{P}_n - P)(\varphi^*_\tau(\cdot;\hat{\eta}) - \varphi^*_\tau(\cdot;\eta)) \\ &+R_2(\eta,\hat{\eta}),\end{aligned}$$

where

$$R_2(\eta,\eta') = P_\eta[\varphi^*_\tau(\cdot;\eta')] + \Psi^{\mathrm{obs}}_\tau(\eta') - \Psi^g_\tau(\eta)$$

and $\Psi^g_\tau(\hat{\eta}) = \mathbb{P}_n[\nu_{0,\tau}(1,\cdot)]$. We consider one-step estimation, that is

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

This means that to show that $\hat{\Psi}_n - \Psi^g_\tau(P) = \mathbb{P}_n \varphi^*_\tau(\cdot;\eta) + o_P\left(n^{-\frac{1}{2}}\right)$, we must show that

$$R_2(\eta,\hat{\eta}) = o_P\left(n^{-\frac{1}{2}}\right), \tag{23}$$

and that the empirical process term fulfills

$$(\mathbb{P}_n - P)(\varphi^*_\tau(\cdot;\hat{\eta}) - \varphi^*_\tau(\cdot;\eta)) = o_P\left(n^{-\frac{1}{2}}\right). \tag{24}$$

We first discuss how to show Equation 24. This can be shown (Lemma 19.24 of van der Vaart (1998)) if

1. $\varphi^*_\tau(\cdot;\hat{\eta}) \in \mathcal{Z}$ for some P-Donsker class $\mathcal{Z}$ of functions with probability tending to 1, and
2. $\|\varphi^*_\tau(\cdot;\hat{\eta}) - \varphi^*_\tau(\cdot;\eta)\|_{L^2_P(O)} = o_P(1)$, with $\|f\|_{L^2_P(O)} = (\mathbb{E}_P[f(O)^2])^{\frac{1}{2}}$.

Simple sufficient conditions for this to happen are provided in Lemma **NOT DONE YET**. Alternatively, one may use cross-fitting/sample splitting (Chernozhukov et al. (2018)) to ensure that the empirical process term is negligible.

To obtain the rates in Equation 23, we find the second order remainder term $R_2(\eta_0,\eta)$ and show that it has a product structure (Theorem 6). This allows us to use estimators which need only convert at $L_2(P)$-rates of at least $o_P\left(n^{-\frac{1}{4}}\right)$ under regularity conditions.

Theorem 6 (Second order remainder): Let $\eta_0 = \left(\left\{\Lambda^x_{k,0}\right\}_{k,x}, \left\{\tilde{\Lambda}^c_{k,0}\right\}_k, \{\pi_{0,k}\}_k, \left\{\bar{Q}^g_{k,\tau}\right\}_k, \left\{\bar{Q}^{g,-L}_{k,\tau}\right\}_k, P_{0,L(0)}\right)$ be the true parameter values and let $\eta = \left(\{\Lambda^x_k\}_{k,x}, \left\{\tilde{\Lambda}^c_k\right\}_k, \{\pi_k\}_k, \{\nu_{k,\tau}\}_k, \{\tilde{\nu}_{k,\tau}\}_k, P'_{L(0)}\right)$. The remainder term $R_2(\eta_0, \eta)$ is given by

$$
\begin{aligned}
R_2(\eta_0, \eta) = \sum_{k=1}^{K-1} \int & \mathbb{1}\{t_1 < \dots < t_k < \tau\}\mathbb{1}\{a_0 = \dots = a_k = 1\} \\
& \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \prod_{j=1}^{k} \left(\frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})}\right)^{\mathbb{1}\{d_j = a\}} \frac{1}{\tilde{S}^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} z_k(f_k) P_{0,\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}}(\mathrm{d}f_k) \\
& + \int \mathbb{1}\{a_0 = 1\} z_0(a_0, l_0) P_{0,L(0)}(\mathrm{d}l_0)
\end{aligned}
$$

where

$$
\begin{aligned}
z_k(f_k) = & \left(\left(\frac{\pi_{k,0}(t_k, f_{k-1})}{\pi_k(t_k, f_{k-1})}\right)^{\mathbb{1}\{d_k = a\}} - 1\right)\left(\bar{Q}^g_{k,\tau}(f_k) - \nu_{k,\tau}(f_k)\right) \\
& + \left(\frac{\pi_{k,0}(t_k, f_{k-1})}{\pi_k(t_k, f_{k-1})}\right)^{\mathbb{1}\{d_k = a\}} \int_{t_k}^{\tau} \left(\frac{\tilde{S}^c_0(u - \mid f_k)}{\tilde{S}^c(u - \mid f_k)} - 1\right)\left(\bar{Q}^g_{k,\tau}(du \mid f_k) - \nu^*_{k,\tau}(du \mid f_k)\right) \\
& + \left(\frac{\pi_{k,0}(t_k, f_{k-1})}{\pi_k(t_k, f_{k-1})}\right)^{\mathbb{1}\{d_k = a\}} \int_{t_k}^{\tau} V_{k+1}(u, f_k) \nu^*_{k,\tau}(du \mid f_k),
\end{aligned}
$$

for $k \geq 1$ and for $k = 0$

$$
\begin{aligned}
z_0(1, l_0) = & \left(\frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} - 1\right)\left(\bar{Q}^g_{0,\tau}(1, l_0) - \nu_{0,\tau}(1, l_0)\right) \\
& + \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \int_0^{\tau} \left(\frac{\tilde{S}^c_0(s - \mid 1, l_0)}{\tilde{S}^c(s - \mid 1, l_0)} - 1\right)\left(\bar{Q}^g_{0,\tau}(ds \mid 1, l_0) - \nu^*_{0,\tau}(ds \mid 1, l_0)\right) \\
& + \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \int_0^{\tau} V_1(u, 1, l_0) \nu^*_{0,\tau}(du \mid 1, l_0),
\end{aligned}
$$

and $V_k(u, f_{k-1}) = \int_{(t_{k-1}, u)} \left(\frac{S_0(s - \mid f_{k-1})}{S(s - \mid f_{k-1})} - 1\right) \frac{\tilde{S}^c_0(s - \mid f_{k-1})}{\tilde{S}^c(s \mid f_{k-1})} \left(\tilde{\Lambda}^c_{k,0}(ds \mid f_{k-1}) - \tilde{\Lambda}^c_k(ds \mid f_{k-1})\right)$.

Proof: First define $\varphi^*_{k,\tau}(O; \eta)$ for $k > 0$ to be the k'th term in the efficient influence function given in Equation 19, and let $\varphi^*_{0,\tau}(O; \eta) = \nu_0(1, L(0)) - \Psi^{\text{obs}}_\tau(\eta)$, so that $\varphi^*_\tau(O; P) = \sum_{k=0}^{K} \varphi^*_k(O; P)$.

Then, we first note that

$$
\mathbb{E}_{P_0}\left[\varphi^*_{0,\tau}(O; \eta)\right] + \Psi^g_\tau(\eta) - \Psi^g_\tau(\eta_0) = \mathbb{E}_{P_0}\left[\nu_0(1, L(0)) - \bar{Q}^g_{0,\tau}(1, L(0))\right]. \tag{25}
$$

Apply the law of iterated expectation to the efficient influence function in Equation 19 to get

$$
\begin{aligned}
\mathbb{E}_{P_0}\left[\varphi^*_{k,\tau}(O; P)\right] = \int & \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\}\mathbb{1}\{a_0 = \dots = a_{k-1} = 1\} \frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \\
& \times \prod_{j=1}^{k-1} \left(\frac{\pi_{0,j}(t_k, f_{j-1})}{\pi_j(t_k, f_{j-1})}\right)^{\mathbb{1}\{d_j = a\}} \frac{1}{\tilde{S}^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} \\
& \times \mathbb{E}_P\left[h_k\left(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right] P_{0,\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(\mathrm{d}f_{k-1})
\end{aligned}
$$

where

$$
h_k\left(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) = \bar{Z}^a_{k,\tau}\left(\tilde{S}, \nu_k\right) - \nu_{k-1} + \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) - \mu_{k-1}\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right) \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(u - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \widetilde{M}^c(\mathrm{d}u).
$$

Now note that

$$
\begin{aligned}
&\mathbb{E}_{P_0}\left[h_k\left(\mathcal{F}^{\tilde\beta}_{\bar T_{(k)}}\right) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] \\
&= \mathbb{E}_{P_0}\left[\bar Z^a_{k,\tau}\left(\tilde S^c_0, \bar Q^g_{k,\tau}\right)\right) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}] - \nu_{k-1,\tau}\left(\mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) \\
&+\mathbb{E}_{P_0}\left[\bar Z^a_{k,\tau}\left(\tilde S^c, \nu_k\right) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] - \mathbb{E}_{P_0}\left[\bar Z^a_{k,\tau}\left(\tilde S^c, \bar Q^g_{k,\tau}\right) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] \\
&+\mathbb{E}_{P_0}\left[\bar Z^a_{k,\tau}\left(\tilde S^c, \bar Q^g_{k,\tau}\right) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] - \mathbb{E}_{P_0}\left[\bar Z^a_{k,\tau}\left(\tilde S^c_0, \bar Q^g_{k,\tau}\right) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] \\
&+\mathbb{E}_{P_0}\left[\int_{\bar T_{(k-1)}}^{\tau\wedge \bar T_{(k)}} \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \mu_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right) \frac{1}{\tilde S^c\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)} \widetilde M^c(\mathrm{d}u) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right]
\end{aligned}
\tag{26}
$$

We shall need the following auxilliary result.

Lemma 1:

$$
\mathbb{E}_{P_0}\left[\int_{\bar T_{(k-1)}}^{\tau\wedge \bar T_{(k)}} \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \mu_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right) \frac{1}{\tilde S^c\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)} \widetilde M^c(\mathrm{d}u) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right]
$$

$$
= \int_{\bar T_{(k-1)}}^{\tau} \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \mu_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right) \frac{\tilde S^c_0\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S_0\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)}{\tilde S^c\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)} \left(\tilde\Lambda^c_{k,0}\left(du \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \tilde\Lambda^c_k\left(du \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right)
$$

Proof: We first note that

$$
\begin{aligned}
&\mathbb{E}_{P_0}\left[\int_{\bar T_{(k-1)}}^{\tau\wedge \bar T_{(k)}} \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \mu_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right) \frac{1}{\tilde S^c\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)} \tilde\Lambda^c(\mathrm{d}u) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] \\
&= \mathbb{E}_{P_0}\left[\int_{\bar T_{(k-1)}}^{\tau\wedge \bar T_{(k)}} \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \mu_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right) \frac{1}{\tilde S^c\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)} \tilde\Lambda^c_k\left(\mathrm{d}u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] \\
&= \mathbb{E}_{P_0}\left[\int_{\bar T_{(k-1)}}^{\tau} \mathbb{1}\left\{\bar T_{(k)} \le t\right\} \right. \\
&\qquad \left. \times \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \mu_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right) \frac{1}{\tilde S^c\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)} \tilde\Lambda^c_k\left(\mathrm{d}u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] \\
&= \int_{\bar T_{(k-1)}}^{\tau} \mathbb{E}_{P_0}\left[\mathbb{1}\left\{\bar T_{(k)} \le t\right\} \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right] \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \mu_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right) \frac{1}{\tilde S^c\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)} \tilde\Lambda^c_k\left(\mathrm{d}u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) \\
&= \int_{\bar T_{(k-1)}}^{\tau} \left(\mu_{k-1}\left(\tau \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) - \mu_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)\right) \frac{\tilde S^c_0\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S_0\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)}{\tilde S^c\left(u \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right) S\left(u- \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)} \tilde\Lambda^c_k\left(du \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)
\end{aligned}
\tag{27}
$$

by simply interchanging the integral and the expectation (see for instance Lemma 3.1.4 of Last & Brandt (1995)). Finally, let $A \in \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}$. Then, using the compensator of $\tilde N^c(t)$ under P_0 is $\tilde\Lambda^c_0 = \sum_{k=1}^K \mathbb{1}\left\{\bar T_{(k-1)} < t \le \bar T_{(k)}\right\} \tilde\Lambda^c_{k,0}\left(\mathrm{d}t \mid \mathcal{F}^{\tilde\beta}_{\bar T_{(k-1)}}\right)$ and that $\mathbb{1}\{A\}\mathbb{1}\left\{\bar T_{(k-1)} < s \le \bar T_{(k)}\right\}$ is predictable, we have

$$\mathbb{E}_{P_0}\left[\mathbb{1}\{A\}\mathbb{1}\left\{\bar{T}_{(k-1)}<\tau\right\}\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}}\left(\mu_{k-1}\left(\tau\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\frac{1}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\tilde{N}^c(\mathrm{d}t)\right]$$

$$=\mathbb{E}_{P_0}\left[\mathbb{1}\{A\}\mathbb{1}\left\{\bar{T}_{(k-1)}<\tau\right\}\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}}\left(\mu_{k-1}\left(\tau\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\frac{1}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\tilde{\Lambda}^c_0(\mathrm{d}t)\right]$$

$$=\mathbb{E}_{P_0}\left[\mathbb{1}\{A\}\mathbb{1}\left\{\bar{T}_{(k-1)}<\tau\right\}\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}}\left(\mu_{k-1}\left(\tau\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\frac{1}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\tilde{\Lambda}^c_{k,0}\left(\mathrm{d}t\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right]$$

from which we conclude that

$$\mathbb{E}_{P_0}\left[\mathbb{1}\left\{\bar{T}_{(k-1)}<\tau\right\}\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}}\left(\mu_{k-1}\left(\tau\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\frac{1}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\widetilde{N}^c(\mathrm{d}t)\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right]$$

$$=\mathbb{E}_{P_0}\left[\mathbb{1}\left\{\bar{T}_{(k-1)}<\tau\right\}\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}}\left(\mu_{k-1}\left(\tau\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\frac{1}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\tilde{\Lambda}^c_{0,k}\left(\mathrm{d}t\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right]$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\left(\mu_{k-1}\left(\tau\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\frac{\tilde{S}^c_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\tilde{\Lambda}^c_{k,0}\left(du\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$$

where the last equality follows from the same argument as in Equation 27. □

By an exchange of integrals, it follows that

$$\int_{\bar{T}_{(k-1)}}^{\tau}\left(\mu_{k-1}\left(\tau\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\frac{\tilde{S}^c_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}^c_{k,0}\left(du\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}^c_k\left(du\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\int_u^{\tau}\mu_{k-1}\left(\mathrm{d}s\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\frac{\tilde{S}^c_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}^c_{k,0}\left(du\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}^c_k\left(du\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\int_{(\bar{T}_{(k-1)},s)}\frac{\tilde{S}^c_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}^c_{k,0}\left(\mathrm{d}u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}^c_k\left(\mathrm{d}u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\int_{(\bar{T}_{(k-1)},s)}\left(\frac{S_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}-1\right)\frac{\tilde{S}^c_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}^c_{k,0}\left(\mathrm{d}u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}^c_k\left(\mathrm{d}u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$$

$$+\int_{\bar{T}_{(k-1)}}^{\tau}\int_{(\bar{T}_{(k-1)},s)}\frac{\tilde{S}^c_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}^c_{k,0}\left(\mathrm{d}u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}^c_k\left(\mathrm{d}u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\int_{(\bar{T}_{(k-1)},s)}\left(\frac{S_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{S\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}-1\right)\frac{\tilde{S}^c_0\left(u-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}^c_{k,0}\left(\mathrm{d}u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}^c_k\left(\mathrm{d}u\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$$

$$-\int_{\bar{T}_{(k-1)}}^{\tau}\left(\frac{\tilde{S}^c_0\left(s-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(s-\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}-1\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)$$

where we apply the Duhamel equation (taking the left limits of Equation (2.6.5) of Andersen et al. (1993)) in the last equality. Since

$$\mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right] - \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c_0, \bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right]$$
$$= \int_{\bar{T}_{(k-1)}}^{\tau} \left(\frac{\tilde{S}^c_0\left(u - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} - 1 \right) \bar{Q}^g_{k-1,\tau}\left(\mathrm{d}u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right),$$

it follows from Equation 26 and Lemma 1 that

$$\begin{aligned}
&\mathbb{E}_{P_0}\left[h_k\left(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right] \\
&= \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c_0, \bar{Q}^g_{k,\tau}\right)\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}] - \nu_{k-1,\tau}\left(\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \\
&\quad + \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \nu_k\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right] - \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right] \\
&\quad + \int_{\bar{T}_{(k-1)}}^{\tau} \left(\frac{\tilde{S}^c_0\left(u - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} - 1 \right) \left(\bar{Q}^g_{k-1,\tau}\left(du \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) - \mu_{k-1,\tau}\left(du \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right) \qquad (28) \\
&\quad + \int_{\bar{T}_{(k-1)}}^{\tau} \int_{\left(\bar{T}_{(k-1)}, u\right)} \left(\frac{S_0\left(s - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{S\left(s - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} - 1 \right) \frac{\tilde{S}^c_0\left(s - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(s \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \left(\tilde{\Lambda}^c_{k,0}\left(ds \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) - \tilde{\Lambda}^c_k\left(ds \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)\right) \mu_{k-1,\tau}\left(du \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)
\end{aligned}$$

Since it also holds for $k \geq 1$ that,

$$\begin{aligned}
&\int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\} \mathbb{1}\{a_0 = \dots = a_{k-1} = 1\} \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{1}{\prod_{j=1}^{k-1} \tilde{S}^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} \\
&\qquad \times \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \nu_k\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right] - \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right] P_{0,\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(df_{k-1}) \\
&= \int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\} \mathbb{1}\{a_0 = \dots = a_{k-1} = 1\} \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{1}{\prod_{j=1}^{k-1} \tilde{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\} \mathbb{1}\{a_k = 1\} \frac{1}{\tilde{S}^c(t_k - \mid f_{k-1})} \\
&\qquad \times \sum_{d_k = a, \ell} \left(\nu_k(a_k, l_k, t_k, d_k, f_{k-1}) - \bar{Q}^g_{k,\tau}(a_k, l_k, t_k, d_k, f_{k-1})\right) P_{0,\left(A(\bar{T}_k), L(\bar{T}_k), \bar{T}_{(k)}, \bar{\Delta}_{(k)}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(\mathrm{d}f_k \mid f_{k-1}) P_{0,\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(df_{k-1}) \\
&= \int \mathbb{1}\{t_1 < \dots < t_k < \tau\} \mathbb{1}\{a_0 = \dots = a_k = 1\} \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{1}{\prod_{j=1}^{k} \tilde{S}^c(t_j - \mid f_{j-1})} \mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
&\qquad \times \left(\nu_k(f_k) - \bar{Q}^g_{k,\tau}(f_k)\right) P_{0,\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}}(df_k)
\end{aligned}$$

we have that

$$
\begin{aligned}
&\int \mathbb{1}\{t_1 < \dots < t_k < \tau\}\mathbb{1}\{a_0 = \dots = a_k = 1\}\frac{\pi_{0,0}(l_0)}{\pi_0(l_0)}\prod_{j=1}^{k}\left(\frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})}\right)^{\mathbb{1}\{d_j=a\}}\frac{1}{\prod_{j=1}^{k}\tilde{S}^c(t_j - \mid f_{j-1})}\mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
&\qquad \times \left(\mathbb{E}_{P_0}\left[\bar{Z}^a_{k+1,\tau}\left(\tilde{S}^c_0, \bar{Q}^g_{k+1,\tau}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_k\right] - \nu_{k,\tau}(f_k)\right)P_{0,\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}}(df_k) \\
&+ \int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\}\mathbb{1}\{a_0 = \dots = a_{k-1} = 1\}\frac{\pi_{0,0}(l_0)}{\pi_0(l_0)}\prod_{j=1}^{k-1}\left(\frac{\pi_{0,j}(t_k, f_{j-1})}{\pi_j(t_k, f_{j-1})}\right)^{\mathbb{1}\{d_j=a\}}\frac{1}{\prod_{j=1}^{k-1}\tilde{S}^c(t_j - \mid f_{j-1})}\mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_{k-1} \in \{\ell, a\}\} \\
&\qquad \times \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \nu_k\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right] - \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right]P_{0,\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(df_{k-1}) \qquad (29) \\
&= \int \mathbb{1}\{t_1 < \dots < t_k < \tau\}\mathbb{1}\{a_0 = \dots = a_k = 1\}\frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \\
&\qquad \prod_{j=1}^{k-1}\left(\frac{\pi_{0,j}(t_j, f_{k-1})}{\pi_j(t_j, f_{j-1})}\right)^{\mathbb{1}\{d_k=a\}}\left(\left(\frac{\pi_{0,k}(k_j, f_{k-1})}{\pi_k(t_k, f_{k-1})}\right)^{\mathbb{1}\{d_k=a\}} - 1\right)\frac{1}{\prod_{j=1}^{k}\tilde{S}^c(t_j - \mid f_{j-1})}\mathbb{1}\{d_1 \in \{\ell, a\}, \dots, d_k \in \{\ell, a\}\} \\
&\qquad \times \left(\bar{Q}^g_{k,\tau}(f_k) - \nu_{k,\tau}(f_k)\right)P_{0,\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}}(df_k).
\end{aligned}
$$

By combining Equation 25, Equation 28 and Equation 29, we are done.

7.2 Algorithm for the calculation of censoring martingale

In this subsection, we present an algorithm for computing the martingale term in Equation 21 along a specified time grid $\{t_1, \dots, t_m\}$ at iteration k of the influence function estimation procedure. In Steps 6, 8, 10, and 11 of the algorithm, we may use coarse approximations for the survival function and the associated integrals. For example, one may approximate the survival function using the exponential function or apply numerical integration techniques such as Simpson's rule to simplify computation. Note that we integrate over time on the interarrival scale. This means that we usually select $t_1 = 0$ and $t_m \leq \tau - \min_i \bar{T}_{k+1,i}$.

$\underline{\text{CENSORINGMARTINGALE}}(k,\ \{t_1, ..., t_m\},\ \{\bar{T}_{k,i}, \bar{T}_{k+1,i}\},\ \left\{\mathcal{F}_{\bar{T}_{k,i}}\right\},\ \left\{\hat{\Lambda}^x_{k+1}\right\}_x,\ \tilde{\nu}_{k+1},\ \{A(\bar{T}_{k,i})\},\ \left\{\bar{\Delta}_{k+1,i}\right\})$:

for $i = 1, ..., n$:

$\quad j_{\max} \leftarrow \max\{v \mid t_v \leq \tau - \bar{T}_{k,i}\}$

$\quad \hat{\nu}^y_\tau(0) \leftarrow \hat{\nu}^a_\tau(0) \leftarrow \hat{\nu}^\ell_\tau(0) \leftarrow t_0 \leftarrow \widehat{M}^c(0) \leftarrow 0$

$\quad \hat{S}_0 \leftarrow 1$

$\quad$ **for** $j = 1, ..., j_{\max}$

$\quad\quad \hat{S}(s-) \leftarrow \prod_{v \in [t_{j-1}, s)} \left(1 - \sum_{x=a,l,d,y} \hat{\Lambda}^x_{k+1}\left(\mathrm{d}v \mid \mathcal{F}_{\bar{T}_{k,i}}\right)\right)$

$\quad\quad \hat{S}_j \leftarrow \hat{S}_{j-1} \cdot S(t_j)$

$\quad\quad \hat{\nu}^y_\tau(t_j) \leftarrow \hat{\nu}^y_\tau(t_{j-1}) + \hat{S}_{j-1} \int_{(t_{j-1}, t_j]} \hat{S}(s-) \hat{\Lambda}^y_{k+1}\left(\mathrm{d}s \mid \mathcal{F}_{\bar{T}_{k,i}}\right)$

$\quad\quad$ **if** $k < K_\tau$:

$\quad\quad\quad \hat{\nu}^a_\tau(t_j) \leftarrow \hat{\nu}^a_\tau(t_{j-1}) + \hat{S}_{j-1} \int_{(t_{j-1}, t_j)} \hat{S}(s-) \tilde{\nu}_{k+1}\left(1, s + \bar{T}_{k+1,i}, a, \mathcal{F}_{\bar{T}_{k,i}}\right) \hat{\Lambda}^a_{k+1}\left(\mathrm{d}s \mid \mathcal{F}_{\bar{T}_{k,i}}\right)$

$\quad\quad\quad \hat{\nu}^\ell_\tau(t_j) \leftarrow \hat{\nu}^\ell_\tau(t_{j-1}) + \hat{S}_{j-1} \int_{(t_{j-1}, t_j)} \hat{S}(s-) \tilde{\nu}_{k+1}\left(A(\bar{T}_{k,i}), s + \bar{T}_{k+1,i}, \ell, \mathcal{F}_{\bar{T}_{k,i}}\right) \hat{\Lambda}^\ell_{k+1}\left(\mathrm{d}s \mid \mathcal{F}_{\bar{T}_{k,i}}\right)$

$\quad\quad$ **else**:

$\quad\quad\quad \hat{\nu}^a_\tau(t_j) \leftarrow \hat{\nu}^a_\tau(t_j) \leftarrow 0$

$\quad\quad \hat{\nu}_\tau(t_j) \leftarrow \hat{\nu}^y_\tau(t_j) + \hat{\nu}^a_\tau(t_j) + \hat{\nu}^\ell_\tau(t_j)$

$\quad\quad \widehat{M}^c(t_j) \leftarrow \mathbb{1}\left\{\bar{\Delta}_i = c, \bar{T}_{k+1,i} - \bar{T}_{k,i} \leq t_j\right\} - \hat{\Lambda}^c_{k+1}\left(t_j \mid \mathcal{F}_{\bar{T}_{k,i}}\right)$

$\quad\quad \hat{S}^c(t_j) \leftarrow \prod_{v \in (0, t_j]} \left(1 - \hat{\Lambda}^c_{k+1}\left(\mathrm{d}v \mid \mathcal{F}_{\bar{T}_{k,i}}\right)\right)$

$\quad k_i \leftarrow \{v \mid t_v \leq \tau \wedge \bar{T}_{k+1,i} - \bar{T}_{k,i}\}$

$\quad \widehat{\mathrm{MG}}_i \leftarrow \sum_{j=1}^{k_i} \left(\hat{\nu}_\tau\left(t_{j_{\max}} \mid \mathcal{F}_{\bar{T}_{k,i}}\right) - \hat{\nu}_\tau\left(t_j \mid \mathcal{F}_{\bar{T}_{k,i}}\right)\right) \frac{1}{\hat{S}^c(t_j) \hat{S}_j} \left(\widehat{M}^c(t_j) - \widehat{M}^c(t_{j-1})\right)$

return $\widehat{\mathrm{MG}}$

8 Real data application

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

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

We also want to compare with other methods.

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

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

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`.

9 Simulation study

The data generating mechanism should be based on real data given in Section 8. Note that the simulation procedure follows the DAG in Figure 4. Depending on the results from the data application, we should consider:

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

10 Discussion

There is one main issue with the method that we have not discussed yet: In the case of irregular data, we may have few people with many events. For example there may only be 5 people in the data with a censoring event as their 4′th event. In that case, we can hardly estimate $\lambda_4^c\left(\cdot \mid \mathcal{F}_{T_{(3)}}\right)$ 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 $T_{K_i+1}^i = \tau_{\max}$). Thus

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

and that

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

So we see that the maximization problem corresponds exactly to finding the maximum likelihood estimator on a pooled data set! One may then apply a flexible method based on the likelihood, e.g., HAL to say a model that pools across all time points. One may then proceed greedily, using Donsker-class conditions, computing the validation based error of a model (likelihood) that pools across all event points except one. If the second model then performs better within some margin, we accept the new model and compare that with a model that pools all events points except two. Theory may be based on Theorem 1 of Schuler et al. (2023). In the machine learning literature, this is deemed "early stopping".

Other methods provide means of estimating the cumulative intensity Λ^x directly instead of splitting it up into K separate parameters using the event-specific cause-specific cumulative hazard functions. 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). Others choices include flexible parametric models/highly adaptive LASSO using piece-wise constant intensity models and the likelihood is based on Poisson regression.

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 5 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

Bang, H., & Robins, J. M. (2005). Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, *61*(4), 962–973. https://doi.org/10.1111/j.1541-0420.2005.00377.x

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., van der Laan, M. J., & Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures, counter-examples. *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, 255–294.

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.

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

Schuler, A., Li, Y., & Laan, M. van der. (2023,). *Lassoed Tree Boosting*. https://arxiv.org/abs/2205.10697

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

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

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.

11 Appendix

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

Theorem 7 (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

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

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 $i = 1, ..., d$, 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,

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

Letting

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

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

completes the proof of (ii).

□

11.2 Simulating the data

One classical causal inference perspective requires that we know how the data was generated up to unknown parameters (NPSEM) (Pearl, 2009). This approach has only to a small degree been discussed in 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 3. 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 4). 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 3, 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).

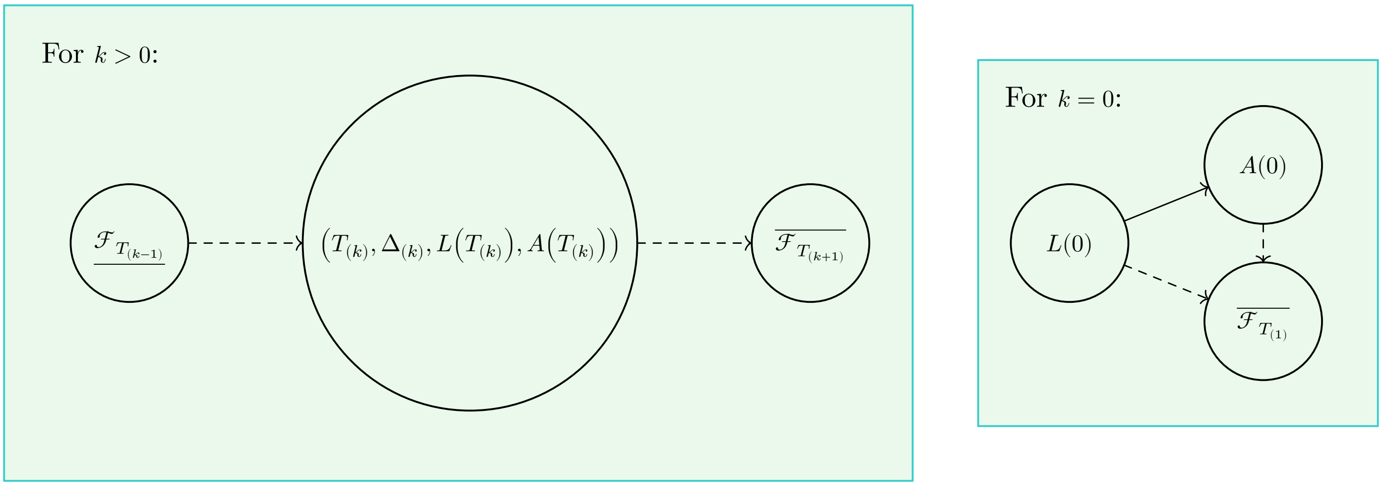

Figure 3: A DAG representing the relationships between the variables of O. The dashed lines indicate multiple edges from the dependencies in the past and into the future.

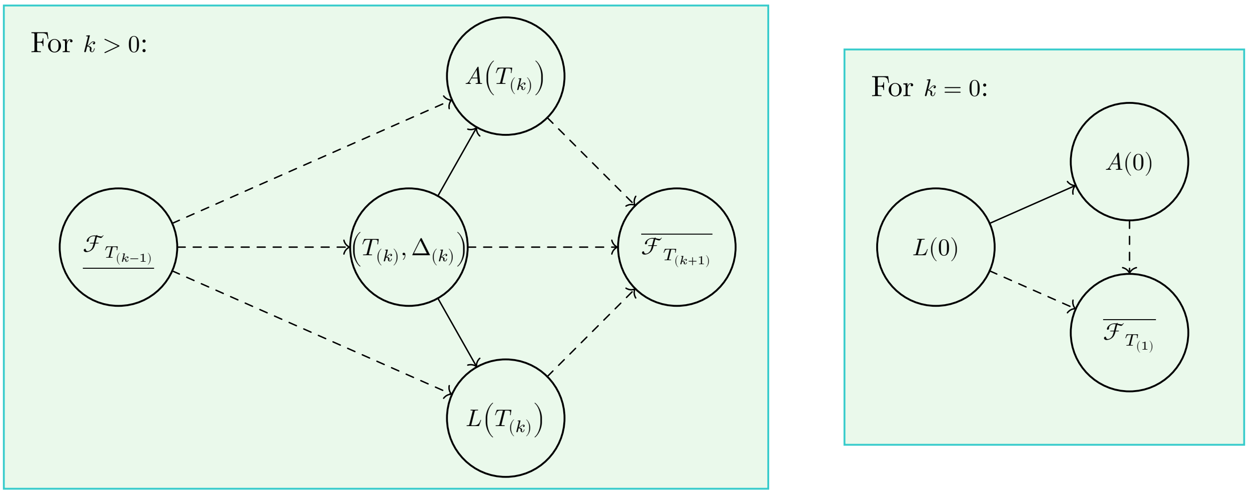

Figure 4: A DAG for simulating the data generating mechanism. The dashed lines indicate multiple edges from the dependencies in the past and into the future. Here $\underline{\mathcal{F}_{T_{(k)}}}$ is the history up to and including the k'th event and $\overline{\mathcal{F}_{T_{(k)}}}$ is the history after and including the k'th event.

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

Let

$$B_{k-1}(u) = \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(u)\right) \frac{1}{\tilde{S}^c(u) S(u)}$$

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)^{\mathbb{1}(\Delta_{(j)}=a)} \frac{\mathbb{1}\left(\Delta_{(k-1)} \in \{\ell, a\}, T_{(k-1)} \le \tau\right)}{\exp\left(-\sum_{1 \le 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)} \mathbb{1}\left(T_{(k)} \le \tau, \Delta_{(k)} = \ell, k < K\right) \Big(\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] \Big) \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^g_\tau(P)
\end{aligned}$$

We find immediately that

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

Now note that

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

Let us calculate the second integral

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

where the last line holds by the Duhamel equation (2.6.5) 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(\mathrm{d}u, \mathcal{F}_{T_{(k-1-1)}}\Big) \\
&= \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 \left(\int_{\mathcal{A}} \bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k-1)}\Big), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\Big) \pi^*_k\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \\
&+ \int_0^u S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^\ell_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1-1)}}\Big) \\
&\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(\mathrm{d}u, \mathcal{F}_{T_{(k-1-1)}}\Big) \\
&= \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(\mathrm{d}u, \mathcal{F}_{T_{(k-1-1)}}\Big) \Big[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 \left(\int_{\mathcal{A}} \bar{Q}^g_{k+1,\tau}\Big(L\Big(T_{(k-1)}\Big), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\Big) \pi^*_k\Big(s, \mathcal{F}_{T_{(k-1)}}\Big) \nu_A(\mathrm{d}a_k)\right) \\
&+ S\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \Lambda^\ell_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1-1)}}\Big) \\
&\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)\Big]
\end{aligned}
$$

Now note that

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

Setting this into the previous integral, we get

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

Thus, we find

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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