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

Johan Sebastian Ohlendorff[a,*], Anders Munch[a] and Thomas Alexander Gerds[a]

[a]*University of Copenhagen*
[a]*University of Copenhagen*
[a]*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 on a medical outcome such as the time to an event of interest.

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

[*]Corresponding author. E-mail address: johan.ohlendorff@sund.ku.dk

July 18, 2025

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. This is well known for inducing bias due to the introduction of time-varying confounding that is unaccounted for (Adams et al. (2020); Ferreira Guerra et al. (2020); Kant & Krijthe (2025); Ryalen et al. (2019); Sofrygin et al. (2019); Sun & Crawford (2023)).

In this paper, we consider a longitudinal continuous-time framework similar to that of Rytgaard et al. (2022) and Røysland (2011). Like Rytgaard et al. (2022), we identify the parameter of interest nonparametrically and focus on estimation and inference through the efficient influence function, yielding nonparametrically locally efficient estimators via a one-step procedure (Bickel et al. (1993); Tsiatis (2006); van der Vaart (1998)).

To this end, we propose an inverse probability of censoring iterative conditional expectation (ICE-IPCW) estimator, which, like the iterative regression of Rytgaard et al. (2022), iteratively updates nuisance parameters by regressing back through the history. Both methods extend the original discrete-time iterative regression method introduced by Bang & Robins (2005).

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. This then allows us to apply simple regression techniques for the nuisance parameters. 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 1 and 2. To the best of our knowledge, no general estimation procedure has yet been proposed for the components involved in the efficient influence function. Another advantage of using iterative regressions is that the resulting estimator will be less sensitive to/near practical positivity violations.

For electronic health records (EHRs), the number of registrations for each patient can be enormous. However, for finely discretized time grids in discrete time, it has been demonstrated that inverse probability of treatment weighting (IPW) estimators become increasingly biased and inefficient as the number of time points increases whereas iterative regression methods appear to be less sensitive to this issue (Adams et al. (2020)). Yet, many existing methods for estimating causal effects in continuous time apply inverse probability of treatment weighting (IPW) to identify the target parameter (see e.g., Røysland (2011); Røysland et al. (2024)).

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. We shall likewise employ a change of measure to define the target parameter. 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.

In Section 2, we introduce the setting and notation used throughout the paper. In Section 3, we present the estimand of interest and provide the iterative representation of the target parameter. In Section 4, we introduce right-censoring, discuss the implications for inference, and present the algorithm for estimation, as well as example usage. In Section 5, we use the Gateaux derivative to find the efficient influence function and present the debiased ICE-IPCW estimator. In Section 6 we present the results of a simulation study and in Section 7 we apply the method to electronic health records data from the Danish registers.

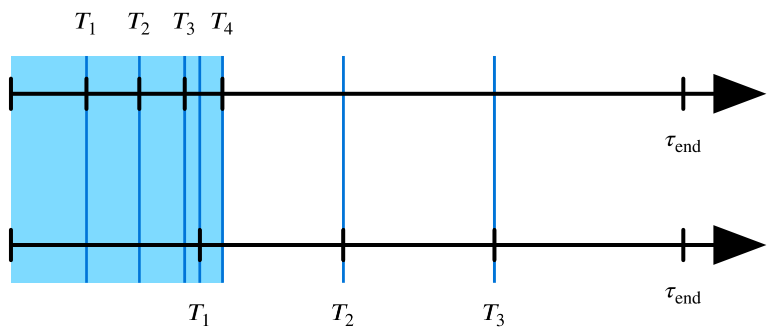

Figure 1: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let $t_1 < \ldots < t_m$ be all the event times in the sample. Let P^{G^*} denote the interventional probability measure. Then, given $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{t_r}\right]$,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

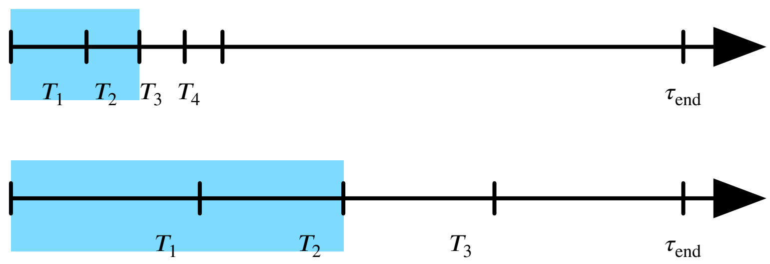

Figure 2: The figure illustrates the sequential regression approach proposed in this article. For each event number k in the sample, we regress back on the history $\mathcal{F}_{T_{(k-1)}}$. Let P^{G^*} denote the interventional probability measure. That is, given $\mathbb{E}_{P^{G^*}}\left[Y \mid \mathcal{F}_{T_{(k)}}\right]$, 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 $\Delta N^\ell(t) \neq 0$ or $\Delta N^a(t) \neq 0$. Note that we allow, for practical reasons, some of the covariate values to change at the same time as the treatment values. This can occur if registrations occur only on a daily level if, for example, a patient visits the doctor, gets a blood test, and receives a treatment all on the same day. This means that we can practically assume that $\Delta N^a \Delta N^\ell \equiv 0$. We *emphasize* the importance of the former 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 at time τ_{end} for the processes L and A satisfies $N^a(\tau_{\text{end}}) + N^\ell(\tau_{\text{end}}) \leq K - 1$ P -a.s. for some $K \geq 1$

We have a counting process representing the event of interest $(N^y(t))_{t\geq 0}$ and the competing event $(N^d(t))_{t\geq 0}$.

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. 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 $(\mathbb{R}_+ \times (\{a, \ell, y, d, c\} \times \{0,1\} \times \mathscr{L}))$. It follows that the filtration $(\mathscr{F}_t)_{t \geq 0}$ 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 $O = (T_{(K)}, \Delta_{(K)}, A(T_{(K-1)}), L(T_{(K-1)}), T_{(K-1)}, \Delta_{(K-1)}, \dots, A(0), L(0)) \sim P \in \mathscr{M}$ where $\mathscr{M}$ is the statistical model, i.e., a set of probability measures and obtain a sample $O = (O_1, \dots, O_n)$ 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 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 2, representing the data set in the form of O .

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

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

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 which 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 $\pi_k\left(t, L(T_{(k)}), \mathcal{F}_{T_{(k-1)}}\right)$ be the probability of being treated at the k ‘th event given $\Delta_{(k)} = a$, $T_{(k)} = t$, $L(T_{(k)})$, 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 $k = 1, \ldots, K-1$ 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,

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

$$\times \frac{\pi_0^*(L(0))^{A(0)}\left(1-\pi_0^*(L(0))\right)^{1-A(0)}}{\pi_0(L(0))^{A(0)}\left(1-\pi_0(L(0))\right)^{1-A(0)}}, \tag{1b}$$

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,

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

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 $k = 0, \ldots, K-1$, then our target parameter is given by the mean interventional cumulative incidence function at time τ,

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

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 $k = 1, \ldots, K-1$, and the methods we present can easily be extended to more complex treatment regimes and contrasts. Note that alternatively, we can interpret the target parameter $\Psi_\tau^{g,K}(P)$ as the counterfactual cumulative incidence function of the event of interest y at time τ, when the intervention enforces treatment as part of the $K-1$ first events. Henceforth, we will assume that (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

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

where $\prod_{s \in (0,t]}$ is the product integral over the interval $(0,t]$ (Gill & Johansen (1990)). We discuss more thoroughly the implications for inference of this representation, the algorithm for estimation and examples in 4 where we also deal with right-censoring.

Theorem 1: Let $H_k = \left(L(T_{(k)}), T_{(k)}, \Delta_{(k)}, A(T_{(k-1)}), L(T_{(k-1)}), T_{(k-1)}, \Delta_{(k-1)}, \dots, A(0), L(0)\right)$ be the history up to and including the k 'th event, but excluding the k 'th treatment values for $k > 0$. For $k = 0$, let $H_0 = L(0)$. Let $\bar{Q}^g_{K,\tau} : (a_k, h_k) \mapsto 0$ and recursively define for $k = K-1, \dots, 1$,

$$Z^a_{k+1,\tau}(u) = \mathbb{1}\{T_{(k+1)} \le u, T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell)\bar{Q}^g_{k+1,\tau}(\tau, A(T_{(k)}), H_{k+1}) \tag{3a}$$

$$+\mathbb{1}\{T_{(k+1)} \le u, T_{(k+1)} < \tau, \Delta_{(k+1)} = a)\bar{Q}^g_{k+1,\tau}(\tau, 1, H_{k+1}) + \mathbb{1}\{T_{(k+1)} \le u, \Delta_{(k+1)} = y\}, \tag{3b}$$

and

$$\bar{Q}^g_{k,\tau} : (u, a_k, h_k) \mapsto \mathbb{E}_P\left[Z^a_{k+1,\tau}(u) \mid A(T_{(k)}) = a_k, H_k = h_k\right], \tag{4}$$

for $u \le \tau$. Then,

$$\Psi^g_\tau(P) = \mathbb{E}_P\left[\bar{Q}^g_{0,\tau}(\tau, 1, L(0))\right]. \tag{5}$$

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

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

for $k = 0, \dots, K$ satisfies the desired property of (4). First, we find

$$\bar{Q}^g_{k,\tau} = \mathbb{E}_P\left[\sum_{j=k+1}^{K} W_{k,j}\mathbb{1}\{T_{(j)} \le \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k)}}\right] \tag{6a}$$

$$= \mathbb{E}_P\Bigg(\mathbb{1}\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\}W_{k,k+1} \tag{6b}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} \in \{a, \ell\}\} \tag{6c}$$

$$\times \mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\{T_{(j)} \le \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right]\Bigg) \tag{6d}$$

$$= \mathbb{E}_P\Bigg(\mathbb{1}\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\}W_{k,k+1} \tag{6e}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\} \tag{6f}$$

$$\times \mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\{T_{(j)} \le \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \tag{6g}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\} \tag{6h}$$

$$\times \mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\Big)\,\middle|\, \mathcal{F}_{T_{(k)}}\right] \tag{6i}$$

$$= \mathbb{E}_P\Bigg(\mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} \tag{6j}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\} \tag{6k}$$

$$\times \mathbb{E}_P\left[W_{k,k+1}\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \tag{6l}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\} \tag{6m}$$

$$\times \mathbb{E}_P\left[\mathbb{E}_P\left[\sum_{j=k+2}^{K} W_{k+1,j}\mathbb{1}\{T_{(j)} \leq \tau, \Delta_{(j)} = y\} \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \,\middle|\, \mathcal{F}_{T_{(k)}}\Bigg) \tag{6n}$$

$$= \mathbb{E}_P\Big(\mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} \tag{6o}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\} \tag{6p}$$

$$\times \mathbb{E}_P\left[W_{k,k+1}\bar{Q}^g_{k+1,\tau}\left(A(T_{(k+1)}), L(T_{(k)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k)}}\right) \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \tag{6q}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\} \tag{6r}$$

$$\times \mathbb{E}_P\left[\bar{Q}^g_{k+1,\tau}\left(A(T_{(k)}), L(T_{(k+1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k)}}\right) \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \mid \mathcal{F}_{T_{(k)}}\Big) \tag{6s}$$

$$= \mathbb{E}_P\Big(\mathbb{1}\{T_{(k+1)} \leq \tau, \Delta_{(k+1)} = y\} \tag{6t}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\} \tag{6u}$$

$$\times \mathbb{E}_P\left[W_{k,k+1}\bar{Q}^g_{k+1,\tau}\left(A(T_{(k+1)}), L(T_{(k)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k)}}\right) \mid T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \tag{6v}$$

$$+\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\} \tag{6w}$$

$$\times \bar{Q}^g_{k+1,\tau}\left(A(T_{(k)}), L(T_{(k+1)}), T_{(k)}, \Delta_{(k)}, \mathcal{F}_{T_{(k)}}\right) \mid \mathcal{F}_{T_{(k)}}\Big) \tag{6x}$$

by the law of iterated expectations and that

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

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

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

by the law of iterated expectations in the second step from which (4) follows. A similar calculation shows that $\Psi^g_\tau(P) = \mathbb{E}_P\left[\bar{Q}^g_{0,\tau}(1, L(0))\right]$ and so (5) follows. This shows the first statement.

□

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). (4) 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 (4) 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 $(A(T_{(k-1)}), L(T_{(k-1)}), T_{(k-1)}, \Delta_{(k-1)}, \ldots, A(0), L(0))$ to obtain an estimator of $\bar{Q}^g_{k-1,\tau}$. which involves iterative integration, as this method becomes computationally infeasible even for small values of K .

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 . Let N^c be the censoring process given by $N^c(t) = \mathbb{1}\{C \leq t\}$.

We will introduce the notation necessary to discuss the algorithm for the the ICE-IPCW estimator in 5 and later discuss the assumptions necessary for consistency of the ICE-IPCW estimator in 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 $(T_{(k)}, \Delta_{(k)}, A(T_{(k)}), L(T_{(k)}))$ be the event times and marks for the N^α process.

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

$$\bar{T}_{(k)} = C \wedge T_{(k)} \tag{7a}$$

$$\bar{\Delta}_{(k)} = \begin{cases} \Delta_{(k)} \text{ if } C > T_{(k)} \\ \text{c if } C \leq T_{(k)} \text{ and } \bar{\Delta}_{(k-1)} \neq c \\ \text{Ø otherwise} \end{cases} \tag{7b}$$

$$A(\bar{T}_k) = \begin{cases} A(T_{(k)}) \text{if } C > T_{(k)} \\ \text{Ø otherwise} \end{cases} \tag{7c}$$

$$L(\bar{T}_k) = \begin{cases} L(T_{(k)}) \text{ if } C > T_{(k)} \\ \text{Ø otherwise} \end{cases} \tag{7d}$$

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

$$\mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = \sigma(\bar{T}_{(k)}, \bar{\Delta}_{(k)}, A(\bar{T}_k), L(\bar{T}_k), \ldots, \bar{T}_{(1)}, \bar{\Delta}_{(1)}, A(\bar{T}_1), L(\bar{T}_1), A(0), L(0)), \tag{8}$$

defining the observed history up to and including the k 'th event. Thus $O = (\bar{T}_{(1)}, \bar{\Delta}_{(1)}, A(\bar{T}_1), L(\bar{T}_1), \dots, \bar{T}_{(K)}, \bar{\Delta}_{(K)}, A(\bar{T}_K), L(\bar{T}_K))$ is the observed data and a sample consists of $O = (O_1, \dots, O_n)$ for n independent and identically distributed observations with $O_i \sim P$. We will formally show (8) later.

Define $\tilde{\Lambda}_k^c\left(dt, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)$ 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 $\tilde{S}^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) = \prod_{s \in (T_{(k-1)}, t]}\left(1 - \tilde{\Lambda}_k^c\left(ds, \mathcal{F}^{\beta}_{\bar{T}_{(k-1)}}\right)\right)$. 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. Ideally, the model for iterative regressions 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 also be chosen such that the predictions are $[0, 1]$-valued. Histories are high-dimensional and should probably be reduced to some low-dimensional representation if many events occur in the sample.

Recall that Theorem 1 states that the target parameter may be identified via iterative regressions. We suppose that we are given an estimator of the censoring compensator $\hat{\Lambda}^c$. In particular, for $\bar{Q}^g_{k,\tau}, k = 0, \dots, K-1$, we start the algorithm at $k = K-1$ by calculating $\hat{S}^c_k(\bar{T}_{(k),i} - \mid \bar{A}_{(k-1),i}, H_{k-1,i}) = \prod_{s \in (\bar{T}_{(k-1),i}, \bar{T}_{(k),i}-)}(1 - \hat{\Lambda}^c_i(s))$. Given an estimator of $\bar{Q}^g_{k+1,\tau}$ denoted by $\hat{\nu}_{(k+1),\tau}$, we then calculate the pseudo-outcome $\hat{Z}^a_{k,i}$ as follows

- If $\bar{\Delta}_{(k),i} = y$, we calculate $\hat{Z}^a_{k,i} = \frac{1}{\hat{S}^c_k(\bar{T}_{(k),i} - \mid \bar{A}_{(k-1),i}, H_{k-1,i})} \mathbb{1}\{\bar{T}_{(k),i} \le \tau\}$.
- If $\bar{\Delta}_{(k),i} = a$, evaluate $\hat{\nu}_{(k+1),\tau}(1, H_{k,i})$ and calculate $\hat{Z}^a_{k,i} = \frac{1}{\hat{S}^c_k(\bar{T}_{(k),i} - \mid \bar{A}_{(k-1),i}, H_{k-1,i})} \mathbb{1}\{\bar{T}_{(k),i} < \tau\} \hat{\nu}_{(k+1),\tau}(1, \bar{H}_{k,i})$.
- If $\bar{\Delta}_{(k),i} = \ell$, evaluate $\hat{\nu}_{(k+1),\tau}(\bar{A}_{(k-1),i}, H_{k,i})$, and calculate $\hat{Z}^a_{k,i} = \frac{1}{\hat{S}^c_k(\bar{T}_{(k),i} - \mid \bar{A}_{(k-1),i}, H_{k-1,i})} \mathbb{1}\{\bar{T}_{(k),i} < \tau\} \hat{\nu}_{(k+1),\tau}(\bar{A}_{(k-1),i}, \bar{H}_{k,i})$.

Then regress $\hat{Z}^a_{k,i}$ on $(\bar{A}_{(k-1),i}, H_{k-1,i})$ for the observations with with $\bar{T}_{(k-1),i} < \tau$ and $\bar{\Delta}_{(k-1),i} \in \{a, \ell\}$ to obtain a prediction function $\hat{\nu}_k$. Finally, we compute $\hat{\Psi}_n = \frac{1}{n}\sum_{i=1}^n \hat{\nu}_0(1, L_i(0))$.

We mention how one may obtain an estimator of the censoring compensator, but this is a wider topic that we will not concern ourselves with here. We provide a model for the censoring that can provide estimates of the hazard measure $\frac{1}{P(\bar{T}_{(k)} \ge t \mid A(\bar{T}_{k-1}), \bar{H}_{k-1})} P(\bar{T}_{(k)} \in dt, \bar{\Delta}_{(k)} = c \mid A(\bar{T}_{k-1}), \bar{H}_{k-1})$, which is always estimable from observed data. Then, regress $\bar{E}_{(k),i} = \bar{T}_{(k),i} - \bar{T}_{(k-1),i}$, known as the k 'th *interarrival* time, with the censoring as the cause of interest on $(\bar{A}_{(k-1),i}, \bar{H}_{k-1,i})$ among the patients who are still at risk after $k-1$ events, that is for i with $\bar{\Delta}_{k-1,i} \in \{a, \ell\}$ if $k > 1$ and otherwise all $i = 1, \dots, n$. This gives an estimator of the cause-specific cumulative hazard function $\hat{\Lambda}^c_k$. This then gives an estimator of the compensator as follows

$$\hat{\Lambda}_i^c(t) = \sum_{k=1}^{K} \mathbb{1}\{\bar{T}_{(k-1),i} < t \leq \bar{T}_{(k),i}\} \hat{\Lambda}_k^c\left(t - \bar{T}_{(k-1),i} \mid \bar{A}_{(k-1),i}, \bar{H}_{(k-1),i}\right)$$

Note that in the algorithm, we only forced the current value of $A(\bar{T}_k)$ to 1, instead of replacing all prior treatment values with 1. The latter is certainly closer to the iterative conditional expectation estimator as proposed by Bang & Robins (2005), but is mathematically equivalent. This follows from standard properties of the conditional expectation (see e.g., Theorem A3.13 of Last & Brandt (1995)).

6. Consistency of the ICE-IPCW Estimator

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

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

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 $\mathcal{F}_t^{\tilde{\beta}} = \sigma(\beta(s \wedge C \wedge T^e) \mid s \leq t)$. The observed data is then given by (7). Importantly, we have in fact*.

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

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 $\mathcal{F}_{T_{(k)}} = \sigma\left(T_{(k)}, \Delta_{(k)}, A\left(T_{(k)}\right), L\left(T_{(k)}\right), \dots, T_{(1)}, \Delta_{(1)}, A\left(T_{(1)}\right), L\left(T_{(1)}\right), A(0), L(0)\right)$. 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

$$P\left(T_{(k)} \in [t, t + \mathrm{d}t), \Delta_{(k)} = x, A\left(T_{(k)}\right) = m, L\left(T_{(k)}\right) = l \mid \mathcal{F}_{T_{(k-1)}}, T_{(k)} \geq t\right) \tag{9a}$$

$$= P\left(\bar{T}_{(k)} \in [t, t + \mathrm{d}t), \bar{\Delta}_{(k)} = x, A\left(\bar{T}_k\right) = m, L\left(\bar{T}_k\right) = l \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}, \bar{T}_{(k)} \geq t\right), \qquad x \neq c. \tag{9b}$$

*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.

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 (9) (e.g., Gill et al. (1997)).

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 $\mathcal{F}_t$. If

1. $\Delta\tilde{\Lambda}_k^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) + \sum_x \Delta\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.} \Rightarrow \Delta\tilde{\Lambda}_{k+1}^c\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.} \vee$ $\sum_x \Delta\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right) = 1 \quad P-\text{a.s.}$.
2. $\tilde{S}^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) > \eta$ for all $t \in (0, \tau]$ and $n \in \{1, \dots, K\}$ P-a.s. for some $\eta > 0$.

Let

$$
\begin{aligned}
\bar{Z}_k^a(u) = \frac{1}{\tilde{S}^c\left(\bar{T}_{(k)} - \mid A\left(\bar{T}_{k-1}\right), \bar{H}_{k-1}\right)} \Big(& \mathbb{1}\left\{\bar{T}_{(k)} \le u, \bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = a\right\} \bar{Q}_{k,\tau}^g\left(A\left(\bar{T}_k\right), \bar{H}_k\right) \\
& + \mathbb{1}\left\{\bar{T}_{(k)} \le u, \bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} = \ell\right\} \bar{Q}_{k,\tau}^g\left(A\left(\bar{T}_k\right), \bar{H}_k\right) \\
& + \mathbb{1}\left\{\bar{T}_{(k)} \le u, \bar{\Delta}_{(k)} = y\right\} \Big)
\end{aligned}
$$

Then,

$$
\bar{Q}_{k,\tau}^g\left(u, a_k, h_k\right) = \mathbb{E}_P\left[\bar{Z}_{k+1}^a(u) \mid A\left(\bar{T}_k\right) = a_k, \bar{H}_k = h_k\right]
$$

Theorem 3: Assume that the compensator Λ^α of N^α with respect to the filtration $\mathscr{F}_t^\beta$ is also the compensator with respect to the filtration $\mathscr{F}_t$. Then for uncensored histories, we have

$$\mathbb{1}\{\bar{\Delta}_{(n-1)} \neq c\} P\Big(\big(\bar{T}_n, \bar{\Delta}_n, A(\bar{T}_n), L(\bar{T}_n)\big) \in d(t, m, a, l) | \, \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}} \Big) \tag{10a}$$

$$= \mathbb{1}\{\bar{T}_{n-1} < t, \bar{\Delta}_{(n-1)} \neq c\} \Bigg(\tilde{S}\Big(t - | \, \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}} \Big) \sum_{x=a,\ell,d,y} \delta_x(\mathrm{d}m) \psi_{n,x}(t, d(a,l)) \Lambda_n^x\Big(\mathrm{d}t, \mathscr{F}_{T_{(n-1)}}\Big) \tag{10b}$$

$$+ \delta_{(c, A(T_{(n-1)}), L(T_{(n-1)}))}(\mathrm{d}(m,a,l)) \tilde{\Lambda}_n^c\Big(\mathrm{d}t, \mathscr{F}^{\beta}_{\bar{T}_{(n-1)}}\Big) \Bigg) \tag{10c}$$

where

$$\psi_{n,x}\Big(t, \mathscr{F}_{T_{(n-1)}}, \mathrm{d}(m,a,l)\Big) = \mathbb{1}\{x = a\} \Big(\delta_1(\mathrm{d}a) \pi_n\Big(t, L\big(T_{(n)}\big), \mathscr{F}_{T_{(n-1)}}\Big) + \delta_0(\mathrm{d}a) \Big(1 - \pi_n\Big(t, L\big(T_{(n)}\big), \mathscr{F}_{T_{(n-1)}}\Big)\Big)\Big) \delta_{L(T_{(n-1)})}(\mathrm{d}l) \tag{11a}$$

$$+ \mathbb{1}\{x = \ell\} \mu_n\Big(dl, t, \mathscr{F}_{T_{(n-1)}}\Big) \delta_{A(T_{(n-1)})}(\mathrm{d}a) \tag{11b}$$

$$+ \mathbb{1}\{x \in \{y, d\}\} \delta_{A(T_{(n-1)})}(\mathrm{d}a) \delta_{L(T_{(n-1)})}(\mathrm{d}l). \tag{11c}$$

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

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 $\mathscr{F}_t^\beta$, can be given by Theorem 4.2.2 (ii) of Last & Brandt (1995),

$$\Lambda^\alpha(\,\mathrm{d}(t, m, a, l)) = K'\big((L(0), A(0)), N^\alpha, t, \mathrm{d}(m, a, l)\big) V'\big((A(0), L(0)), N^\alpha, \mathrm{d}t\big) \tag{12}$$

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

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

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

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

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

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

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

and similarly, we see that

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

Let $T^*_{(k)}$ denote the ordered event times of the process N^β. With $S \coloneqq T^e \wedge C \wedge T_{(k)}$, 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

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

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

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

and we are done. From this, we get (10). Applying this to the right hand side of (13) shows that it is equal to (4). $\square$

Further supposee that:

1. $\mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\} \tilde{S}\left(t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) = \mathbb{1}\left\{\bar{\Delta}_{(n-1)} \neq c\right\} \tilde{S}^c\left(t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(n-1)}}\right) S\left(t \mid \mathscr{F}_{T_{(n-1)}}\right)$.

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

Note that the theorem can now be used to show the consistency of the ICE-IPCW estimator. What we claim is that,

$$\bar{Q}^g_{k-1,\tau}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) = \mathbb{E}_P\left[\bar{Z}^a_{k-1}\left(A(T_{(k-1)}), L(T_{(k-1)}), T_{(k-1)}, \Delta_{(k-1)}, \mathcal{F}_{T_{(k-1)}}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right] \tag{13}$$

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

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

We proceed by backwards induction.

$$\begin{aligned}
&\bar{Q}^g_{k,\tau}\left(\tau \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}\right)\\
&= \mathbb{E}_P\Big[\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\}\bar{Q}^g_{k+1,\tau}\left(1, L(T_{(k)}), \bar{T}_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\\
&\qquad + \mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\}\bar{Q}^g_{k+1,\tau}\left(A(T_{(k)}), L(T_{(k+1)}), T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\\
&\qquad + \mathbb{1}\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\} \mid \mathcal{F}_{T_{(k)}}\Big]\\
&\overset{*}{=} \mathbb{E}_P\Big[\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\}\bar{Q}^g_{k+1,\tau}\left(1, L(T_{(k)}), \bar{T}_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\\
&\qquad + \mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\}\bar{Q}^g_{k+1,\tau}\left(A(T_{(k)}), L(T_{(k+1)}), T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\\
&\qquad + \mathbb{1}\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\} \mid \mathcal{F}_{T_{(k)}}, \mathscr{C}_k\Big]\\
&= \mathbb{E}_P\Big[\mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = a\}\bar{Q}^g_{k+1,\tau}\left(1, L(T_{(k)}), \bar{T}_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\\
&\qquad + \mathbb{1}\{T_{(k+1)} < \tau, \Delta_{(k+1)} = \ell\}\bar{Q}^g_{k+1,\tau}\left(A(T_{(k)}), L(T_{(k+1)}), T_{(k+1)}, \Delta_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\\
&\qquad + \mathbb{1}\{T_{(k+1)} \le \tau, \Delta_{(k+1)} = y\} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k)}}, \mathscr{C}_k\Big],\\
&= \bar{Q}^g_{k-1,\tau}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) = \mathbb{E}_P\left[\bar{Z}^a_{k-1}\left(A(T_{(k-1)}), L(T_{(k-1)}), T_{(k-1)}, \Delta_{(k-1)}, \mathcal{F}_{T_{(k-1)}}\right) \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right]
\end{aligned}$$

where $\mathscr{C}_k = (\nexists k : \Delta_{(k)} = c)$. In $(*)$, we use local independence since the corresponding event lies in the sigma-algebra $\mathcal{F}^{\beta}_{T_{(k)}}$, and so we can use the same mark and conditional event distribution as in the first line. In the last line, we apply (10) since the conditioning set is a part of $\mathcal{F}^{\beta}_{T_{(k)}}$.

Note that (10) also ensures that all hazards (other than censoring) and mark probabilities are identifiable from censored data if we can show that the censoring survival factorizes. We provide two criteria for this. Theorem 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 4: Assume independent censoring as in Theorem 2. Then the left limit of the survival function factorizes on $(0,\tau]$, i.e.,

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

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

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

Proof: First, we argue that for every $t \in (0,\tau]$ with $\tilde{S}\left(t - \mid \mathscr{F}^{\tilde{\beta}}_{\tilde{T}_{(k-1)}}\right) > 0$ (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

$$\begin{aligned} 0 = \sum_{k=1}^K \mathbb{1}\{T_{(k-1)} \wedge C < t \leq T_{(k)} \wedge C\} \Bigg(\int_{(T_{(k-1)} \wedge C, t]} \Delta\tilde{\Lambda}_k^c\left(s, \mathscr{F}^{\beta}_{\tilde{T}_{(k-1)}}\right) \left(\sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s, \mathscr{F}_{T_{(k-1)}}\right)\right) \\ + \sum_{j=1}^{k-1} \int_{(T_{(j-1)} \wedge C, T_{(j)} \wedge C]} \Delta\tilde{\Lambda}_c\left(s \mid \mathscr{F}^{\tilde{\beta}}_{\tilde{T}_{(j-1)}}\right) \left(\sum_{x=a,\ell,y,d} \Lambda_k^x\left(\mathrm{d}s, \mathscr{F}_{T_{(k-1)}}\right)\right)\Bigg) \end{aligned}$$

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

$$\mathbb{1}\{T_{(k-1)} \wedge C < t\} \tilde{S}\left(t - \mid \mathscr{F}^{\tilde{\beta}}_{\tilde{T}_{(k-1)}}\right) \sum_{\tilde{T}_{(k)} < s \leq t} \Delta\tilde{\Lambda}_k^c\left(s, \mathscr{F}^{\beta}_{\tilde{T}_{(k-1)}}\right) \left(\sum_{x=a,\ell,y,d} \Delta\Lambda_k^x\left(s, \mathscr{F}_{T_{(k-1)}}\right)\right) = 0, \tag{15}$$

where we also use that $\Delta\tilde{\Lambda}_k^c\left(s, \mathscr{F}^{\beta}_{\tilde{T}_{(k-1)}}\right) \neq 0$ 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 (15). It follows that for every t with $\tilde{S}\left(t - \mid \mathscr{F}^{\tilde{\beta}}_{\tilde{T}_k}\right) > 0$,

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

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

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

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

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

from which the result follows. (**NOTE**: We already the seen implication of the first part to the second part since $\Delta\gamma(u)+\Delta\zeta(u)\le 1$; 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^g_τ 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 debias our initial ICE-IPCW estimator either through double/debiased machine learning (Chernozhukov et al. (2018)), obtaining a one-step estimator that for given estimators of the nuisance parameters appearing in the efficient influence function,s

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

We derive the efficient influence function using the iterative representation given in (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 (17) is the efficient influence function, but rather a heuristic argument. To proceed, we introduce additional notation and define

$$\bar{Q}^g_{k,\tau}\left(u \mid \mathcal{F}_{T_{(k)}}\right) = p_{ka}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k\ell}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{ky}\left(u \mid \mathcal{F}_{T_{(k-1)}}\right), u < \tau. \tag{16}$$

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 $P \in \mathcal{M}$, $\tilde{\Lambda}_k^c\left(t \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}; P\right)$ 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 $\tilde{\Lambda}_k^c\left(\tau_{\text{end}} \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}; P\right) \leq C$ for all $k = 1, \dots, K$ and every $P \in \mathcal{M}$. The efficient influence function is then given by

$$\varphi_\tau^*(P) = \frac{\mathbb{1}\{A(0) = 1\}}{\pi_0(L(0))} \sum_{k=1}^K \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{A(\bar{T}_j) = 1\}}{\pi_j\left(\bar{T}_{(j)}, L(T_{(j)}), \mathcal{F}_{T_{(j-1)}}\right)} \right)^{\mathbb{1}\{\bar{\Delta}_{(j)} = a\}} \frac{1}{\prod_{j=1}^{k-1} \tilde{S}^c\left(\bar{T}_{(j)} - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(j-1)}}\right)} \tag{17a}$$

$$\times \mathbb{1}\{\bar{\Delta}_{(k-1)} \in \{\ell, a\}, \bar{T}_{(k-1)} < \tau\} \Bigg(\left(\bar{Z}^a_{k,\tau} - \bar{Q}^g_{k-1,\tau} \right) \tag{17b}$$

$$+ \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \left(\bar{Q}^g_{k-1,\tau}\left(\tau \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) - \bar{Q}^g_{k-1,\tau}\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) \right) \frac{1}{\tilde{S}^c\left(u \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) S\left(u - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right)} \widetilde{M}^c(du) \Bigg) \tag{17c}$$

$$+ \bar{Q}^g_{0,\tau} - \Psi^g_\tau(P), \tag{17d}$$

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

Proof: Define

$$\bar{Z}^a_{k,\tau}(P \mid s, t_k, d_k, l_k, a_k, f_{k-1}) = \frac{\mathbb{1}\{t_k \leq s, t_k < \tau, d_k = \ell\}}{\tilde{S}^c_P\left(t_k - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}) \bar{Q}^g_{k,\tau}(P \mid a_{k-1}, l_k, t_k, d_k, f_{k-1}) \tag{18a}$$

$$+ \frac{\mathbb{1}\{t_k \leq s, t_k < \tau, d_k = a\}}{\tilde{S}^c_P\left(t_k - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \bar{Q}^g_{k,\tau}(P \mid 1, l_{k-1}, t_k, d_k, f_{k-1}) \tag{18b}$$

$$+ \frac{\mathbb{1}\{t_k \leq s, d_k = y\}}{\tilde{S}^c_P\left(t_k - \mid \mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}), s \leq \tau \tag{18c}$$

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 $\Psi_\tau^g(P_\varepsilon)$ 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)}}}\left(f_{k-1}\right)}{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)} \\
&\quad - \frac{\delta_{\mathcal{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}\left(f_{k-1}\right)}{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)}}}\left(f_{k-1}\right)}{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)}}}\left(f_{k-1}\right)}{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

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

$$= -\frac{1}{1-\Delta\tilde{\Lambda}_k^c\left(t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)} \prod_{u\in(s,t]}\left(1-\tilde{\Lambda}_k^c\left(\mathrm{d}t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right)\int_{(s,t)} \frac{1}{1-\Delta\tilde{\Lambda}_k^c\left(u \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\tilde{\Lambda}_k^c\left(\mathrm{d}u \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)$$
$$-\frac{1}{1-\Delta\tilde{\Lambda}_k^c\left(t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)} \prod_{u\in(s,t]}\left(1-\tilde{\Lambda}_k^c\left(\mathrm{d}t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right)\int_{\{t\}} \frac{1}{1-\Delta\tilde{\Lambda}_k^c\left(u \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\tilde{\Lambda}_k^c\left(\mathrm{d}u \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)$$
$$+\prod_{u\in(s,t]}\left(1-\tilde{\Lambda}_k^c\left(\mathrm{d}t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right)\frac{1}{\left(1-\Delta\tilde{\Lambda}_k^c\left(t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)\right)^2}\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\Delta\tilde{\Lambda}_{k,\varepsilon}^c\left(t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k)}} = f_{k-1}\right)$$
$$= -\prod_{u\in(s,t)}\left(1-\tilde{\Lambda}_k^c\left(\mathrm{d}t \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\right)\int_{(s,t)} \frac{1}{1-\Delta\tilde{\Lambda}_k^c\left(u \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\tilde{\Lambda}_k^c\left(\mathrm{d}u \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right).$$

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

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

Now, we are ready to recursively calculate the derivative

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

which yields,

$$\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\bar{Q}^a_{k-1,\tau}\left(P_\varepsilon \mid a_{k-1}, l_{k-1}, t_{t-1}, d_{k-1}, f_{k-2}\right)$$
$$= \frac{\delta_{\mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}}(f_{k-1})}{P\left(\mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}\Bigg(\bar{Z}^a_{k,\tau} - \bar{Q}^g_{k-1,\tau}\left(\tau, \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}}\right) +$$
$$+\int_{\bar{T}_{(k-1)}}^{\tau}\bar{Z}^a_{k,\tau}\left(\tau, t_k, d_k, l_k, a_k, f_{k-1}\right)\int_{(\bar{T}_{(k-1)},t_k)}\frac{1}{1-\Delta\tilde{\Lambda}_k^c\left(s \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}\frac{1}{\tilde{S}\left(s- \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)}\left(\tilde{N}^c(\mathrm{d}s) - \tilde{\Lambda}^c(\mathrm{d}s)\right)$$

$$P_{(\bar{T}_{(k)},\bar{\Delta}_{(k)},L(\bar{T}_k),A(\bar{T}_k))}\left(\mathrm{d}\left(t_k,d_k,l_k,a_k\right) \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)\Bigg)$$

$$+\int_{\bar{T}_{(k-1)}}^{\tau} \frac{\mathbb{1}\{t_k < \tau, d_k \in \{a, \ell\}\}}{\tilde{S}^c\left(t_k - \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\frac{\mathbb{1}\{a_k = 1\}}{\pi_k\left(t_k, L(T_{(k)}), \mathscr{F}_{T_{(k-1)}}\right)}\right)^{\mathbb{1}\{d_k=a\}} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \bar{Q}^g_{k,\tau}(P_\varepsilon \mid a_k, l_k, t_k, d_k, f_{k-1})$$
$$P_{(\bar{T}_{(k)},\bar{\Delta}_{(k)},L(\bar{T}_k),A(\bar{T}_k))}\left(\mathrm{d}\left(t_k,d_k,l_k,a_k\right) \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)$$

Now note for the second term, we can write

$$\int_{\bar{T}_{(k-1)}}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{(\bar{T}_{(k-1)},t_k)} \frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(s \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{1}{\tilde{S}\left(s - \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(\mathrm{d}s) - \tilde{\Lambda}^c(\mathrm{d}s)\right)$$
$$P_{(\bar{T}_{(k)},\bar{\Delta}_{(k)},L(\bar{T}_k),A(\bar{T}_k))}\left(\mathrm{d}\left(t_k,d_k,l_k,a_k\right) \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right))$$
$$= \int_{\bar{T}_{(k-1)}}^{\tau} \int_{s}^{\tau} \bar{Z}^a_{k,\tau}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) P_{(\bar{T}_{(k)},\bar{\Delta}_{(k)},L(\bar{T}_k),A(\bar{T}_k))}\left(\mathrm{d}\left(t_k,d_k,l_k,a_k\right) \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)$$
$$\frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(s \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{1}{\tilde{S}\left(s - \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(\mathrm{d}s) - \tilde{\Lambda}^c(\mathrm{d}s)\right)$$
$$= \int_{\bar{T}_{(k-1)}}^{\tau} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right)$$
$$\frac{1}{1 - \Delta\tilde{\Lambda}^c_k\left(s \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \frac{1}{\tilde{S}\left(s - \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} \left(\widetilde{N}^c(\mathrm{d}s) - \tilde{\Lambda}^c(\mathrm{d}s)\right)$$
$$= \int_{\bar{T}_{(k-1)}}^{\tau} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(s)\right)$$
$$\frac{1}{\tilde{S}^c\left(s \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right) S\left(s - \mid \mathscr{F}^{\tilde{\beta}}_{\bar{T}_{(k-1)}} = f_{k-1}\right)} (\widetilde{N}^c(\mathrm{d}s) - \tilde{\Lambda}^c(\mathrm{d}s)$$

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. □

Theorem 6 (Adaptive selection of K): Let $K_{nc} = \max_{v: \sum_{i=1}^n \mathbb{1}\{N_{\tau i}(\tau) \geq v\} > c}$ denote the maximum number of events with at least c people at risk. Suppose that we have the decomposition of the estimator $\hat{\Psi}_n$, such that

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

Suppose that there is a number $K_{\lim} \in \mathbb{N}$, such that $P\left(N_\tau = K_{\lim}\right) > 0$, but $P\left(N_\tau > K_{\lim}\right) = 0$. Then, the estimator $\hat{\Psi}_n$ satisfies

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

Proof: We find the following decomposition,

$$\begin{aligned}
\hat{\Psi}_n - \Psi_\tau^{g,K_{\lim}}(P) &= \hat{\Psi}_n - \Psi_\tau^{g,K_{nc}}(P) + \Psi_\tau^{g,K_{nc}}(P) - \Psi_\tau^{g,K_{\lim}}(P) \\
&= \left(\mathbb{P}_n - P\right)\varphi_\tau^{*,K_{nc}}(\cdot; P) + \Psi_\tau^{g,K_{nc}}(P) - \Psi_\tau^{g,K_{\lim}}(P) + o_P\left(n^{-\frac{1}{2}}\right) \\
&= \left(\mathbb{P}_n - P\right)\varphi_\tau^{*,K_{\lim}}(\cdot; P) + \left(\mathbb{P}_n - P\right)\left(\varphi_\tau^{*,K_{nc}}(\cdot; P) - \varphi_\tau^{*,K_{\lim}}(\cdot; P)\right) + \Psi_\tau^{g,K_{nc}}(P) - \Psi_\tau^{g,K_{\lim}}(P) + o_P\left(n^{-\frac{1}{2}}\right)
\end{aligned}$$

We will have shown the result if

1. $\Psi_\tau^{g,K_{nc}}(P) - \Psi_\tau^{g,K_{\lim}}(P) = o_P\left(n^{-\frac{1}{2}}\right)$.
2. $\left(\mathbb{P}_n - P\right)\left(\varphi_\tau^{*,K_{nc}}(\cdot; P) - \varphi_\tau^{*,K_{\lim}}(\cdot; P)\right) = o_P\left(n^{-\frac{1}{2}}\right)$ and

Assume that we have shown that $P\left(K_{nc} \neq K_{\lim}\right) \to 0$ as $n \to \infty$. We have that

$$\sqrt{n}\left(\Psi_\tau^{g,K_{nc}}(P) - \Psi_\tau^{g,K_{\lim}}(P)\right) = \sqrt{n}\mathbb{1}\left\{K_{nc} \neq K_{\lim}\right\}\left(\Psi_\tau^{g,K_{nc}}(P) - \Psi_\tau^{g,K_{\lim}}(P)\right) := E_n$$

and

$$P\left(|E_n| > \varepsilon\right) \leq P\left(K_{nc} \neq K_{\lim}\right) \to 0,$$

as $n \to \infty$. A similar conclusion holds for 2. Returning to showing that $P\left(K_{nc} \neq K_{\lim}\right) \to 0$, first define that $K_n = \max_i N_{\tau i}$. Then,

$$P\left(K_n < K_{\lim}\right) = P\left(N_\tau < K_{\lim}\right)^n \to 0$$

Then we can certainly write that

$$K_{nc} - K_{\lim} = K_{nc} - K_n + K_n - K_{\lim},$$

so we have

$$P\big(K_{nc} \neq K_n\big) = P\left(\cup_{v=1}^{K_{\lim}} \left(\sum_{i=1}^{n} \mathbb{1}\big\{N_{\tau i} \geq v\big\} \leq c\right)\right) \leq \sum_{v=1}^{K_{\lim}} P\left(\sum_{i=1}^{n} \mathbb{1}\big\{N_{\tau i} \geq v\big\} \leq c\right) \to 0$$

as $n \to \infty$. Here, we use that

$$\sum_{i=1}^{n} \mathbb{1}\big\{N_{\tau i} \geq v\big\}$$

converges almost surely to ∞ . Too see this, note that $\sum_{i=1}^{n} \mathbb{1}\big\{N_{\tau i} \geq v\big\}$ is almost surely monotone in n , and

$$\sum_{i=1}^{n} P\big(N_{\tau i} \geq v\big) = nP\big(N_{\tau} \geq v\big) \to \infty$$

From this and Kolmogorov's three series theorem, we conclude that

$$\sum_{i=1}^{n} \mathbb{1}\big\{N_{\tau i} \geq v\big\} \to \infty$$

almost surely as $n \to \infty$ and that

$$\sum_{i=1}^{n} \mathbb{1}\big\{N_{\tau i} \geq v\big\} \leq c$$

has probability tending to zero as $n \to \infty$. □

8. Real data application

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

NOTE: It is natural to assume that the doctor may take decide the treatment at the same time as some of the time-varying covariates are measured. Therefore, we can actually redefine $\Delta_k = a$ to be an event for which a treatment decision is possible and for which some of the covariates may change. For instance, we may assume that the doctor makes the treatment decisions at HbA1c measurement times.

Should we apply the methods to trial data? In the LEADER trial, the times at which the dosages change may be quite irregular.

Should we apply the methods to electronic health records data? We do not directly observe when the patient is treated, and can only monitor them through their history of purchases. In the case, we could define the visitation times as purchase dates and use information about dosages to calculate the times at which they would have no more medicine available. One problem with this approach is that a person that is not treated anymore is unable to not be treated at the next visitation time. For an emulated target trial in Diabetes research, a study on the effect of DPP4 may be the easiest, as that type of medication only comes in one recommended dosage.

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 8.

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.

Other methods provide means of estimating the cumulative intensity Λ^x directly instead of splitting it up into κ 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 piecewise 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. 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)).

References

Adams, R., Saria, S., & Rosenblum, M. (2020). The impact of time series length and discretization on longitudinal causal estimation methods. *Arxiv Preprint Arxiv:2011.15099*.

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

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., & Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models* (Vol. 4). Johns Hopkins University Press Baltimore.

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

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

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.

Kant, W. M., & Krijthe, J. H. (2025). Irregular measurement times in estimating time-varying treatment effects: Categorizing biases and comparing adjustment methods. *Arxiv Preprint Arxiv:2501.11449*.

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

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

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

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

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

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

Sofrygin, O., Zhu, Z., Schmittdiel, J. A., Adams, A. S., Grant, R. W., van der Laan, M. J., & Neugebauer, R. (2019). Targeted learning with daily EHR data. *Statistics in Medicine*, *38*(16), 3073–3090.

Sun, J., & Crawford, F. W. (2023). The role of discretization scales in causal inference with continuous-time treatment. *Arxiv Preprint Arxiv:2306.08840*.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. Springer.

van der Laan, M. J., & Gruber, S. (2010). Collaborative double robust targeted maximum likelihood estimation. *The International Journal of Biostatistics*, *6*(1).

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

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

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 $\Delta N_j(T_{(k)}) \neq 0$ and let $\mathbb{F}_k = \mathcal{W} \times (\mathbb{R}_+ \times \{1, \ldots, d\} \times \mathcal{X})^k$.

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

(i). We have $\mathscr{F}_{T_{(k)}} = \sigma\big(T_{(k)}, \Delta_{(k)}, X(T_{(k)})\big) \vee \mathscr{F}_{T_{(k-1)}}$ and $\mathscr{F}_0 = \sigma(W)$. Furthermore, $\mathscr{F}_t^{\bar{N}} = \sigma\big(\bar{N}((0,s],\cdot) \mid s \le t\big) \vee \sigma(W) = \mathscr{F}_t$, where

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

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,

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

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 $X_i(T_{(k)})$ given $\mathscr{F}_{T_{(k-1)}}$, $T_{(k)}$ 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 $\mathscr{F}_t^N = \sigma(N((0,s],\cdot) \mid s \le t) \vee \sigma(W) = \mathscr{F}_t$ where

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

Thus $\mathscr{F}_{T_{(k)}} = \sigma\Big(T_{(k)}, \tilde{X}(T_{(k)})\Big) \vee \mathscr{F}_{T_{(k-1)}}$ and $\mathscr{F}_0 = \sigma(W)$ 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 $N^i(T_{(k)})$ for $i = 1, \ldots, d$, implying that $\bar{N}$ generates the same filtration as N, i.e., $\mathscr{F}_t^N = \mathscr{F}_t^{\bar{N}}$ for all $t \ge 0$.

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

$$\Lambda(d(t,m,x)) = \sum_{k:T_{(k)}<\infty} \mathbb{1}\{T_{(k-1)} < t \le T_{(k)}\} \frac{P\Big(\big(T_{(k)}, \Delta_{(k)}, X(T_{(k)})\big) \in d(t,m,x) \mid \mathscr{F}_{T_{(k-1)}}\Big)}{P\Big(T_{(k)} \ge t \mid \mathscr{F}_{T_{(k-1)}}\Big)}$$

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

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

We then have,

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

Letting

$$\zeta_{k,j}\big(dx, t, f_{k-1}\big) := P\Big(X_j\big(T_{(k)}\big) \in dx \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}, T_{(k)} = t, \Delta_{(k)} = j\Big)$$
$$\Lambda_{k,j}\big(dt, f_{k-1}\big) := \frac{P\Big(T_{(k)} \in dt, \Delta_{(k)} = j \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)}{P\Big(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)}$$

completes the proof of (ii).

□

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

Let us define in the censored setting

$$W^g(t) = \prod_{k=1}^{N_t} \frac{\mathbb{1}\{A\big(T_{(k)}\big) = 1\}}{\pi_k\Big(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big)} \frac{\mathbb{1}\{A(0) = 1\}}{\pi_0(L(0))} \prod_{k=1}^{N_t} \frac{\mathbb{1}\{\Delta_{(k)} \neq c\}}{\prod_{u \in (T_{(k-1)}, T_{(k)})}\Big(1 - \Lambda_k^c\Big(\mathrm{d}u, \mathcal{F}_{T_{(k-1)}}\Big)\Big)}$$

in alignment with (1). We verify that our efficient influence function is the same as Rytgaard et al. (2022) in the case with continuous compensators. The efficient influence function of Rytgaard et al. (2022) is given in Theorem 1 of Rytgaard et al. (2022) in our notation by

$$\begin{aligned}
\varphi_\tau^*(P) &= \mathbb{E}_{P^{G^*}}\big[N_y(\tau) \mid \mathcal{F}_0\big] - \Psi_\tau(P) \\
&+ \int_0^\tau W^g(t-)(\mathbb{E}_{P^{G^*}}\big[N_y(\tau) \mid L(t), N^\ell(t), \mathcal{F}_{t-}\big] - \mathbb{E}_{P^{G^*}}\big[N_y(\tau) \mid N^\ell(t), \mathcal{F}_{t-}\big])N^\ell(\mathrm{d}t) \\
&+ \int_0^\tau W^g(t-)(\mathbb{E}_{P^{G^*}}\big[N_y(\tau) \mid \Delta N^\ell(t) = 1, \mathcal{F}_{t-}\big] - \mathbb{E}_{P^{G^*}}\big[N_y(\tau) \mid \Delta N^\ell(t) = 0, \mathcal{F}_{t-}\big])M^\ell(\mathrm{d}t) \\
&+ \int_0^\tau W^g(t-)(\mathbb{E}_{P^{G^*}}\big[N_y(\tau) \mid \Delta N^a(t) = 1, \mathcal{F}_{t-}\big] - \mathbb{E}_{P^{G^*}}\big[N_y(\tau) \mid \Delta N^a(t) = 0, \mathcal{F}_{t-}\big])M^a(\mathrm{d}t) \\
&+ \int_0^\tau W^g(t-)(1 - \mathbb{E}_{P^{G^*}}\big[N_y(\tau) \mid \Delta N^y(t) = 0, \mathcal{F}_{t-}\big])M^y(\mathrm{d}t)
\end{aligned}$$

$$+\int_0^\tau W^g(t-)(0-\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mid \Delta N^d(t)=0,\mathcal{F}_{t-}\right])M^d(\mathrm{d}t).$$

We note, for instance, for $x=\ell$ that

$$\begin{aligned}
&\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mid \Delta N^x(t)=1,\mathcal{F}_{t-}\right]\\
&=\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mid T_{(k)}=t,\Delta_{(k)}=x,\mathcal{F}_{T_{(k-1)}}\right]\\
&=\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\lim_{\varepsilon\to 0}\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mid T_{(k)}\in(t,t+\varepsilon),\Delta_{(k)}=x,\mathcal{F}_{T_{(k-1)}}\right]\\
&=\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\lim_{\varepsilon\to 0}\frac{\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mathbb{1}\left\{T_{(k)}\in(t,t+\varepsilon),\Delta_{(k)}=x\right\}\mid \mathcal{F}_{T_{(k-1)}}\right]}{\mathbb{E}_{P^{G^*}}\left[\mathbb{1}\left\{T_{(k)}\in(t,t+\varepsilon),\Delta_{(k)}=x\right\}\mid \mathcal{F}_{T_{(k-1)}}\right]}\\
&=\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\\
&\qquad\lim_{\varepsilon\to 0}\frac{\mathbb{E}_{P^{G^*}}\left[\mathbb{E}_{P^{G^*}}\left[\bar{Q}^g_{k,\tau}\left(\mathcal{F}_{T_{(k)}}\right)\mid T_{(k)},\Delta_{(k)}=x,\mathcal{F}_{T_{(k-1)}}\right]\mathbb{1}\left\{T_{(k)}<\tau\right\}\mathbb{1}\left\{T_{(k)}\in(t,t+\varepsilon),\Delta_{(k)}=x\right\}\mid \mathcal{F}_{T_{(k-1)}}\right]}{\lambda^x_k\left(t\mid \mathcal{F}_{T_{(k-1)}}\right)}\\
&\overset{*}{=}\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\frac{\mathbb{E}_P\left[\bar{Q}^g_{k,\tau}\left(\mathcal{F}_{T_{(k-1)}}\right)\mid T_{(k)}=t,\Delta_{(k)}=x,\mathcal{F}_{T_{(k-1)}}\right]\lambda^x_k\left(t\mid \mathcal{F}_{T_{(k-1)}}\right)}{\lambda^x_k\left(t\mid \mathcal{F}_{T_{(k-1)}}\right)}\\
&=\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\mathbb{E}_P\left[\bar{Q}^g_{k,\tau}\left(\mathcal{F}_{T_{(k-1)}}\right)\mid T_{(k)}=t,\Delta_{(k)}=x,\mathcal{F}_{T_{(k-1)}}\right]
\end{aligned}$$

where we, in $(*)$, apply dominated convergence. Similarly, we may find that

$$\begin{aligned}
\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mid \Delta N^y(t)=1,\mathcal{F}_{t-}\right]&=1,\\
\mathbb{E}_{P^{G^*}}[N_y(\tau)\mid \Delta N^d(t)=1,\mathcal{F}_{t-}]&=0,\\
\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mid \Delta N^a(t)=1,\mathcal{F}_{t-}\right]&=\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\bar{Q}^g_{k,\tau}\left(\tau\mid \mathcal{F}^1_{T_{(k-1)}}\right)
\end{aligned}$$

For the first term, we find that

$$\begin{aligned}
&\mathbb{E}_{P^{G^*}}[N_y(\tau)\mid L(t),N^\ell(t),\mathcal{F}_{t-}]-\mathbb{E}_{P^{G^*}}[N_y(\tau)\mid N^\ell(t),\mathcal{F}_{t-}]\\
&=\mathbb{E}_{P^{G^*}}[N_y(\tau)\mid L(t),\Delta N^\ell(t)=0,\mathcal{F}_{t-}]-\mathbb{E}_{P^{G^*}}[N_y(\tau)\mid \Delta N^\ell(t)=0,\mathcal{F}_{t-}]\\
&+\mathbb{E}_{P^{G^*}}[N_y(\tau)\mid L(t),\Delta N^\ell(t)=1,\mathcal{F}_{t-}]-\mathbb{E}_{P^{G^*}}[N_y(\tau)\mid \Delta N^\ell(t)=1,\mathcal{F}_{t-}]\\
&=0+\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\left(\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mid L\left(T_{(k)}\right),T_{(k)}=t,\Delta_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\right]-\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mid T_{(k)}=t,\Delta_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\right]\right)\\
&=\sum_{k=1}^K \mathbb{1}\left\{T_{(k-1)}<t\leq T_{(k)}\right\}\mathbb{1}\left(T_{(k)}<\tau,\Delta_{(k)}=\ell,k<K\right)\left(\bar{Q}^g_{k,\tau}\left(L\left(T_{(k)}\right),A\left(T_{(k-1)}\right),T_{(k)},\ell,\mathcal{F}_{T_{(k-1)}}\right)\right.\\
&\qquad\left.-\mathbb{E}_P\left[\bar{Q}^g_{k,\tau}\left(L\left(T_{(k)}\right),A\left(T_{(k-1)}\right),T_{(k)},\Delta_{(k)},\mathcal{F}_{T_{(k-1)}}\right)\mid T_{(k)}=t,\Delta_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\right]\right)]
\end{aligned}$$

Next, note that

$$
\begin{aligned}
&\mathbb{E}_{P^{G^*}}\left[N_y(\tau) \mid \Delta N^x(t)=0, \mathscr{F}_{t-}\right] \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \mathbb{E}_{P^{G^*}}\left[N_y(\tau) \mid \left(T_{(k)} > t \vee T_{(k)} = t, \Delta_{(k)} \neq x\right), \mathscr{F}_{T_{(k-1)}}\right] \\
&\overset{**}{=} \sum_{k=1}^{K} \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \mathbb{E}_{P^{G^*}}\left[N_y(\tau) \mid T_{(k)} > t, \mathscr{F}_{T_{(k-1)}}\right] \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \frac{\mathbb{E}_{P^{G^*}}\left[N_y(\tau)\mathbb{1}\left\{T_{(k)} > t\right\} \mid \mathscr{F}_{T_{(k-1)}}\right]}{\mathbb{E}_{P^{G^*}}\left[\mathbb{1}\left\{T_{(k)} > t\right\} \mid \mathscr{F}_{T_{(k-1)}}\right]} \\
&= \sum_{k=1}^{K} \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \frac{\bar{Q}^g_{k,\tau}\left(\tau \mid \mathscr{F}_{T_{(k-1)}}\right) - \bar{Q}^g_{k,\tau}\left(t \mid \mathscr{F}_{T_{(k-1)}}\right)}{S\left(t \mid \mathscr{F}_{T_{(k-1)}}\right)},
\end{aligned}
$$

where in $(**)$, we use that the event $(T_{(k)} = t, \Delta_{(k)} \neq x)$ has probability zero. Let

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

Combining these facts, we find that the efficient influence function can also be written as:

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

We find immediately that

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

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

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

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

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

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

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

Now note that

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

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

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

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

Let us calculate the second integral

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

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

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

Now note that

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

Setting this into the previous integral, we get

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

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

Thus, we find

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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