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

Johan Sebastian Ohlendorff
johan.ohlendorff@sund.ku.dk
University of Copenhagen

Abstract

In medical research, causal effects of treatments that may change over time on an outcome can be defined in the context of an emulated target trial. We are concerned with estimands that are defined as contrasts of the absolute risk that an outcome event occurs before a given time horizon τ under prespecified treatment regimens. Most of the existing estimators based on observational data require a projection onto a discretized time scale Rose & van der Laan (2011). We consider a recently developed continuous-time approach to causal inference in this setting Rytgaard et al. (2022), which theoretically allows preservation of the precise event timing on a subject level. Working on a continuous-time scale may improve the predictive accuracy and reduce the loss of information. However, continuous-time extensions of the standard estimators comes at the cost of increased computational burden. We will discuss a new sequential regression type estimator for the continuous-time framework which estimates the nuisance parameter models by backtracking through the number of events. This estimator significantly reduces the computational complexity and allows for efficient, single-step targeting using machine learning methods from survival analysis and point processes, enabling robust continuous-time causal effect estimation.

1 Introduction

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

Assumption 1 (Bounded number of events): In the time interval $[0, \tau_{\text{end}}]$ there are at most $K - 1 < \infty$ many changes of treatment and covariates in total for a single individual. Without loss of register data applications, we assume that the maximum number of treatment and covariate changes of an individual is bounded by $K = 10,000$. Practically, we shall adapt K to our data and our target parameter. We let $K - 1$ be given by the maximum number of non-terminal events for any individual in the data.

Assumption 2 (No simultaneous jumps): The counting processes N^a, N^ℓ, N^y, N^d, and N^c have with probability 1 no jump times in common.

Let $\kappa_i(\tau)$ be the number of events for individual i up to time τ. In Rytgaard et al. (2022), the authors propose a continuous-time LTMLE for the estimation of causal effects in which a single step of the targeting procedure must update each of the nuisance estimators $\sum_{i=1}^n \kappa_i(\tau)$ times. We propose an estimator where the number of nuisance parameters is reduced to $\sim \max_i \kappa_i(\tau)$ in total, and, in principle, only one step of the targeting procedure is needed to update all nuisance parameters. We provide an iterative conditional expectation formula that, like Rytgaard et al. (2022), iteratively updates the nuisance parameters. The key difference is that the estimation of the nuisance parameters can be performed by going back in the number of events instead of going back in time. The different approaches are illustrated in Figure 2 and Figure 3 for an outcome Y of interest. Moreover, we argue that the nuisance components can be estimated with existing machine learning algorithms from the survival analysis and point process literature. As always let $(\Omega, \mathcal{F}, P)$ be a probability space on which all processes and random variables are defined.

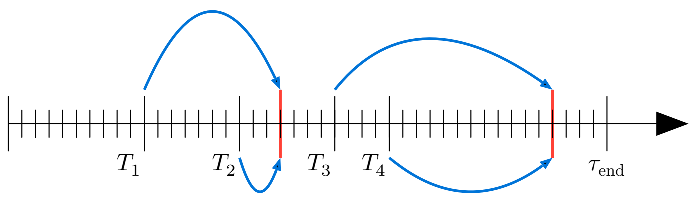

Figure 1: The "usual" approach where time is discretized. Each event time and its corresponding mark is rolled forward to the next time grid point, that is the values of the observations are updated based on the on the events occuring in the previous time interval.

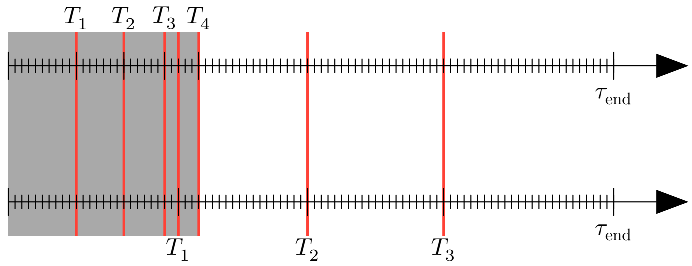

Figure 2: The figure illustrates the sequential regression approach given in Rytgaard et al. (2022) for two observations: Let $t_1 < ... < t_m$ be all the event times in the sample. Then, given $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{t_r}\right]$, we regress back to $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{t_{r-1}}\right]$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

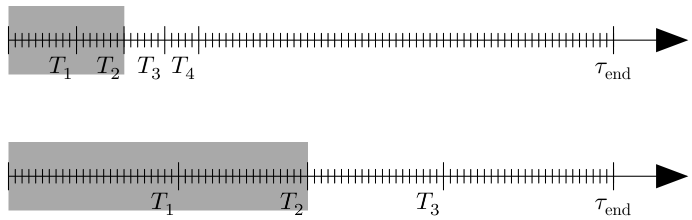

Figure 3: The figure illustrates the sequential regression approach proposed in this article. For each event k in the sample, we regress back on the history $\mathcal{F}_{T_{(k-1)}}$. That is, given $\mathbb{E}_Q\left[Y \mid \mathcal{F}_{T_{(k)}}\right]$, we regress back to $\mathbb{E}_Q\left[Y \mid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

2 Setting and Notation

First, we assume that at baseline, we observe the treatment A_0 and the time-varying confounders at time 0, L_0. The time-varying confounders can in principle include covariates which do not change over time, but for simplicity of notation, we will include them among those that do change over time. Also, we assume that we have two treatment options, $A(t) = 0, 1$ (e.g., placebo and active treatment). The time-varying confounders and treatment are assumed to take values in $\mathbb{R}^m$ and $\mathbb{R}$, and that $L(t) : \Omega \to \mathbb{R}^m$ and $A(t) : \Omega \to \mathbb{R}$ are measurable for each $t \geq 0$, respectively. These processes are assumed to be càdlàg, i.e., right-continuous with left limits. Furthermore, the times at which the treatment and covariates may only change at the jump times of the counting processes N^a and N^ℓ, respectively which makes $L(t)$ and $A(t)$ into jump processes (Last & Brandt (1995)). The jump times of these counting processes thus represent visitation times.

We are interested in the cumulative incidence function, so we also observe N^y and N^d corresponding to the counting processes for the primary and competing event, respectively. Finally, let N^c be the counting process for the censoring counting process. Our the outcome of interest is $Y_\tau = I(T \leq \tau, \Delta = y)$, where T is the time of the terminal event and $\Delta \in \{y, d\}$ is the indicator for which terminal event occurred. We assume that the jump times differ with probability 1 (Assumption 2). Moreover, we assume that only a bounded number of events occur for each individual in the time interval $[0, \tau_{\text{end}}]$ (Assumption 1).

We consider the framework in Rytgaard et al. (2022) and cast it into the framework of marked point processes. To this end, we can define the jump process M as

$$M(s) = \left(N^y(s), N^d(s), N^c(s), L(s), N^\ell(s), A(s), N^a(s)\right) \tag{1}$$

and consider its corresponding natural filtration by

$$\mathcal{F}_t^M = \sigma\left(N^y(s), N^d(s), N^c(s), L(s), N^\ell(s), A(s), N^a(s) \mid s \leq t\right) \tag{2}$$

and the corresponding point process given by

$$(\pi_n(M), k_n(M)) \tag{3}$$

where $\pi_n(M) = T_{(n)}$ is the n'th jump time of M and

$$k_n(M) = \begin{cases} \left(N^y\left(T_{(n)}\right), N^d\left(T_{(n)}\right), N^c\left(T_{(n)}\right), L\left(T_{(n)}\right), N^\ell\left(T_{(n)}\right), A\left(T_{(n)}\right), N^a\left(T_{(n)}\right)\right) \text{ if } T_{(n)} < \infty \\ \nabla \text{ otherwise} \end{cases} \tag{4}$$

Consider the counting process N of M given by

$$N(dt, dx) = \sum_{n=1}^{K} \delta_{(\pi_n(M), k_N(M))}(dt, dx) \tag{5}$$

By reparametrization and Assumption 2, we can essentially use the random measure

$$\widetilde{N}(dt, dx) = \sum_{n=1}^{K} \delta_{\pi_n(M)}(dt)\delta_{\left(D_n, L\left(T_{(n)}\right), A\left(T_{(n)}\right)\right)}(dx) \tag{6}$$

instead, since their histories are the same $\left(\mathcal{F}_t = \sigma(N((0,s],\cdot) \mid s \leq t) \vee \mathcal{F}_0 = \widetilde{\mathcal{F}}_t = \sigma\left(\widetilde{N}((0,s],\cdot) \mid s \leq t\right)\right) \vee \mathcal{F}_0$. Moreover its natural filtration (Theorem 2.5.10 of Last & Brandt (1995) under so-called *minimality* which we will just assume) satisfies,

$$\mathcal{F}_t^N = \sigma\left(\widetilde{N}((0,s],\cdot) \mid s \leq t\right) \vee \mathcal{F}_0 = \mathcal{F}_t^M \tag{7}$$

for $\mathcal{F}_0 = \sigma(L_0, A_0)$. Since N is a marked point process, we may assume the filtration to be right-continuous. Then $\mathcal{F}_{T_{(k)}} = \left(T_{(k)}, D_{(k)}, L_{(k)}, A_{(k)}\right) \vee \mathcal{F}_{T_{(k-1)}}$ is the history up to the k'th event. Our observations can thus be assumed to be on the form $O = \mathcal{F}_{T_{(K)}}$.

Assumption 3 (Conditional distributions of jumps and marks): We assume that the conditional distributions $P\left(T_{(k)} \in \cdot \mid \mathcal{F}_{T_{(k-1)}}\right) \ll m$ P-a.s., and $P\left(A\left(T_{(k)}\right) \in \cdot \mid T_{(k)} = t, D_{(k)} = a, \mathcal{F}_{T_{(k-1)}}\right) \ll \nu_a$ P-a.s. and $P\left(L\left(T_{(k)}\right) \in \cdot \mid T_{(k)} = t, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right) \ll \nu_\ell$ P-a.s., where m is the Lebesgue measure on $\mathbb{R}_+$, ν_a is a measure on $\mathcal{A}$, and ν_ℓ is a measure on $\mathcal{L}$.

Theorem 1 (Existence of compensator): Let $\mathbb{F}_k = (\mathbb{R}_+ \times \{a, \ell, c, d, y\} \times \mathcal{A} \times \mathcal{L})^k$. Under Assumption 1, Assumption 2, and Assumption 3, there exists functions for $k = 1, ..., K$, functions $\lambda_k^x(\cdot, \cdot) : \mathbb{R}_+ \times \mathbb{F}_k \to \mathbb{R}_+$, $\pi_k(\cdot, \cdot, \cdot) : \mathbb{R}_+ \times \mathcal{A} \times \mathbb{F}_k \to \mathbb{R}_+$, and $\mu_k(\cdot, \cdot, \cdot) : \mathbb{R}_+ \times \mathcal{L} \times \mathbb{F}_k \to \mathbb{R}_+$ such that

$$\begin{aligned}\Lambda(dt, dm, da, dl) = \sum_{k=1}^{K} &\mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\}\delta_a(dm)\lambda_{k-1}^a\left(t, \mathcal{F}_{T_{(k-1)}}\right)\pi_{k-1}\left(t, da, \mathcal{F}_{T_{(k-1)}}\right) \\ &+\mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\}\delta_\ell(dm)\lambda_{k-1}^\ell\left(t, \mathcal{F}_{T_{(k-1)}}\right)\mu_{k-1}\left(t, dl, \mathcal{F}_{T_{(k-1)}}\right) \\ &+\mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\}\delta_y(dm)\lambda_{k-1}^y\left(t, \mathcal{F}_{T_{(k-1)}}\right) \\ &+\mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\}\delta_d(dm)\lambda_{k-1}^d\left(t, \mathcal{F}_{T_{(k-1)}}\right) \\ &+\mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\}\delta_c(dm)\lambda_{k-1}^c\left(t, \mathcal{F}_{T_{(k-1)}}\right)\end{aligned} \tag{8}$$

is a P-$\mathcal{F}_t$-compensator measure of N. As a consequence we have,

$$\begin{aligned}&P\left(T_{(k)} \leq s, D_{(k)} = x, L\left(T_{(k)}\right) \in dl, A\left(T_{(k)}\right) \in da \mid \mathcal{F}_{T_{(k-1)}}\right) \\ &= \int_0^s \underbrace{\exp\left(-\sum_{x=y,d,\ell,a,c} \int_0^t \lambda_{k-1}^x\left(s, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}s\right)}_{\text{probability of surviving up to } t.} \underbrace{\lambda_k^x\left(t, \mathcal{F}_{T_{(k)}}\right)}_{\text{probability that it was an event of type } x} \\ &\left(\underbrace{\int_{\mathbb{L}} \mu_{k-1}\left(t, x, \mathcal{F}_{T_{(k-1)}}\right)\nu_\ell(\mathrm{d}x)}_{\text{probability of } L\left(T_{(k)}\right) \in \mathbb{L} \text{ given } D_{(k)}=\ell \text{ and } T_{(k)}=t} \delta_{\left(\ell, A\left(T_{(k-1)}\right)\right)}(\{x\}, da) + \right. \\ &\left. + \underbrace{\int_{\mathbb{A}} \pi_{k-1}\left(t, x, \mathcal{F}_{T_{(k-1)}}\right)\nu_a(\mathrm{d}x)}_{\text{probability of } A\left(T_{(k)}\right) \in \mathbb{A} \text{ given } D_{(k)}=a \text{ and } T_{(k)}=t} \delta_{\left(a, L\left(T_{(k-1)}\right)\right)}(\{x\}, dl) + \delta_{(\{x\}, \emptyset, \emptyset)}(\{d, y, c\}, dl, da) \right) \mathrm{d}t.\end{aligned} \tag{9}$$

on the event $T_{(k-1)} < s$.

Proof: Simply use Theorem 4.1.11 of Last & Brandt (1995) which states that

$$\Lambda(dt, dm, da, dl) = \sum_{k: T_{(k-1)} < \infty} \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \frac{P\left(\left(T_{(k)}, D_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right) \in (dt, dm, da, dl) \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)} \tag{10}$$

is a P-$\mathcal{F}_t$-compensator measure of N. Now rewrite $P\left(\left(T_{(k)}, D_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right) \in (dt, dm, da, dl) \mid \mathcal{F}_{T_{(k-1)}}\right) = P\left(\left(D_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right) \in (dm, da, dl) \mid T_{(k)} = t, \mathcal{F}_{T_{(k-1)}}\right) P\left(T_{(k)} = dt \mid \mathcal{F}_{T_{(k-1)}}\right)$. Under Assumption 3, we can write

$$P\left(D_{(k)} = x \mid T_{(k)} = t, \mathcal{F}_{T_{(K-1)}}\right) \frac{P\left(T_{(k)} \in dt \mid \mathcal{F}_{T_{(k-1)}}\right)}{P\left(T_{(k)} \geq t \mid \mathcal{F}_{T_{(k-1)}}\right)} = \lambda_{k-1}^x\left(t, \mathcal{F}_{T_{(k-1)}}\right) dt \tag{11}$$

which is simply the cause-specific hazard function of the k'th event. Also, we can define

$$\begin{aligned}\pi_{k-1}\left(t, da, \mathcal{F}_{T_{(k-1)}}\right) &= P\left(\left(L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right) \in \left(\left\{L_{(k-1)}\right\}, da\right) \mid T_{(k)} = t, D_{(k)} = a, \mathcal{F}_{T_{(k-1)}}\right) \\ &= P\left(A\left(T_{(k)}\right) \in da \mid T_{(k)} = t, D_{(k)} = a, \mathcal{F}_{T_{(k-1)}}\right) \\ \mu_{k-1}\left(t, dl, \mathcal{F}_{T_{(k-1)}}\right) &= P\left(\left(L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right) \in \left(dl, \left\{A_{(k-1)}\right\}\right) \mid T_{(k)} = t, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right) \\ &= P\left(L\left(T_{(k)}\right) \in dl \mid T_{(k)} = t, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right)\end{aligned} \tag{12}$$

and the result follows. The latter conclusion can be seen by Theorem 4.3.8 of Last & Brandt (1995). □

3 A pragmatic approach to continuous-time causal inference

One classical causal inference perspective requires that we know how the data was generated up to unknown parameters (NPSEM) (Pearl, 2009). This approach has only to a small degree been discussed in continous-time causal inference literature (Røysland et al. (2024)). This is initially considered in the uncensored case, but is later extended to the censored case. For a moment, we ponder how the data was generated given a DAG. The DAG given in the section provides a useful tool for simulating continuous-time data, but not for drawing causal inference conclusions. For the event times, we can draw a figure representing the data generating mechanism which is shown in Figure 4. Some, such as Chamapiwa (2018), write down this DAG, but with an arrow from $T_{(k)}$ to $L(T_{(k)})$ and $A(T_{(k)})$ instead of displaying a multivariate random variable which they deem the "time-as-confounder" approach to allow for irregularly measured data (see Figure 5). Fundamentally, this arrow would only be meaningful if the event time was known prior to the treatment and covariate value measurements, which they might not be. This can make sense if the event is scheduled ahead of time, but for, say, a stroke the time is not measured prior to the event. Because a cause must precede an effect, this makes the arrow invalid from a philosophical standpoint. On the other hand, DAGs such as the one in Figure 4, are not informative about the causal relationships between the variables are. This issue with simultaneous events is likely what has led to the introduction of local independence graphs (Didelez (2008)) but is also related to the notion that the treatment times are not predictable (that is knowable just prior to the event) as in Ryalen (2024).

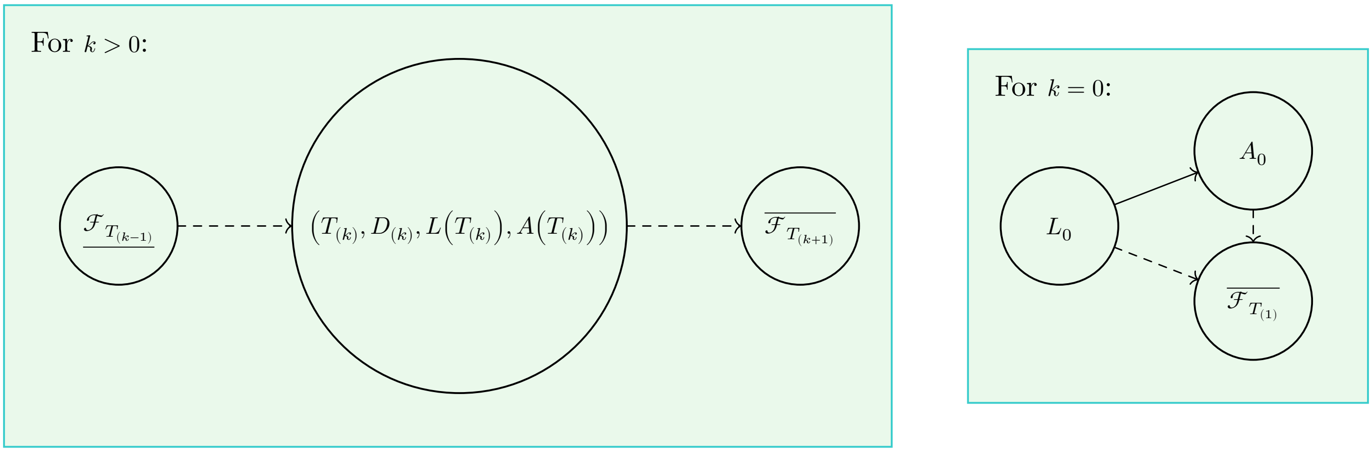

Figure 4: A DAG representing the relationships between the variables of O. The dashed lines indicate multiple edges from the dependencies in the past and into the future. Here $\underline{\mathcal{F}_{T_{(k)}}}$ is the history up to and including the k'th event and $\overline{\mathcal{F}_{T_{(k)}}}$ is the history after and including the k'th event.

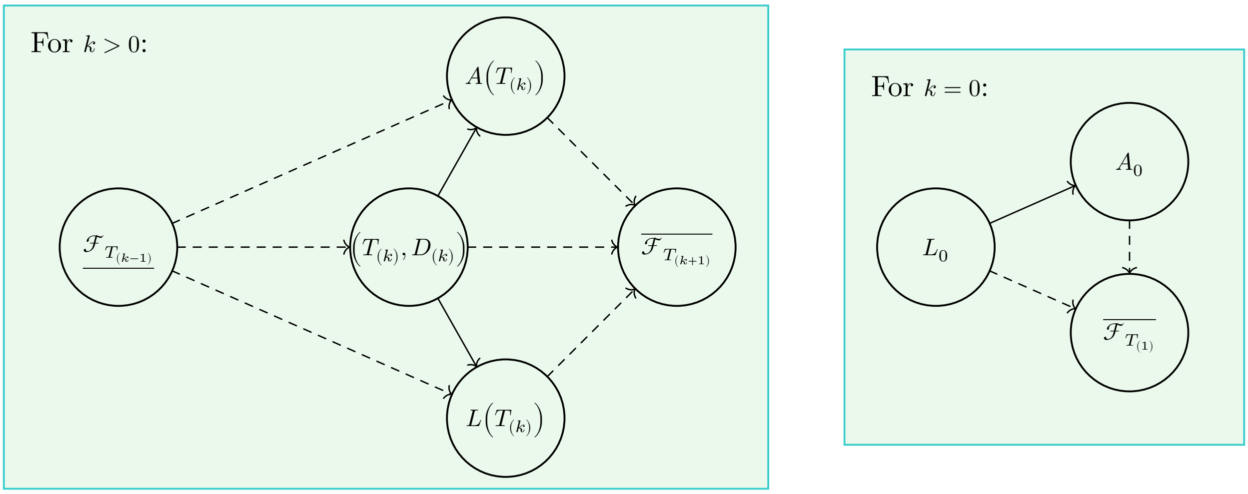

Figure 5: A DAG for simulating the data generating mechanism or such as those that may be found in Chamapiwa (2018). The dashed lines indicate multiple edges from the dependencies in the past and into the future. Here $\underline{\mathcal{F}_{T_{(k)}}}$ is the history up to and including the k'th event and $\overline{\mathcal{F}_{T_{(k)}}}$ is the history after and including the k'th event.

We now take an interventionalist stance to causal inference such as the one given in Ryalen (2024). In the interventionalist school of thought, one tries to emulate a randomized controlled trial. In the continuous-time longitudinal setting, this can e.g., correspond to a trial in which there is perfect compliance. We reformulate the conditions of Ryalen (2024) to our setting, stating the conditions directly in terms of the events instead of using martingales. For simplicity, we presuppose that there are two treatment levels (0/1). As in randomized trials, we suppose that there is a treatment plan g_k at each event point which specifies the treatment that the person observation should have at teach event point which is a treatment event point that is $g_k : \mathbb{R}_+ \times \mathbb{F}_{k-1} \to \{0, 1\}$. Specifically, the plan specifies that $A(T_{(k)}) = g_k(T_{(k)}, \mathcal{F}_{T_{(k-1)}})$ if $T_{(k)} < \infty$ and $D_{(k)} = a$.

We assume the existence of potential outcome process

$$\tilde{O}^g = \left(L_0, \tilde{A}(0), \tilde{T}_{(1)}, \tilde{D}_{(1)}, \tilde{A}\left(\tilde{T}_{(1)}\right), \tilde{L}\left(\tilde{T}_{(1)}\right), ..., \tilde{T}_{(K)}, \tilde{D}_{(K)}\right)$$
$$\text{with } \tilde{A}\left(\tilde{T}_{(k)}\right) = \begin{cases} g_k\left(\tilde{T}_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \text{ if } \tilde{T}_k < \infty \text{ and } D_{(k)} = a \\ \tilde{A}\left(\tilde{T}_{(k-1)}\right) \text{ if } \tilde{T}_k < \infty \text{ and } D_{(k)} \neq a \\ \nabla \text{ otherwise} \end{cases} \tag{13}$$

We choose not to index the random variables with g when it can clearly be inferred from the context. Our potential outcome, which is part of this process, is given by

$$\tilde{Y}_t^g = \left(\mathbb{1}\left\{\tilde{T}_1 \leq t, \tilde{D}_1 = y\right\}, ..., \mathbb{1}\left\{\tilde{T}_K \leq t, \tilde{D}_K = y\right\}\right) \tag{14}$$

We are then interested in estimating the causal parameter given in Definition 1.

Definition 1 (Target parameter): Our target parameter $\Psi_\tau^g : \mathcal{M} \to \mathbb{R}$ is the mean interventional potential outcome at time τ given the intervention plan g,

$$\Psi_\tau^g(P) = \mathbb{E}_P\left[\sum_{k=1}^K \mathbb{1}\left\{\tilde{T}_k \leq \tau, \tilde{D}_k = y\right\}\right] \tag{15}$$

The three identifying conditions for the target parameter are as follows:

1. For all $k = 1, ..., K$, $\mathbb{1}\left\{\tilde{T}_k \leq \tau, \tilde{D}_k = y\right\} = \mathbb{1}\{T_k \leq \tau, D_k = y\}$ P-a.s. if there does not exist a treatment plan $j < k$ such that $A(T_{(j)}) \neq g_j\left(T_{(j)}, \mathcal{F}_{T_{(j-1)}}\right)$ with $D_{(j)} = a$ and $T_{(j)} < \infty$ and $A_0 = g_0(L_0)$ (consistency).
2. For all $k = 1, ..., K$, $A(T_{(k)}) \perp \left(\mathbb{1}\left\{\tilde{T}_{k+1} \leq t, \tilde{D}_{k+1} = y\right\}, ..., \mathbb{1}\left\{\tilde{T}_K \leq t, \tilde{D}_K = y\right\}\right) \mid D_{(k)} = a, \mathcal{F}_{T_{(k-1)}}$ (exchangeability).

3. Let $w_{k-1}(l_0, a_0, f_1, ..., f_k) = \frac{\mathbb{1}\{a_0=g(l_0)\}}{\pi_0(g(l_0))} \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{a_j=g_j(t_j,f_{j-1})\}}{\pi_j(t_j,g_j(t_j,f_{j-1}))}\right)^{\mathbb{1}\{d_j=a\}} \mathbb{1}\{t_1 < ... < t_k\}$. The measure $Q_\tau = \left(\sum_{k=1}^K w_k\right) \cdot P$ is a valid probability measure (positivity).

Then we have the following theorem

Theorem 2 (Identification via inverse probability weights): Under the conditions of 1., 2., and 3., the target parameter is identified by

$$\Psi_\tau^g(P) = \mathbb{E}_P\left[\sum_{k=1}^K w_{k-1}\mathbb{1}\{T_k \le \tau, D_k = y\}\right] \tag{16}$$

Proof: We will show this by proving that $\psi_{k,\tau}(P) = \mathbb{E}_P[w_{k-1}\mathbb{1}\{T_k \le \tau, D_k = y\}] = \mathbb{E}_P\left[\mathbb{1}\left\{\tilde{T}_k \le \tau, \tilde{D}_k = y\right\}\right]$. Let $Y_{k,j}^* = \mathbb{E}_P\left[\mathbb{1}\left\{\tilde{T}_k \le \tau, \tilde{D}_k = y\right\} \mid \left(T_{(j-1)}, D_{(j-1)}, L\left(T_{(j-1)}\right), \mathcal{F}_{T_{(j-2)}}^{-a}\right)\right]$

Let $g_k^* = g_k$ if $D_{(k)} = a$ and $T_{(k)} < \infty$ and $g_k^* = g_{k-1}^*$ otherwise. We use the law of iterated expectations to find that

$$\begin{aligned}
\psi_{k,\tau}(P) &= \mathbb{E}_P\left[w_{k-1}\mathbb{E}_P\left[\mathbb{1}\{T_k \le \tau, D_k = y\} \mid \mathcal{F}_{T_{(k-1)}}\right]\right] \\
&= \mathbb{E}_P\left[w_{k-1}\mathbb{1}\{t_1 < ... < t_{k-1}\}\mathbb{E}_P\left[\mathbb{1}\{T_k \le \tau, D_k = y\} \mid \left(T_{(k-1)}, D_{(k-1)}, L\left(T_{(k-1)}\right), A\left(T_{(k-1)}\right) = g_k^*\left(T_{(k-1)}, \mathcal{F}_{T_{(k-2)}}\right), ..., A_0 = g_0(L_0), L_0\right)\right]\right] \\
&= \mathbb{E}_P\left[w_{k-1}\mathbb{1}\{t_1 < ... < t_{k-1}\}\mathbb{E}_P\left[\mathbb{1}\left\{\tilde{T}_k \le \tau, \tilde{D}_k = y\right\} \mid \left(T_{(k-1)}, D_{(k-1)}, L\left(T_{(k-1)}\right), A\left(T_{(k-1)}\right) = g_k^*\left(T_{(k-1)}, \mathcal{F}_{T_{(k-2)}}\right), ..., A_0 = g_0(L_0), L_0\right)\right]\right] \\
&= \mathbb{E}_P[w_{k-1}\mathbb{1}\{t_1 < ... < t_{k-1}\}\mathbb{E}_P\left[\mathbb{1}\left\{\tilde{T}_k \le \tau, \tilde{D}_k = y\right\} \mid \left(T_{(k-1)}, D_{(k-1)}, L\left(T_{(k-1)}\right), \mathcal{F}_{T_{(k-2)}}^{-a}\right)\right] \\
&= \mathbb{E}_P[w_{k-2}\mathbb{1}\{t_1 < ... < t_{k-1}\}Y_k^*\mathbb{E}_P\left[\left(\frac{\mathbb{1}\left\{A\left(T_{(k-1)}\right) = g_{k-1}\left(T_{(k-1)}, \mathcal{F}_{T_{(k-1)}}^g\right)\right\}}{\pi_{k-1}\left(T_{(k-1)}, g_{k-1}\left(T_{(k-1)}, \mathcal{F}_{T_{(k-1)}}\right)\right)}\right)^{\mathbb{1}\{D_{(k-1)}=a\}} \mid \left(T_{(k-1)}, D_{(k-1)}, L\left(T_{(k-1)}\right), \mathcal{F}_{T_{(k-2)}}^{-a}\right)\right] \\
&= \mathbb{E}_P[w_{k-2}\mathbb{1}\{t_1 < ... < t_{k-1}\}Y_k^*] \\
&= \mathbb{E}_P[w_{k-2}\mathbb{1}\left\{T_{(1)} < ... < T_{(k-2)}\right\}\mathbb{E}_P\left[Y_k^*\mathbb{1}\left\{T_{(k-2)} < T_{(k-1)}\right\} \mid \mathcal{F}_{\left(T_{(k-2)}\right)}^{-a}\right] \\
&= \mathbb{E}_P\left[w_{k-2}\mathbb{1}\left\{T_{(1)} < ... < T_{(k-2)}\right\}Y_{k,k-1}^*\right] \\
&\dots \\
&= \mathbb{E}_P\left[\mathbb{1}\left\{\tilde{T}_k \le \tau, \tilde{D}_k = y\right\}\right]
\end{aligned} \tag{17}$$

□

We first state and prove a formula for at target parameter that is not causal, but we will use it to identify the causal parameter. This will be useful for the derivation of the efficient influence function.

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

$$\begin{aligned}
\bar{Q}_{k-1} = \mathbb{E}_P\Big[&I\left(T_{(k)} \le \tau, D_{(k)} = \ell\right)\bar{Q}_k\left(A\left(T_{(k-1)}\right), L\left(T_{(k)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \\
&+ I\left(T_{(k)} \le \tau, D_{(k)} = a\right)\mathbb{E}_P\left[\bar{Q}_k\left(A\left(T_{(k)}\right), L\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right] \\
&+ I\left(T_{(k)} \le \tau, D_{(k)} = y\right) \Big| \mathcal{F}_{T_{(k-1)}}\Big]
\end{aligned} \tag{18}$$

for $k = K, ..., 1$. Thus, $\mathbb{E}_P\left[\sum_{k=1}^K I\left(T_{(k)} \le \tau, D_{(k)} = y\right)\right] = \mathbb{E}_P\left[\bar{Q}_0\right]$.

Proof: We find

$$
\begin{aligned}
\bar{Q}_k &= \mathbb{E}_P\left[\sum_{j=k+1}^{K} I\big(T_{(j)} \le \tau, D_{(j)} = y\big) \mid \mathcal{F}_{T_{(k)}}\right] \\
&= \mathbb{E}_P\left[\mathbb{E}_P\left[\mathbb{E}_P\left[\sum_{j=k+1}^{K} I\big(T_{(j)} \le \tau, D_{(j)} = y\big) \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, D_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \mid \mathcal{F}_{T_{(k)}}\right] \\
&= \mathbb{E}_P\Bigg[I\big(T_{(k+1)} \le \tau, D_{(k+1)} = y\big) \\
&\qquad\qquad + \mathbb{E}_P\left[\mathbb{E}_P\left[\sum_{j=k+2}^{K} I\big(T_{(j)} \le \tau, D_{(j)} = y\big) \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, D_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \Bigg| \mathcal{F}_{T_{(k)}}\Bigg] \qquad (19) \\
&= \mathbb{E}_P\Bigg[I\big(T_{(k+1)} \le \tau, D_{(k+1)} = y\big) \\
&+ I\big(T_{(k+1)} \le \tau, D_{(k+1)} = a\big)\mathbb{E}_P\left[\mathbb{E}_P\left[\sum_{j=k+2}^{K} I\big(T_{(j)} \le \tau, D_{(j)} = y\big) \mid \mathcal{F}_{T_{(k+1)}}\right] \mid T_{(k+1)}, D_{(k+1)}, \mathcal{F}_{T_{(k)}}\right] \mid \mathcal{F}_{T_{(k)}}\Bigg] \\
&+ \mathbb{E}_P\left[I\big(T_{(k+1)} \le \tau, D_{(k+1)} = \ell\big)\mathbb{E}_P\left[\sum_{j=k+2}^{K} I\big(T_{(j)} \le \tau, D_{(j)} = y\big) \mid \mathcal{F}_{T_{(k+1)}}\right] \mid \mathcal{F}_{T_{(k)}}\right]
\end{aligned}
$$

by the law of iterated expectations and that

$$
\big(T_{(k)} \le \tau, D_{(k)} = y\big) \subseteq \big(T_{(j)} \le \tau, D_{(j)} \in \{a, \ell\}\big) \qquad (20)
$$

for all $j = 1, ..., k-1$ and $k = 1, ..., K$. $\square$

Theorem 3 (Identification via g-formula): Let $\bar{Q}^a_{k,\tau} = \bar{Q}_k(Q)$ be defined as in the previous theorem for Q. Let

$$\begin{aligned}
&p_{ka}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) \\
&= \int_{T_{(k)}}^{t} \exp\left(-\sum_{x\in\{\ell,a,d,y\}} \int_{T_{(k)}}^{s} \lambda^x_k\left(u, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}u\right)\lambda^a_k\left(s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad\qquad \times \left(\int_{\mathcal{A}} \bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k-1)}\right), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\right)\pi^*_k\left(s, a_k, \mathcal{F}_{T_{(k)}}\right)\nu_A(\mathrm{d}a_k)\right)\mathrm{d}s \\
&p_{k\ell}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) \\
&= \int_{T_{(k)}}^{t} \exp\left(-\sum_{x\in\{\ell,a,d,y\}} \int_{T_{(k)}}^{s} \lambda^x_k\left(u, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}u\right)\lambda^\ell_k\left(s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad\qquad \times \left(\int_{\mathcal{L}} \bar{Q}^a_{k+1,\tau}\left(l_k, A\left(T_{(k-1)}\right), s, \ell, \mathcal{F}_{T_{(k-1)}}\right)\mu_k\left(s, l_k, \mathcal{F}_{T_{(k)}}\right)\nu_L(\mathrm{d}l_k)\right)\mathrm{d}s \\
&= \int_{T_{(k)}}^{t} \exp\left(-\sum_{x\in\{\ell,a,d,y\}} \int_{T_{(k)}}^{s} \lambda^x_k\left(u, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}u\right)\lambda^\ell_k\left(s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad\qquad \times \left(\mathbb{E}_P\left[\bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\right)\mathrm{d}s \\
&p_{ky}\left(t \mid \mathcal{F}_{T_{(k-1)}}\right) \\
&= \int_{T_{(k)}}^{t} \exp\left(-\sum_{x\in\{\ell,a,d,y\}} \int_{T_{(k)}}^{s} \lambda^x_k\left(u, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}u\right)\lambda^y_k\left(s, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}s
\end{aligned} \tag{21}$$

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

$$\bar{Q}^a_{k,\tau} = p_{ka}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{k\ell}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) + p_{ky}\left(\tau \mid \mathcal{F}_{T_{(k-1)}}\right) \tag{22}$$

Alternatively, we can apply inverse probability of censoring weighting to obtain

$$\begin{aligned}
\bar{Q}^a_{k-1,\tau} = \mathbb{E}_P\Bigg[&\frac{I\left(T_{(k)} \le \tau, D_{(k)} = \ell\right)}{\exp\left(-\int_{T_{(k-1)}}^{T_{(k)}} \lambda^c\left(s \mid \mathcal{F}_{T_{(k)}}\right)\mathrm{d}s\right)}\bar{Q}^a_{k,\tau}\left(A\left(T_{(k-1)}\right), L\left(T_{(k)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \\
&+\frac{I\left(T_{(k)} \le \tau, D_{(k)} = a\right)}{\exp\left(-\int_{T_{(k-1)}}^{T_{(k)}} \lambda^c\left(s \mid \mathcal{F}_{T_{(k)}}\right)\mathrm{d}s\right)} \\
&\qquad \times \int \bar{Q}^a_{k,\tau}\left(a_k, L\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right)\pi^*_{k-1}\left(T_{(k)}, a_k, \mathcal{F}_{T_{(k-1)}}\right)\nu_A(\mathrm{d}a_k) \\
&+\frac{I\left(T_{(k)} \le \tau, D_{(k)} = y\right)}{\exp\left(-\int_{T_{(k-1)}}^{T_{(k)}} \lambda^c\left(s \mid \mathcal{F}_{T_{(k)}}\right)\mathrm{d}s\right)} \Bigg| \mathcal{F}_{T_{(k-1)}}\Bigg]
\end{aligned} \tag{23}$$

for $k = K-1, ..., 1$. This is Method 3. Then,

$$\Psi_\tau(Q) = \mathbb{E}_P\left[\int \bar{Q}^a_{0,\tau}(a, L_0)\pi^*_0(0, a, \mathcal{F}_0)\nu_A(\mathrm{d}a)\right]. \tag{24}$$

Proof: The theorem is an immediate consequence of Theorem 1 and Lemma 1 (the sets $\left(T_{(k)} \le t, D_{(k)} = x, L\left(T_{(k)}\right) \in \mathbb{L}, A\left(T_{(k)}\right) \in \mathbb{A}\right)$ fully determine the regular conditional distribution of $\left(T_{(k)}, D_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right)\right)$ given $\mathcal{F}_{T_{(k-1)}}$). □

Interestingly, Equation 22 corresponds exactly with the target parameter of Rytgaard et al. (2022) and Gill & Robins (2023) by plugging in the definitions of $\bar{Q}^a_{k,\tau}$ and simplifying (to be shown).

We recommend combining Method 2 to deal with high-dimensional confounding initially. Afterwards when there is sparsity use Method 1.

4 Implementation of Method 2

First, we rewrite to the interarrival times. Note that the hazard of the interarrival time $S_k = T_k - T_{k-1}$ is $\lambda^x\left(w + T_{(k-1)} \mid \mathcal{F}_{T_{(k-1)}}\right)$

$$\begin{aligned}
\bar{Q}^a_{k,\tau} &= \int_{T_{(k)}}^{\tau} \exp\left(-\sum_{x\in\{\ell,a,d,y\}} \int_{T_{(k)}}^{s} \lambda^x_k\left(u, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}u\right)\lambda^a_k\left(s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad\times\left(\int_{\mathcal{A}} \bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k-1)}\right), a_k, s, a, \mathcal{F}_{T_{(k)}}\right)\pi^*_k\left(s, a_k, \mathcal{F}_{T_{(k)}}\right)\nu_A(\mathrm{d}a_k)\right)\mathrm{d}s \\
&+\int_{T_{(k)}}^{\tau} \exp\left(-\sum_{x\in\{\ell,a,d,y\}} \int_{T_{(k)}}^{s} \lambda^x_k\left(u, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}u\right)\lambda^\ell_k\left(s, \mathcal{F}_{T_{(k)}}\right) \\
&\qquad\times\left(\mathbb{E}_P\left[\bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\right)\mathrm{d}s \\
&+\int_{T_{(k)}}^{\tau} \exp\left(-\sum_{x\in\{\ell,a,d,y\}} \int_{T_{(k)}}^{s} \lambda^x_k\left(u, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}u\right)\lambda^y_k\left(s, \mathcal{F}_{T_{(k)}}\right)\mathrm{d}s \\
&= \int_0^{\tau - T_{(k-1)}} \exp\left(-\sum_{x=a,\ell,y,d} \int_0^s \lambda^x\left(w + T_{(k-1)} \mid \mathcal{F}_{T_{(k-1)}}\right)\mathrm{d}w\right) \\
&\times\Bigg(\lambda^a_k\left(T_{(k-1)} + s, \mathcal{F}_{T_{(k)}}\right)\int_{\mathcal{A}} \bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k-1)}\right), a_k, T_{(k-1)} + s, a, \mathcal{F}_{T_{(k)}}\right)\pi^*_k\left(T_{(k-1)} + s, a_k, \mathcal{F}_{T_{(k)}}\right)\nu_A(\mathrm{d}a_k) \\
&+\lambda^\ell_k\left(T_{(k-1)} + s, \mathcal{F}_{T_{(k)}}\right)\left(\mathbb{E}_P\left[\bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = T_{(k-1)} + s, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}}\right]\right) \\
&+\lambda^y_k\left(T_{(k-1)} + s, \mathcal{F}_{T_{(k)}}\right)\Bigg))
\end{aligned} \tag{25}$$

Now by using a piecewise estimator of the cumulative hazard, such as the Nelson-Aalen estimator, Cox model, or a random survival forest. Let $M^x_k = \left\{i \in \{1, ..., n\} \mid D_{(k),i} = x\right\}$. Let $S_{(k),i}$ be the time of the i'th event (ordered) in the sample. Then we can estimate the cumulative hazard for the interarrival times as

$$\widehat{\Lambda^x_{k,t}}(f_{k-1}) = \sum_{i\in M^x_k} I\left(S_{(k),i-1} < t \le S_{(k),i}\right)\hat{A}^x_{i,k}(f_{k-1}) \tag{26}$$

for values $\hat{A}^x_{i,k}$ that are estimated from the data. Let $c^x_k(t) = \sup\{i \in M^x_k \mid T_{k,i} \le t\}$. Then we can estimate the integrals in Equation 22, for example the integral corresponding to covariate change,

$$\begin{aligned}
&\sum_{i=1}^{c_k(\tau)} \exp\left(-\sum_{x=a,\ell,y,d} \hat{A}^x_{i,k}(f_{k-1})\right)\left(\widehat{\mathbb{E}_P}\left[\bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = t_{k-1} + S_{(k),i}, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} = f_{k-1}\right]\right) \\
&\left(\hat{A}^\ell_{i,k} - \hat{A}^\ell_{i-1,k}\right)(f_{k-1})
\end{aligned} \tag{27}$$

for given covariate history.

We begin with an algorithm for Method 2, in which it is initially assumed that all relevant models can be fitted.

4.1 Surrogate/approximate modeling of the $\bar{Q}$ function

Let K_τ be the maximal number of non-terminal events that occur before time τ. We suppose we are given estimators $\hat{\Lambda}_k(t)^x$ of the cause-specific hazard functions for the interarrival times $S_k = T_{(k)} - T_{(k-1)}$, which are piecewise constant.

1. Let $\mathcal{R}^y_{K,\tau} = \left\{i \in \{1, ..., n\} \mid D_{(K),i} \in \{a, \ell\}, T_{(K),i} \le \tau, D_{(K),i} = y\right\}$.
2. Calculate for each $j \in \mathcal{R}_{K,\tau}$, $\hat{p}_{Ky} = \sum_{i=1}^{c^y_K(\tau)} \exp\left(-\sum_{x=y,d} \hat{\Lambda}^x_K\left(\mathcal{F}^j_{T_K}\right)\right)\left(\hat{\Lambda}^y_K\left(\mathcal{F}^j_{T_K}\right) - \hat{\Lambda}^y_K\left(\mathcal{F}^j_{T_{K-1}}\right)\right)$.

3. Use surrogate model $\tilde{p}_{Ky}$ for $\hat{p}_{Ky}$ to learn p_{Ka} and thus $\bar{Q}_K$.

For each event point $k = K-1, ..., 1$:
1. For each $j \in \mathcal{R}_{k,\tau}$, calculate $\hat{p}_{ky}\left(\tau \mid \mathcal{F}^j_{T_{k-1}}\right)$ as earlier. Also, calculate $\hat{p}_{ka}\left(\tau \mid \mathcal{F}^j_{T_{k-1}}\right)$ and $\hat{p}_{k\ell}\left(\tau \mid \mathcal{F}^j_{T_{k-1}}\right)$ based on the surrogate model $\tilde{Q}_{k+1}$.

At baseline $k = 0$:
1. Calculate $\hat{\Psi}_n = \frac{1}{n}\sum_{i=1}^n \bar{Q}^a_{0,\tau}(A_0, L^i_0)$.

5 Implementation of the Iterative Conditional Expectations formula

We assume that K_τ is the 1 + the maximal number of non-terminal events that occur before time τ. For now, we assume that this is number is fixed and does not depend on the sample. Let $\tilde{Y}_k(t) = I\left(T_{(k-1)} < t \leq T_{(k)}\right)$.

For $k = K_\tau - 1, ..., 0$:
- We want a prediction function $\bar{Q}^a_{k,\tau}$ of the history up to the k'th event, that is $\bar{Q}^a_{k,\tau} : \mathcal{H}_k \longrightarrow \mathbb{R}$, given that we have one for the $(k+1)$'th event, i.e., $\bar{Q}^a_{k+1,\tau} : \mathcal{H}_k \longrightarrow \mathbb{R}$ (note that for $k = K_\tau$, we have $\bar{Q}^a_{K_\tau,\tau} = I\left(T_{(K_\tau)} \leq \tau, D_{(K_\tau)} = y\right).$). We consider the data set $\mathcal{D}_{k,n}$ that is obtained from the original data $\mathcal{D}_n$ by only considering the observations that have had k non-terminal events, that is $D_{(k)} \in \{a, \ell\}$ for $j = 1, ..., k$. On this data:
  - We estimate $\lambda^c_{k+1}\left(\cdot, \mathcal{F}_{T_{(k+1)}}\right)$ by using $T_{(k+1)}$ as the time-to-event and $D_{(k+1)}$ as the event indicator on the data set $\mathcal{D}_{k,n}$, regressing on $\mathcal{F}_{T_{(k)}} = \left(L\left(T_{(k)}\right), A\left(T_{(k)}\right), T_{(k)}, D_{(k)}, ..., L_0, A_0\right)$[1]

  We are now able provide estimated values for the integrand in Equation 23. These values are provided on the smaller data set $\mathcal{D}_{k,n,\tau}$ of $\mathcal{D}_{k,n}$ where we only consider the observations with $T_{(k)} \leq \tau$. This is done as follows:
  1. For observations in $\mathcal{D}_{k,n,\tau}$ with $D_{(k+1)} = \ell$ and $T_{(k+1)} \leq \tau$, use the previous function to predict values $\bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k+1)}\right), A\left(T_{(k)}\right), T_{(k+1)}, D_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)$.
  2. For observations in $\mathcal{D}_{k,n,\tau}$ with $D_{(k+1)} = a$ and $T_{(k+1)} \leq \tau$, integrate using the previous function $\int_{\mathcal{A}} \bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), a_k, T_{(k+1)}, D_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\pi^*_k\left(T_{(k+1)}, a_k, \mathcal{F}_{T_{(k)}}\right)\nu_A(\mathrm{d}a_k)$. If for example the intervention sets the treatment to 1, then $\int_{\mathcal{A}} \bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), a_k, T_{(k+1)}, D_{(k+1)}, \mathcal{F}_{T_{(k)}}\right)\pi^*_k\left(T_{(k)}, 1, \mathcal{F}_{T_{(k)}}\right)\nu_A(\mathrm{d}a_k) = \bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), 1, T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k)}}\right)$. This gives predicted values for this group.
  3. For observations in $\mathcal{D}_{k,n,\tau}$ with $D_{(k+1)} = y$ and $T_{(k+1)} \leq \tau$, simply put the values equal to 1.
  4. For all other observations put their values equal to 0.

  For all the observations, divide the corresponding values by estimates of censoring survival function $\exp\left(-\int_{T_{(k)}}^{T_{(k+1)}} \lambda^c\left(s \mid \mathcal{F}_{T_{(k)}}\right)\mathrm{d}s\right)$. We then regress the values on $\mathcal{F}_{T_{(k)}} = \left(L\left(T_{(k)}\right), A\left(T_{(k)}\right), T_{(k)}, D_{(k)}, ..., L_0, A_0\right)$. From this regression, we set $\bar{Q}^a_{k,\tau}$ to be the predicted values of the function from the regression.
- If $k = 0$: We estimate the target parameter via $\mathbb{P}_n\left[\sum_{k=1}^{K_\tau} \bar{Q}^a_{0,\tau}(\cdot, a_0)\nu_A(\mathrm{d}a_0)\right]$.

Note: The $\bar{Q}^a_{k,\tau}$ have the interpretation of the heterogenous causal effect after k events.

For now, we recommend Equation 23 for estimating $\bar{Q}^a_{k,\tau}a$: For estimators of the hazard that are piecewise constant, we would need to compute integrals for each unique pair of history and event times occurring in the sample at each event k. On the other hand, the IPCW approach is very sensitive to the specification of the censoring distribution. Something very similar can be written down when we use Equation 22.

5.1 Alternative nuisance parameter estimators

An alternative is to estimate the entire cumulative hazards Λ^x at once instead of having K separate parameters: There are very few methods for marked point process estimation but see Liguori et al. (2023) for methods mostly

[1] We abuse the notation a bit by writing $\mathcal{F}_{T_{(k)}}$ here, but it is actually a σ-algebra.

based on neural networks or Weiss & Page (2013) for a forest-based method. As a final alternative, we can use temporal difference learning to avoid iterative estimation of $\bar{Q}^a, \tilde{Q}$ Shirakawa et al. (2024). Most point process estimators are actually on the form given in terms of ref:intensity.

6 The efficient influence function

We want to use machine learning estimators of the nuisance parameters, so to get inference we need to debias our estimate with the efficient influence function, e.g., double/debiased machine learning Chernozhukov et al. (2018) or targeted minimum loss estimation van der Laan & Rubin (2006). We use Equation 23 for censoring to derive the efficient influence function, because it will contain fewer martingale terms. Let $N^c_k(t) = N_t(\{c\} \times \mathcal{L} \cup \{\emptyset\} \times \mathcal{A} \cup \{\emptyset\})$.

Theorem 4 (Efficient influence function): Let $N^x_k = N_t(\{x\} \times \mathcal{L} \cup \{\emptyset\} \times \mathcal{A} \cup \{\emptyset\})$ and $\tilde{Y}_{k-1}(t) = I\left(T_{(k-1)} < t \leq T_{(k)}\right)$. The efficient influence function is given by

$$\begin{aligned}\varphi^*_\tau(P) = \sum_{k=1}^K \prod_{j=0}^{k-1} & \left(\frac{\pi^*_{j-1}\left(T_{(j)}, A\left(T_{(j)}\right), \mathcal{F}_{T_{(j-1)}}\right)}{\pi_{j-1}\left(T_{(j)}, A\left(T_{(j)}\right), \mathcal{F}_{T_{(j-1)}}\right)}\right)^{I\left(D_{(j)}=a\right)} \frac{I\left(D_{(k-1)} \in \{\ell, a\}, T_{(k-1)} \leq \tau\right)}{\exp\left(-\sum_{1 \leq j < k} \int_{T_{(j-1)}}^{T_{(j)}} \lambda^c_{j-1}\left(s, \mathcal{F}_{T_{(j-1)}}\right) \mathrm{d}s\right)} \\ & \times \Bigg[\left(\bar{Z}^a_{k,\tau} - \bar{Q}^a_{k-1,\tau}\left(\tau, \mathcal{F}_{T_{(k-1)}}\right)\right) \\ & + \int_{T_{(k-1)}}^\tau \left(\bar{Q}^a_{k-1,\tau}(\tau) - \bar{Q}^a_{k-1,\tau}(u)\right) \frac{1}{\exp\left(-\int_{T_{(k-1)}}^u \sum_{x=a,\ell,d,y,c} \lambda^x_{k-1}\left(s, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}s\right)} \left(N^c_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^c_{k-1}\left(s, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}s\right)\Bigg] \\ & + \int \bar{Q}^a_{1,\tau}(a, L_0) \pi^*_0(0, a, \mathcal{F}_0) \nu_A(\mathrm{d}a) - \Psi_\tau(P)\end{aligned} \tag{28}$$

(we take the empty sum to be zero and define $T_0 = 0$, $D_{(0)} = a$ and $\mathcal{F}_{T_{(-1)}} = L(0)$.)

Proof: Define (sorry about the notation!)

$$\begin{aligned}\bar{Z}^a_{k,\tau}(s, t_k, d_k, l_k, a_k, f_{k-1}) = & \frac{I(t_k \leq s, d_k = \ell)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)} \bar{Q}^a_{k,\tau}(a_{k-1}, l_k, t_k, d_k, f_{k-1}) \\ & + \frac{I(t_k \leq s, d_k = a)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)} \\ & \quad \times \int \bar{Q}^a_{k,\tau}(\tilde{a}_k, l_{k-1}, t_k, d_k, f_{k-1}) \pi^*_{k-1}\left(t_k, \tilde{a}_k, \mathcal{F}_{T_{(k-1)}}\right) \nu_A(\mathrm{d}\tilde{a}_k) \\ & + \frac{I(t_k \leq s, d_k = y)}{\exp\left(-\int_{t_{k-1}}^{t_k} \lambda^c(s \mid f_{k-1}) \mathrm{d}s\right)}, s \leq \tau\end{aligned} \tag{29}$$

and let

$$\bar{Q}^a_{k-1,\tau}(s) = \mathbb{E}_P\left[\bar{Z}^a_{k,s}\left(\tau, T_{(k)}, D_{(k)}, L\left(T_{(k)}\right), A\left(T_{(k)}\right), \mathcal{F}_{T_{(k-1)}}\right) \mid \mathcal{F}_{T_{(k-1)}}\right], s \leq \tau \tag{30}$$

We compute the efficient influence function by taking the Gateaux derivative of the above with respect to P, by discretizing the time. We will use two well-known "results" for the efficient influence function.

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon} \int_{T_{(k-1)}}^{t} \lambda_{\varepsilon}^{x}\Big(s \mid \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}s \\
&= \frac{\delta_{\mathcal{F}_{T_{(k-1)}}}(f_{k-1})}{P\Big(\mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)} \int_{T_{(k-1)}}^{t} \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda_{k-1}^{x}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big)} \Big(N_{k}^{x}(\mathrm{d}s) - \tilde{Y}_{k-1}(s)\lambda_{k-1}^{x}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big)
\end{aligned} \tag{31}
$$

and

$$
\frac{\partial}{\partial \varepsilon} \mathbb{E}_{(1-\varepsilon)P + \varepsilon\delta_{(Y,X)}}[Y \mid X = x]\Big|_{\varepsilon=0} = \frac{\delta_X(x)}{P(X=x)}(Y - \mathbb{E}_P[Y \mid X = x]) \tag{32}
$$

We will recursively calculate the derivative,

$$
\frac{\partial}{\partial \varepsilon} \bar{Q}_{k-1,\tau}^{a,\varepsilon}(a_{k-1}, l_{k-1}, t_{t-1}, d_{k-1}, f_{k-2})\Big((1-\varepsilon)P + \varepsilon\delta_{\mathcal{F}_{T_{(k-1)}}}\Big)\Big|_{\varepsilon=0} \tag{33}
$$

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

$$
\begin{aligned}
&\frac{\partial}{\partial \varepsilon} \bar{Q}_{k-1,\tau}^{a,\varepsilon}(a_{k-1}, l_{k-1}, t_{t-1}, d_{k-1}, f_{k-2})\Big((1-\varepsilon)P + \varepsilon\delta_{\mathcal{F}_{T_{(k-1)}}}\Big)\Big|_{\varepsilon=0} \\
&= \frac{\delta_{\mathcal{F}_{T_{(k-1)}}}(f_{k-1})}{P\Big(\mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)} \Bigg(\bar{Z}_{k,\tau}^{a} - \bar{Q}_{k-1,\tau}^{a}\Big(\tau, \mathcal{F}_{T_{(k-1)}}\Big) + \\
&+ \int_{T_{(k-1)}}^{\tau} \bar{Z}_{k,\tau}^{a}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{T_{(k-1)}}^{t_k} \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda_{k-1}^{x}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big)} \Big(N_{k}^{c}(\mathrm{d}s) - \tilde{Y}_{k-1}(s)\lambda_{k-1}^{c}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big) \\
&\qquad P_{(T_{(k)}, D_{(k)}, L(T_{(k)}), A(T_{(k)}))}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)\Bigg) \\
&+ \int_{T_{(k-1)}}^{\tau} \Bigg(\frac{I(t_k \le \tau, d_k \in \{a, \ell\})}{\exp\Big(-\int_{T_{(k-1)}}^{t_k} \lambda^{c}(s \mid f_{k-1})\mathrm{d}s\Big)} \cdot \left(\frac{\pi_{k-1}^{*}\Big(t_k, a_k, \mathcal{F}_{T_{(k-1)}}\Big)}{\pi_{k-1}\Big(t_k, a_k, \mathcal{F}_{T_{(k-1)}}\Big)}\right)^{I(d_k = a)} \frac{\partial}{\partial \varepsilon} \bar{Q}_{k-1,\tau}^{a,\varepsilon}(a_k, l_k, t_k, d_k, f_{k-1})\Big((1-\varepsilon)P + \varepsilon\delta_{\mathcal{F}_{T_{(k)}}}\Big)\Big|_{\varepsilon=0} \\
&\qquad P_{(T_{(k)}, D_{(k)}, L(T_{(k)}), A(T_{(k)}))}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)
\end{aligned} \tag{34}
$$

Now note for the second term, we can write

$$
\begin{aligned}
&\int_{T_{(k-1)}}^{\tau} \bar{Z}_{k,\tau}^{a}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) \int_{T_{(k-1)}}^{t_k} \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda_{k-1}^{x}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big)} \Big(N_{k}^{c}(\mathrm{d}s) - \tilde{Y}_{k-1}(s)\lambda_{k-1}^{c}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big) \\
&\qquad P_{(T_{(k)}, D_{(k)}, L(T_{(k)}), A(T_{(k)}))}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big)) \\
&= \int_{T_{(k-1)}}^{\tau} \int_{s}^{\tau} \bar{Z}_{k,\tau}^{a}(\tau, t_k, d_k, l_k, a_k, f_{k-1}) P_{(T_{(k)}, D_{(k)}, L(T_{(k)}), A(T_{(k)}))}\Big(\mathrm{d}t_k, \mathrm{d}d_k, \mathrm{d}l_k, \mathrm{d}a_k \mid \mathcal{F}_{T_{(k-1)}} = f_{k-1}\Big) \\
&\qquad \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda_{k-1}^{x}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big)} \Big(N_{k}^{c}(\mathrm{d}s) - \tilde{Y}_{k-1}(s)\lambda_{k-1}^{c}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big)) \\
&= \int_{T_{(k-1)}}^{\tau} \Big(\bar{Q}_{k-1,\tau}^{a}(\tau) - \bar{Q}_{k-1,\tau}^{a}(s)\Big) \\
&\qquad \frac{1}{\exp\Big(-\sum_{x=a,\ell,c,d,y} \int_{T_{(k-1)}}^{s} \lambda_{k-1}^{x}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big)} \Big(N_{k}^{c}(\mathrm{d}s) - \tilde{Y}_{k-1}(s)\lambda_{k-1}^{c}\Big(s, \mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\Big))
\end{aligned} \tag{35}
$$

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

For now, we recommend using the one step estimator and not the TMLE because the martingales are computationally intensive to estimate. This means that multiple TMLE updates may not be a good idea.

6.1 Remainder term with IPCW for $K = 1$:

Let $\varphi_{\tau,1}$ be the term of the efficient influence function without the martingale and let $\varphi_{\tau,2}$ be the martingale term. It is given by

$$\begin{aligned}\varphi^*_\tau(P) = &\frac{\pi^*(A(0) \mid L(0))}{\pi(A(0) \mid L(0))} \\ &\times \left[\frac{\mathbb{1}\left\{T_{(1)} \le \tau, D_{(1)} = y\right\}}{\exp\left(-\Lambda^c\left(T_{(1)} \mid A_0, L_0\right)\right)} - F_1(\tau \mid A_0, L_0)\right. \\ &\left.+ \int_0^\tau (F_1(\tau \mid A_0, L_0) - F_1(s \mid A_0, L_0)) \frac{1}{\exp(-(\Lambda^y(s \mid A_0, L_0) + \Lambda^c(s \mid A_0, L_0)))}\left(N^c(\mathrm{d}s) - \tilde{Y}(s)\Lambda^c(ds \mid A(0), L(0))\right)\right] \\ &+ \int F_1(\tau|a, L(0))\pi^*(a \mid L(0))\nu_A(\mathrm{d}a) - \Psi_\tau(P)\end{aligned} \tag{36}$$

Taken the mean with respect to P_0 gives

$$\begin{aligned}\mathbb{E}_{P_0}[\varphi^*_\tau(P)] = &\mathbb{E}_{P_0}\left[\frac{\pi^*(A(0) \mid L(0))}{\pi(A(0) \mid L(0))}\right. \\ &\times \left[\frac{\mathbb{1}\left\{T_{(1)} \le \tau, D_{(1)} = y\right\}}{\exp\left(-\Lambda^c\left(T_{(1)} \mid A_0, L_0\right)\right)} - F_1(\tau \mid A_0, L_0)\right. \\ &\left.+ \int_0^\tau (F_1(\tau \mid A_0, L_0) - F_1(s \mid A_0, L_0)) \frac{1}{\exp(-(\Lambda^y(s \mid A_0, L_0) + \Lambda^c(s \mid A_0, L_0)))}\left(N^c(\mathrm{d}s) - \tilde{Y}(s)\Lambda^c(ds \mid A(0), L(0))\right)\right] \\ &\left.+ \int F_1(\tau|a, L(0))\pi^*(a \mid L(0))\nu_A(\mathrm{d}a) - \Psi_\tau(P)\right] \\ = &\mathbb{E}_{P_0}\left[\frac{\pi^*(A(0) \mid L(0))}{\pi(A(0) \mid L(0))}\right. \\ &\times \left[\left(\int_0^\tau \Lambda^y_0(ds \mid A_0, L(0))(\exp(-\Lambda^y_0(s \mid A_0, L(0)) - \Lambda^c_0(s| A_0, L(0))))\left(\frac{1}{\exp(-\Lambda^c(s \mid A_0, L_0))} - \frac{1}{\exp(-\Lambda^c_0(s \mid A_0, L_0))}\right)\right.\right. \\ &\left.\left.+ \int_0^\tau (F_1(\tau \mid A_0, L_0) - F_1(s \mid A_0, L_0)) \frac{\exp(-(\Lambda^y_0(s \mid A_0, L_0) + \Lambda^c_0(s \mid A_0, L_0)))}{\exp(-(\Lambda^y(s \mid A_0, L_0) + \Lambda^c(s \mid A_0, L_0)))}(\Lambda^c_0(\mathrm{d}s \mid A(0), L(0))) - \Lambda^c(ds \mid A(0), L(0))\right)\right] \\ &\left.+ \int F_1(\tau|a, L(0))\pi^*(a \mid L(0))\nu_A(\mathrm{d}a) - \Psi_\tau(P)\right]\end{aligned} \tag{37}$$

and the two parameters are $\Psi_{\tau(P)} = \mathbb{E}_P\left[\mathbb{E}_P \frac{I(T \le \tau, D=y)}{\exp(-\Lambda)}\right)]$.

Let us try to calculate this remainder for a simple static intervention.

$$\begin{aligned}R_2(P, P_0) &= \Psi_\tau(P) - \Psi_\tau(P_0) + P_0(\varphi^*_\tau(P)) \\ &= \Psi_\tau(P) - \Psi_\tau(P, S^c_0) + P_0\left(\varphi^*_{\tau,1}(P)\right) - P_0\left(\varphi^*_{\tau,1}(P, S^c_0)\right) \\ &+\Psi_\tau(P, S^c_0) - \Psi_\tau(P_0) + P_0\left(\varphi^*_{\tau,1}(P, S^c_0)\right) \\ &+P_0\left(\varphi^*_{\tau,2}(P)\right) - P_0\left(\varphi^*_{\tau,2}(P, S^c_0)\right) + P_0\left(\varphi^*_{\tau,2}(P, S^c_0)\right)\end{aligned} \tag{38}$$

The above calculation shows $P_0\left(\varphi^*_{\tau,2}(P, S^c_0)\right) = 0$. The martingale difference has the desired product structure (almost we're missing like a 1 in the censoring survival functino quotient). We can write for one of the terms in the $P_0\left(\varphi^*_{\tau,1}(P)\right) - P_0\left(\varphi^*_{\tau,1}(P, S^c_0)\right)$:

$$P_0\left[\mathbb{E}_P\left[\frac{I(T \le \tau, D = 1)}{S^c(T| 1, L)} - \frac{I(T \le \tau, D = 1)}{S^c_0(T| 1, L)} \mid A_0 = 1, L\right]\right] \tag{39}$$

For the other term

$$P_0\left[\frac{\mathbb{1}\{A = 1\}}{\pi(A, L)}\mathbb{E}_P\left[\frac{I(T \le \tau, D = 1)}{S^c(T| 1, L)} - \frac{I(T \le \tau, D = 1)}{S^c_0(T| 1, L)} \mid A_0 = 1, L\right]\right] \tag{40}$$

Add these two terms to get

For the other term in the $\varphi_{\tau,1}$, we have $P_0(\varphi^*_{\tau,1}(P,S^c_0))=0$. Apply the well-known remainder when the censoring is known to the second term to get

$$(2)=\int \frac{\pi(1\mid L)-\pi_0(1|L)}{\pi(1|L)}\left(\mathbb{E}_P\left[\frac{I(T\leq \tau, D=1)}{S^c_0(T|\,1,L)}\mid A_0=1,L\right]-\mathbb{E}_{P_0}\left[\frac{I(T\leq \tau, D=1)}{S^c_0(T|\,1,L)}\mid A_0=1,L\right]\right)dP_0(L) \tag{41}$$

We get a term like (2) for the other part.

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

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

$$\begin{aligned}
\varphi^*_\tau(P)=\sum_{k=1}^K\prod_{j=0}^{k-1}&\left(\frac{\pi^*_{j-1}\left(T_{(j)},A\left(T_{(j)}\right),\mathcal{F}_{T_{(j-1)}}\right)}{\pi_{j-1}\left(T_{(j)},A\left(T_{(j)}\right),\mathcal{F}_{T_{(j-1)}}\right)}\right)^{I\left(D_{(j)}=a\right)}\frac{I\left(D_{(k-1)}\in\{\ell,a\},T_{(k-1)}\leq\tau\right)}{\exp\left(-\sum_{1\leq j<k}\int_{T_{(j-1)}}^{T_{(j)}}\lambda^c_{j-1}\left(s,\mathcal{F}_{T_{(j-1)}}\right)\mathrm{d}s\right)}\Bigg[\\
&\int_{T_{(k-1)}}^\tau \frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}\left(\left(\int_{\mathcal{A}}\bar{Q}^a_{k,\tau}\left(L\left(T_{(k-1)}\right),a_k,s,a,\mathcal{F}_{T_{(k)}}\right)\pi^*_k\left(s,a_k,\mathcal{F}_{T_{(k)}}\right)\nu_A(\mathrm{d}a_k)-B_{k-1}(u)\right)M^a_k(du)\\
&+\int_{T_{(k-1)}}^\tau\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}\left(\mathbb{E}_P\left[\bar{Q}^a_{k,\tau}\left(L\left(T_{(k)}\right),A\left(T_{(k-1)}\right),T_{(k)},D_{(k)},\mathcal{F}_{T_{(k-1)}}\right)\mid T_{(k)}=s,D_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\right]-B_{k-1}(u)\right)M^\ell_k(du)\\
&+\int_{T_{(k-1)}}^\tau\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}(1-B_{k-1}(u))M^y_k(du)+\int_{T_{(k-1)}}^\tau\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}(0-B_{k-1}(u))M^d_k(du)\\
&+\frac{1}{S^c\left(T_{(k)}\mid\mathcal{F}_{T_{(k-1)}}\right)}I\left(T_{(k)}\leq\tau,D_{(k)}=\ell,k<K\right)\left(\bar{Q}^a_{k,\tau}\left(L\left(T_{(k)}\right),A\left(T_{(k-1)}\right),T_{(k)},\ell,\mathcal{F}_{T_{(k-1)}}\right)\right.\\
&\qquad\left.-\mathbb{E}_P\left[\bar{Q}^a_{k-1,\tau}\left(L\left(T_{(k)}\right),A\left(T_{(k-1)}\right),\widetilde{T_{(k)}},D_{(k)},\mathcal{F}_{T_{(k-1)}}\right)\mid\widetilde{T_{(k)}}=T_{(k)},D_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\right]\right)\Bigg]\\
&+\int\bar{Q}^a_{1,\tau}(a,L_0)\pi^*_0(0,a,\mathcal{F}_0)\nu_A(\mathrm{da})-\Psi_\tau(P)
\end{aligned}\tag{42}$$

We find immediately that

$$\begin{aligned}
\varphi^*_\tau(P)=\sum_{k=1}^K\prod_{j=0}^{k-1}&\left(\frac{\pi^*_{j-1}\left(T_{(j)},A\left(T_{(j)}\right),\mathcal{F}_{T_{(j-1)}}\right)}{\pi_{j-1}\left(T_{(j)},A\left(T_{(j)}\right),\mathcal{F}_{T_{(j-1)}}\right)}\right)^{I\left(D_{(j)}=a\right)}\frac{I\left(D_{(k-1)}\in\{\ell,a\},T_{(k-1)}\leq\tau\right)}{\exp\left(-\sum_{1\leq j<k}\int_{T_{(j-1)}}^{T_{(j)}}\lambda^c_{j-1}\left(s,\mathcal{F}_{T_{(j-1)}}\right)\mathrm{d}s\right)}\Bigg[\\
&-\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}\left(\int_{\mathcal{A}}\bar{Q}^a_{k,\tau}\left(L\left(T_{(k-1)}\right),a_k,s,a,\mathcal{F}_{T_{(k)}}\right)\pi^*_k\left(s,a_k,\mathcal{F}_{T_{(k)}}\right)\nu_A(\mathrm{da}_k)\right)\Lambda^a_k(du)\\
&-\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}\left(\mathbb{E}_P\left[\bar{Q}^a_{k,\tau}\left(L\left(T_{(k)}\right),A\left(T_{(k-1)}\right),T_{(k)},D_{(k)},\mathcal{F}_{T_{(k-1)}}\right)\mid T_{(k)}=s,D_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\right]\right)\Lambda^\ell_k(du)\\
&-\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}(1)\Lambda^y_k(du)-\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}(0)\Lambda^d_k(du)\\
&-\int_{T_{(k-1)}}^{\tau}\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}B_{k-1}(u)M^\bullet_k(\mathrm{d}u)\\
&+\bar{Z}_{k,\tau}\left(T_{(k)},D_{(k)},L\left(T_{(k)}\right),A\left(T_{(k)}\right),\mathcal{F}_{T_{(k-1)}}\right)+\int_{T_{(k-1)}}^\tau\frac{1}{S^c\left(u\mid\mathcal{F}_{T_{(k-1)}}\right)}B_{k-1}(u)M^c_k(du)\Bigg]\\
&+\int\bar{Q}^a_{1,\tau}(a,L_0)\pi^*_0(0,a,\mathcal{F}_0)\nu_A(\mathrm{da})-\Psi_\tau(P)
\end{aligned}\tag{43}$$

Now note that

$$
\begin{aligned}
&\int_{T_{(k-1)}}^{\tau}\left(\bar{Q}^a_{k-1,\tau}(\tau)-\bar{Q}^a_{k-1,\tau}(u)\right)\frac{1}{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}\left(N^\bullet_k(\mathrm{d}s)-\tilde{Y}_{k-1}(s)\lambda^\bullet_{k-1}\Big(s,\mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\right)\\
&=\left(\bar{Q}^a_{k-1,\tau}(\tau)-\bar{Q}^a_{k-1,\tau}\big(T_{(k)}\big)\right)\frac{1}{S^c\Big(T_{(k)}\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(T_{(k)}\mid\mathcal{F}_{T_{(k-1)}}\Big)}\mathbb{1}\Big\{T_{(k)}\le\tau\Big\}\\
&-\bar{Q}^a_{k-1,\tau}(\tau)\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{1}{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}\lambda^\bullet_{k-1}\Big(s,\mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\\
&+\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{\bar{Q}^a_{k-1,\tau}(u)}{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}\lambda^\bullet_{k-1}\Big(s,\mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s
\end{aligned}
\tag{44}
$$

Let us calculate the second integral

$$
\begin{aligned}
&\bar{Q}^a_{k-1,\tau}(\tau)\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{1}{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}\lambda^\bullet_{k-1}\Big(s,\mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\\
&=\bar{Q}^a_{k-1,\tau}(\tau)\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}{\Big(S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)\Big)^2}\lambda^\bullet_{k-1}\Big(s,\mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}s\\
&=\bar{Q}^a_{k-1,\tau}(\tau)\left(1-\frac{1}{S^c\Big(T_{(k)}\wedge\tau\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(T_{(k)}\wedge\tau\mid\mathcal{F}_{T_{(k-1)}}\Big)}\right)
\end{aligned}
\tag{45}
$$

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

$$
\begin{aligned}
&\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\frac{\bar{Q}^a_{k+1,\tau}(u)}{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}\lambda^\bullet_{k-1}\Big(u,\mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}u\\
&=\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\Bigg[\int_0^u S\Big(s\mid\mathcal{F}_{T_{(k-1)}}\Big)\Lambda^a_{k-1}\Big(ds,\mathcal{F}_{T_{(k-1)}}\Big)\\
&\qquad\times\left(\int_{\mathcal{A}}\bar{Q}^a_{k+1,\tau}\Big(L\big(T_{(k-1)}\big),a_k,s,a,\mathcal{F}_{T_{(k-1)}}\Big)\pi^*_k\Big(s,a_k,\mathcal{F}_{T_{(k)}}\Big)\nu_A(\mathrm{d}a_k)\right)\\
&+\int_0^u S\Big(s\mid\mathcal{F}_{T_{(k-1)}}\Big)\Lambda^\ell_{k-1}\Big(ds,\mathcal{F}_{T_{(k-1)}}\Big)\\
&\qquad\times\left(\mathbb{E}_P\Big[\bar{Q}^a_{k+1,\tau}\Big(L\big(T_{(k)}\big),A\big(T_{(k-1)}\big),T_{(k)},D_{(k)},\mathcal{F}_{T_{(k-1)}}\Big)\mid T_{(k)}=s,D_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\Big]\right)\\
&+\int_0^u S\Big(s\mid\mathcal{F}_{T_{(k-1)}}\Big)\Lambda^y_{k-1}\Big(ds,\mathcal{F}_{T_{(k-1)}}\Big)\Bigg]\\
&\times\frac{1}{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}\lambda^\bullet_{k-1}\Big(u,\mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}u\\
&=\int_{T_{(k-1)}}^{\tau\wedge T_{(k)}}\int_s^{\tau\wedge T_{(k)}}\frac{1}{S^c\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)S\Big(u\mid\mathcal{F}_{T_{(k-1)}}\Big)}\lambda^\bullet_{k-1}\Big(u,\mathcal{F}_{T_{(k-1)}}\Big)\mathrm{d}u\Bigg[S\Big(s\mid\mathcal{F}_{T_{(k-1)}}\Big)\Lambda^a_{k-1}\Big(ds,\mathcal{F}_{T_{(k-1)}}\Big)\\
&\qquad\times\left(\int_{\mathcal{A}}\bar{Q}^a_{k+1,\tau}\Big(L\big(T_{(k-1)}\big),a_k,s,a,\mathcal{F}_{T_{(k-1)}}\Big)\pi^*_k\Big(s,a_k,\mathcal{F}_{T_{(k)}}\Big)\nu_A(\mathrm{d}a_k)\right)\\
&+S\Big(s\mid\mathcal{F}_{T_{(k-1)}}\Big)\Lambda^\ell_{k-1}\Big(ds,\mathcal{F}_{T_{(k-1)}}\Big)\\
&\qquad\times\left(\mathbb{E}_P\Big[\bar{Q}^a_{k+1,\tau}\Big(L\big(T_{(k)}\big),A\big(T_{(k-1)}\big),T_{(k)},D_{(k)},\mathcal{F}_{T_{(k-1)}}\Big)\mid T_{(k)}=s,D_{(k)}=\ell,\mathcal{F}_{T_{(k-1)}}\Big]\right)\\
&+S\Big(s\mid\mathcal{F}_{T_{(k-1)}}\Big)\Lambda^y_{k-1}\Big(ds,\mathcal{F}_{T_{(k-1)}}\Big)\Bigg]
\end{aligned}
\tag{46}
$$

Now note that

$$
\begin{aligned}
&\int_s^{\tau \wedge T_{(k)}} \frac{1}{S^c\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(u \mid \mathcal{F}_{T_{(k-1)}}\Big)} \lambda^{\bullet}_{k-1}\Big(u, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}u \\
&= \frac{1}{S^c(s) S(s)} - \frac{1}{S^c\Big(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big)}
\end{aligned} \tag{47}
$$

Setting this into the previous integral, we get

$$
\begin{aligned}
&\int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda^{\bullet}_{k-1}\Big(u, \mathcal{F}_{T_{(k-1)}}\Big) \mathrm{d}u \Bigg[\Lambda^a_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \\
&\qquad\qquad \times \left(\int_{\mathcal{A}} \bar{Q}^a_{k+1,\tau}\Big(L\Big(T_{(k-1)}\Big), a_k, s, a, \mathcal{F}_{T_{(k-1)}}\Big) \pi^*_k\Big(s, a_k, \mathcal{F}_{T_{(k)}}\Big) \nu_A(\mathrm{d}a_k) \right) \\
&+ \Lambda^{\ell}_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \\
&\qquad\qquad \times \Big(\mathbb{E}_P\Big[\bar{Q}^a_{k+1,\tau}\Big(L\Big(T_{(k)}\Big), A\Big(T_{(k-1)}\Big), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\Big) \mid T_{(k)} = s, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \Big] \Big) \\
&+ \Lambda^{y}_{k-1}\Big(ds, \mathcal{F}_{T_{(k-1)}}\Big) \Bigg] \\
&- \frac{1}{S^c\Big(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big) S\Big(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\Big)} \bar{Q}^a_{k-1,\tau}\Big(\tau \wedge T_{(k)}\Big)
\end{aligned} \tag{48}
$$

Thus, we find

$$
\begin{aligned}
&\int_{T_{(k-1)}}^{\tau} \left(\bar{Q}^a_{k-1,\tau}(\tau) - \bar{Q}^a_{k-1,\tau}(u)\right) \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} \left(N^{\bullet}_k(\mathrm{d}s) - \tilde{Y}_{k-1}(s) \lambda^{\bullet}_{k-1}\left(s, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}s\right) \\
&= \left(\bar{Q}^a_{k-1,\tau}(\tau) - \bar{Q}^a_{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}^a_{k-1,\tau}(\tau) \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} \lambda^{\bullet}_{k-1}\left(s, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}s \\
&+ \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{\bar{Q}^a_{k-1,\tau}(u)}{S^c\left(u \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(u \mid \mathcal{F}_{T_{(k-1)}}\right)} \lambda^{\bullet}_{k-1}\left(s, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}s \\
&= \left(\bar{Q}^a_{k-1,\tau}(\tau) - \bar{Q}^a_{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}^a_{k-1,\tau}(\tau) \left(1 - \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)\right) \\
&+ \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda^{\bullet}_{k-1}\left(u, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}u \Bigg[\Lambda^a_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \\
&\qquad\qquad \times \left(\int_{\mathcal{A}} \bar{Q}^a_{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)}}\right) \nu_A(\mathrm{d}a_k) \right) \\
&+ \Lambda^{\ell}_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \\
&\qquad\qquad \times \left(\mathbb{E}_P\left[\bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \\
&+ \Lambda^{y}_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \Bigg] \\
&- \frac{1}{S^c\left(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\right) S\left(\tau \wedge T_{(k)} \mid \mathcal{F}_{T_{(k-1)}}\right)} \bar{Q}^a_{k-1,\tau}\left(\tau \wedge T_{(k)}\right) \\
&= \left(\bar{Q}^a_{k-1,\tau}(\tau) - \bar{Q}^a_{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\} \\
&+ \frac{\bar{Q}^a_{k-1,\tau}(\tau) - \bar{Q}^a_{k-1,\tau}\left(\tau \wedge T_{(k)}\right)}{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)} \\
&+ \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda^{\bullet}_{k-1}\left(u, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}u \Bigg[\Lambda^a_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \\
&\qquad\qquad \times \left(\int_{\mathcal{A}} \bar{Q}^a_{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)}}\right) \nu_A(\mathrm{d}a_k) \right) \\
&+ \Lambda^{\ell}_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \\
&\qquad\qquad \times \left(\mathbb{E}_P\left[\bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \\
&+ \Lambda^{y}_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \Bigg] - \bar{Q}^a_{k-1,\tau}(\tau) \\
&\overset{?}{=} \int_{T_{(k-1)}}^{\tau \wedge T_{(k)}} \frac{1}{S^c(s)} \lambda^{\bullet}_{k-1}\left(u, \mathcal{F}_{T_{(k-1)}}\right) \mathrm{d}u \Bigg[\Lambda^a_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \\
&\qquad\qquad \times \left(\int_{\mathcal{A}} \bar{Q}^a_{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)}}\right) \nu_A(\mathrm{d}a_k) \right) \\
&+ \Lambda^{\ell}_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \\
&\qquad\qquad \times \left(\mathbb{E}_P\left[\bar{Q}^a_{k+1,\tau}\left(L\left(T_{(k)}\right), A\left(T_{(k-1)}\right), T_{(k)}, D_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \mid T_{(k)} = s, D_{(k)} = \ell, \mathcal{F}_{T_{(k-1)}} \right] \right) \\
&+ \Lambda^{y}_{k-1}\left(ds, \mathcal{F}_{T_{(k-1)}}\right) \Bigg] - \bar{Q}^a_{k-1,\tau}(\tau)
\end{aligned} \tag{49}
$$

Inserting this into the EIF gives the desired result in terms of our EIF.

7 Data-adaptive choice of K

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

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

$$(\mathbb{P}_n - P)(\varphi_\tau^*(P)(K_n) - \varphi_\tau^*(P)(K)) \tag{50}$$

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

8 Issues relating to rare patient histories (postponed)

Consider the following table representing example data:

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

8.1 Pooling

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

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

To estimate the regression for the time-varying covariates, one could do:

- Not intervene on the last two or three time points, letting certain parts of the data generating mechanism be observational, that is $\pi_j^*\left(t, \cdot, \mathcal{F}_{T_{(j)}}\right) = \pi_j\left(t, \cdot, \mathcal{F}_{T_{(j)}}\right)$ for $j = 4, 5$.
- Another is to make a Markov-like assumption in the interventional world, i.e.,

$$\mathbb{E}_Q\left[\sum_{j=1}^{3} I\left(T_{(j)} \leq \tau, D_{(j)} = y\right) \mid \mathcal{F}_0\right] = \mathbb{E}_Q\left[\sum_{j=6}^{8} I\left(T_{(j)} \leq \tau, D_{(j)} = y\right) \mid \mathcal{F}_{T_{(5)}}\right] \tag{51}$$

So we separately estimate the target parameter on the left hand side and use it to estimate the one on the right when we need to, pooling the data from the last three events with the data from the first three events.

Doing this adaptively leads to data-adaptive target parameter (Hubbard et al., 2016).

Other possible methods are:

- Use an estimation procedure that is similar to Shirakawa et al. (2024) or use hazards which are estimated all at once.
- Bayesian methods may be useful since they do not have issues with finite sample size. They are also a natural way of dealing with the missing data problem. However, nonparametric Bayesian methods are not (yet) able to deal with a large number of covariates.

8.2 Other ideas

Some other issues are that the covariates are (fairly) high dimensional. This may yield issues with regression-based methods.

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

8.3 Topics for further research

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

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

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

Bibliography

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

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

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

Gill, R. D., & Robins, J. M. (2023,). *Causal Inference for Complex Longitudinal Data: The Continuous Time g-Computation Formula*. https://arxiv.org/abs/math/0409436

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

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

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

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

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

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

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

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

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

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

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

van der Laan, M. J., & Rubin, D. (2006). Targeted Maximum Likelihood Learning. *The International Journal of Biostatistics*, *2*(1). https://doi.org/10.2202/1557-4679.1043

Weiss, J. C., & Page, D. (2013). Forest-Based Point Process for Event Prediction from Electronic Health Records. In H. Blockeel, K. Kersting, S. Nijssen, & F. Železný (Eds.), *Machine Learning and Knowledge Discovery in Databases: Machine Learning and Knowledge Discovery in Databases*. https://doi.org/10.1007/978-3-642-40994-3_35