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

Johan Sebastian Ohlendorff
University of Copenhagen

Anders Munch
University of Copenhagen

Thomas Alexander Gerds
University of Copenhagen

Abstract

?

1 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 $k = 1, ..., K-1$ 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,

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

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

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

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

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

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

We now present a simple iterated representation of the data target parameter $\Psi_\tau^g(P)$ in the case with no censoring. To do so, define

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

where $\prod_{s\in(0,t]}$ is the product integral over the interval $(0, t]$ (Gill & Johansen (1990)).

Our idea builds upon the works of Bang & Robins (2005) (in discrete time) and Rytgaard et al. (2022) (in continuous time), who suggest the use of so-called *iterated regressions* of the target parameter. where we also deal with right-censoring.

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

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

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

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

Furthermore,

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

where,

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

2 Algorithm for ICE-IPCW Estimator

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

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

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

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

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

     Let $\mathcal{F}^g_{\bar{T}_{(k)}}$ denote the history with $A(\bar{T}_k) = 1$ if $\bar{\Delta}_{(k)} = a$. Otherwise, $\mathcal{F}^g_{\bar{T}_{(k)}} = \bar{\mathcal{F}}_{\bar{T}_{(k)}}$. Then calculate,
     $$\hat{Z}_k^a = \frac{\mathbb{1}\left\{\bar{T}_{(k)} < \tau, \bar{\Delta}_{(k)} \in \{a, \ell\}\right\} \hat{\nu}_k\left(\mathcal{F}^g_{\bar{T}_{(k)}}\right)}{\hat{S}^c\left(\bar{T}_{(k)} - \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)} + \widehat{m}_k.$$
     - If $k = K+1$, put
     $$\hat{Z}_k^a = \widehat{m}_k.$$
  4. Regress $\hat{Z}_k^a$ on $\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}$ with model $\mathcal{L}_o$ on the data with $\bar{T}_{(k-1)} < \tau$ and $\bar{\Delta}_{(k-1)} \in \{a, \ell\}$ to obtain a prediction function $\hat{\nu}_{k-1} : \mathcal{H}_{k-1} \to \mathbb{R}^+$. Here we denote by $\mathcal{H}_{k-1}$ the set of possible histories of the process up to and including the $k-1$'th event.
- At baseline, we obtain the estimate $\hat{\Psi}_n = \frac{1}{n} \sum_{i=1}^n \hat{\nu}_0(1, L_i(0))$.

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

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

We also need to discuss what models should be used for $\bar{Q}^g_{k,\tau}$. Note that
$$\mathbb{1}\left\{T_{(k)} < \tau \wedge \Delta_{(k)} \in \{a, \ell\}\right\} \bar{Q}^g_{k,\tau} = \mathbb{E}_{P^{G^*}}\left[N^y(\tau) \mid \mathcal{F}_{T_{(k)}}\right] \mathbb{1}\left\{T_{(k)} < \tau \wedge \Delta_{(k)} \in \{a, \ell\}\right\}$$
We see thus see that we should use a model for $\bar{Q}^g_{k,\tau}$ that is able to capture the counterfactual probability of the primary event occuring at or before time τ given the history up to and including the k'th event (among the people who are at risk of the event before time τ after k events).

2.1 Example usage of the Algorithm

To help illustrate the algorithm, we present a simple example in Table 1 in the case where $\tau = 5$. Since $K = 2$ in Table 1, we start at $k = 3$.

Iteration $k = 3$

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

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

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

Iteration $k = 2$

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

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

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

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

Iteration $k = 1$

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

Iteration $k = 0$

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

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

Table 1: Example data for illustration of the ICE-IPCW algorithm.

2.2 One-step ICE-IPCW estimator

We found that

$$\begin{aligned}
\varphi_\tau^*(P) = &\frac{\mathbb{1}\{A(0)=1\}}{\pi_0(L(0))} \sum_{k=1}^K \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{A(\bar{T}_j)=1\}}{\pi_j\left(\bar{T}_{(j)}, L(\bar{T}_j), \bar{\mathcal{F}}_{\bar{T}_{(j-1)}}\right)} \right)^{\mathbb{1}\{\bar{\Delta}_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} \tilde{S}^c\left(\bar{T}_{(j)} - \mid \bar{\mathcal{F}}_{\bar{T}_{(j-1)}}\right)} \\
&\times \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \in \{\ell, a\}, \bar{T}_{(k-1)} < \tau\right\} \Bigg(\left(\bar{Z}^a_{k,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(\tau)\right) \\
&\quad + \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \left(\bar{Q}^g_{k-1,\tau}(\tau) - \bar{Q}^g_{k-1,\tau}(u)\right) \frac{1}{\tilde{S}^c\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) S\left(u - \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)} \widetilde{M}^c(\mathrm{d}u) \Bigg) \\
&+ \bar{Q}^g_{0,\tau}(\tau) - \Psi^g_\tau(P),
\end{aligned} \tag{6}$$

is the efficient influence function.

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

$$\begin{aligned}
\varphi_\tau^*(O;\eta) = &\frac{\mathbb{1}\{A(0)=1\}}{\hat{\pi}_0(L(0))} \sum_{k=1}^K \prod_{j=1}^{k-1} \left(\frac{\mathbb{1}\{A(\bar{T}_j)=1\}}{\pi_j\left(\bar{T}_{(j)}, \bar{\mathcal{F}}_{\bar{T}_{(j-1)}}\right)} \right)^{\mathbb{1}\{\bar{\Delta}_{(j)}=a\}} \frac{1}{\prod_{j=1}^{k-1} \hat{S}^c\left(\bar{T}_{(j)} - \mid \bar{\mathcal{F}}_{\bar{T}_{(j-1)}}\right)} \mathbb{1}\left\{\bar{\Delta}_{(k-1)} \in \{\ell, a\}, \bar{T}_{(k-1)} < \tau\right\} \\
&\times \Bigg(\left(\bar{Z}^a_{k,\tau}\left(\hat{S}^c_{k-1}, \nu_{k,\tau}\right) - \nu_{k-1,\tau}\right) \\
&\quad + \int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \left(\mu_{k-1}\left(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) - \mu_{k-1}\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right) \frac{1}{\hat{S}^c\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \hat{S}\left(u - \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)} \left(\widetilde{N}^c(\mathrm{d}u) - \tilde{\Lambda}^c\left(\mathrm{d}u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right) \Bigg) \\
&+ \nu_{0,\tau}(1, L(0)) - \hat{P}_{L(0)}\left[\nu_{0,\tau}(1, \cdot)\right]
\end{aligned} \tag{7}$$

where

$$\mu_k\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right) = \int_{\bar{T}_{(k)}}^{u} \prod_{s\in\left(\bar{T}_{(k)},u\right)} \left(1 - \sum_{x=a,l,d,y} \hat{\Lambda}_k^x\left(\mathrm{d}s \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right)\right) \tag{8}$$
$$\times \left[\hat{\Lambda}_{k+1}^y\left(\mathrm{d}s \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right) + \mathbb{1}\{s<u\}\tilde{\nu}_{k+1,\tau}\left(1, s, a, \mathcal{F}_{T_{(k)}}\right)\hat{\Lambda}_{k+1}^a\left(\mathrm{d}s \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right) + \mathbb{1}\{s<u\}\tilde{\nu}_{k+1,\tau}\left(A\left(T_{(k-1)}\right), s, \ell, \mathcal{F}_{T_{(k)}}\right)\hat{\Lambda}_{k+1}^\ell\left(\mathrm{d}s \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right)\right].$$

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

We will now describe how to estimate the efficient influence function in practice. Overall, we consider the same procedure as in Section 2 with additional steps:

1. For $\{\nu_{k,\tau}(f_k)\}$, use the procedure described in Section 2.
2. For $\{\tilde{\nu}_{k,\tau}(f_k)\}$ use a completely similar procedure to the one given in Section 2 using the estimator $\nu_{k+1,\tau}$ to obtain $\tilde{\nu}_{k,\tau}$. Now we do not include the latest time varying covariate $L(\bar{T}_k)$ in the regression, so that $\tilde{\nu}_{k-1,\tau} = \mathbb{E}_{\hat{P}}\left[\bar{Z}_{k,\tau}^a\left(\hat{S}_{k-1}^c, \nu_{k,\tau}\right) \mid A(\bar{T}_k) = a_k, \bar{\Delta}_{(k)} = d_k, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}} = f_{k-1}\right]$.
3. Find $\left\{\hat{\Lambda}_k^x\right\}$ for $x = a, \ell, d, y$ and $\left\{\hat{\Lambda}_k^c\right\}$ using step 1 in Section 2.
4. Obtain an estimator of the propensity score $\{\pi_k(t, f_{k-1})\}$ by regressing $A(\bar{T}_k)$ on $\left(\bar{T}_{(k)}, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)$ among subjects with $\bar{\Delta}_{(k)} = a$ and $\bar{\Delta}_{(k-1)} \in \{a, \ell\}$ for each k and for $k = 0$ estimate $\pi_0(L(0))$ by regressing $A(0)$ on $L(0)$.
5. Use the estimates $\tilde{\nu}_{k,\tau}(f_k)$ and the estimates of $\hat{\Lambda}_k^x, x = a, \ell, d, y$ to numerically compute μ_{k-1} via Equation 8.
6. Use the estimated survival functions from the cumulative hazards in step 3 to compute the martingale term in Equation 7. See also Section 2.3 for details on how to approximately compute the censoring martingale term.
7. Substitute the rest of the estimates into Equation 7 and obtain the estimate of the efficient influence function.

There are multiple computational aspects of the stated procedure that should be addressed. First note that $\bar{Q}_{k,\tau}^g(\tau)$ is estimated twice. This redundancy is intentional: it ensures both the computability of the terms involved in the censoring martingale and that we can use $\nu_{k,\tau}$ required for subsequent iterations of the algorithm (avoiding the high dimensionality of the integrals as discussed in Section 1).

Our decision to use $\nu_{k,\tau}$ rather than $\mu_{k,\tau}$ as an estimator for the regression terms $(\bar{Z}_{k,\tau}^a - \bar{Q}_{k-1,\tau}^g)$ in Equation 7 is motivated by practical considerations. In particular, numerical integration may yield less accurate results due to the need to compute Λ_k^x for $x = a, \ell, d, y$. In practice, the contribution of the censoring martingale to the efficient influence function is typically small. As such, a simplified procedure that excludes the censoring martingale term or one that computes the censoring martingale term only at a sparse grid of time points may offer substantial computational efficacy with minimal bias.

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

Moreover, the resulting estimators are not guaranteed to be monotone in u which $\bar{Q}_{k,\tau}^g\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right)$ is. Applying flexible machine learning estimator may yet be possible if we apply a method that can handle multivariate, potentially high-dimensional, outcomes, such as neural networks. Note also

$$\frac{\bar{Q}_{k,\tau}^g\left(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) - \bar{Q}_{k,\tau}^g\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{S\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)} = S^c\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\mathbb{E}_P\left[Z_{k,\tau}^a\left(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) - Z_{k,\tau}^a\left(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right) \mid \bar{T}_{(k)} \geq t, \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right],$$

which actually means that regression can be applied to estimate every term in the efficient influence function, having estimated the cumulative cause-specific hazard functions for the censoring.

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

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

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

where

$$R_2(\eta,\eta') = P_\eta[\varphi_\tau^*(\cdot;\eta')] + \Psi_\tau^{\mathrm{obs}}(\eta') - \Psi_\tau^g(\eta)$$

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

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

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

$$R_2(\eta,\hat{\eta}) = o_P\left(n^{-\frac{1}{2}}\right), \tag{9}$$

and that the empirical process term fulfills

$$(\mathbb{P}_n - P)(\varphi_\tau^*(\cdot;\hat{\eta}) - \varphi_\tau^*(\cdot;\eta)) = o_P\left(n^{-\frac{1}{2}}\right). \tag{10}$$

We first discuss how to show Equation 10. This can be shown (Lemma 19.24 of van der Vaart (1998)) if
1. $\varphi_\tau^*(\cdot;\hat{\eta}) \in \mathcal{Z}$ for some P-Donsker class $\mathcal{Z}$ of functions with probability tending to 1, and
2. $\|\varphi_\tau^*(\cdot;\hat{\eta}) - \varphi_\tau^*(\cdot;\eta)\|_{L_P^2(O)} = o_P(1)$, with $\|f\|_{L_P^2(O)} = (\mathbb{E}_P[f(O)^2])^{\frac{1}{2}}$.

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

To obtain the rates in Equation 9, we find the second order remainder term $R_2(\eta_0,\eta)$ and show that it has a product structure (Theorem 2). This allows us to use estimators which need only convert at $L_2(P)$-rates of at least $o_P\left(n^{-\frac{1}{4}}\right)$ under regularity conditions.

Theorem 2 (Second order remainder): Let $\eta_0 = \left(\left\{\Lambda^x_{k,0}\right\}_{k,x}, \left\{\tilde{\Lambda}^c_{k,0}\right\}_k, \left\{\pi_{0,k}\right\}_k, \left\{\bar{Q}^g_{k,\tau}\right\}_k, \left\{\bar{Q}^{g,-L}_{k,\tau}\right\}_k, P_{0,L(0)}\right)$ be the true parameter values and let $\eta = \left(\left\{\Lambda^x_k\right\}_{k,x}, \left\{\tilde{\Lambda}^c_k\right\}_k, \left\{\pi_k\right\}_k, \left\{\nu_{k,\tau}\right\}_k, \left\{\tilde{\nu}_{k,\tau}\right\}_k, P'_{L(0)}\right)$. The remainder term $R_2(\eta_0, \eta)$ is given by

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

where

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

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

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

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

Proof: First define $\varphi^*_{k,\tau}(O; \eta)$ for $k > 0$ to be the k'th term in the efficient influence function given in Equation 6, and let $\varphi^*_{0,\tau}(O; \eta) = \nu_0(1, L(0)) - \Psi^{\text{obs}}_\tau(\eta)$, so that $\varphi^*_\tau(O; P) = \sum_{k=0}^{K} \varphi^*_k(O; P)$.

Then, we first note that

$$
\mathbb{E}_{P_0}\left[\varphi^*_{0,\tau}(O; \eta)\right] + \Psi^g_\tau(\eta) - \Psi^g_\tau(\eta_0) = \mathbb{E}_{P_0}\left[\nu_0(1, L(0)) - \bar{Q}^g_{0,\tau}(1, L(0))\right]. \tag{11}
$$

Apply the law of iterated expectation to the efficient influence function in Equation 6 to get

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

where

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

Now note that

$$
\begin{aligned}
&\mathbb{E}_{P_0}\Big[h_k\Big(\bar{\mathcal{F}}_{\bar{T}_{(k)}}\Big) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big] \\
&= \mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}\Big(\tilde{S}^c_0, \bar{Q}^g_{k,\tau}\Big)\Big) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big] - \nu_{k-1,\tau}\Big(\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) \\
&+\mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}\Big(\tilde{S}^c, \nu_k\Big) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big] - \mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}\Big(\tilde{S}^c, \bar{Q}^g_{k,\tau}\Big) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big] \\
&+\mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}\Big(\tilde{S}^c, \bar{Q}^g_{k,\tau}\Big) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big] - \mathbb{E}_{P_0}\Big[\bar{Z}^a_{k,\tau}\Big(\tilde{S}^c_0, \bar{Q}^g_{k,\tau}\Big) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big] \\
&+\mathbb{E}_{P_0}\left[\int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \Big(\mu_{k-1}\Big(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) - \mu_{k-1,\tau}\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)\Big) \frac{1}{\tilde{S}^c\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) S\Big(u- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)} \widetilde{M}^c(\mathrm{d}u) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]
\end{aligned} \tag{12}
$$

We shall need the following auxilliary result.

Lemma 1:

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

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

Proof: We first note that

$$
\begin{aligned}
&\mathbb{E}_{P_0}\left[\int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \Big(\mu_{k-1}\Big(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) - \mu_{k-1,\tau}\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)\Big) \frac{1}{\tilde{S}^c\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) S\Big(u- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)} \tilde{\Lambda}^c(\mathrm{d}u) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] \\
&= \mathbb{E}_{P_0}\left[\int_{\bar{T}_{(k-1)}}^{\tau \wedge \bar{T}_{(k)}} \Big(\mu_{k-1}\Big(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) - \mu_{k-1,\tau}\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)\Big) \frac{1}{\tilde{S}^c\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) S\Big(u- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)} \tilde{\Lambda}^c_k\Big(\mathrm{d}u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] \\
&= \mathbb{E}_{P_0}\left[\int_{\bar{T}_{(k-1)}}^{\tau} \mathbb{1}\Big\{\bar{T}_{(k)} \leq t\Big\}\right. \\
&\qquad \left. \times \Big(\mu_{k-1}\Big(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) - \mu_{k-1,\tau}\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)\Big) \frac{1}{\tilde{S}^c\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) S\Big(u- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)} \tilde{\Lambda}^c_k\Big(\mathrm{d}u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right] \\
&= \int_{\bar{T}_{(k-1)}}^{\tau} \mathbb{E}_{P_0}\Big[\mathbb{1}\Big\{\bar{T}_{(k)} \leq t\Big\} \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big] \Big(\mu_{k-1}\Big(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) - \mu_{k-1,\tau}\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)\Big) \frac{1}{\tilde{S}^c\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) S\Big(u- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)} \tilde{\Lambda}^c_k\Big(\mathrm{d}u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) \\
&= \int_{\bar{T}_{(k-1)}}^{\tau} \Big(\mu_{k-1}\Big(\tau \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) - \mu_{k-1,\tau}\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)\Big) \frac{\tilde{S}^c_0\Big(u- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) S_0\Big(u- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)}{\tilde{S}^c\Big(u \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big) S\Big(u- \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)} \tilde{\Lambda}^c_k\Big(du \mid \bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\Big)
\end{aligned} \tag{13}
$$

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

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

$$=\mathbb{E}_{P_0}\left[\mathbb{1}\{A\}\mathbb{1}\left\{\bar{T}_{(k-1)}<\tau\right\}\int_{\bar{T}_{(k-1)}}^{\tau\wedge\bar{T}_{(k)}}\left(\mu_{k-1}\left(\tau\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)\frac{1}{\tilde{S}^c\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\tilde{\Lambda}_0^c(\mathrm{d}t)\right]$$

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

from which we conclude that

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

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

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\left(\mu_{k-1}\left(\tau\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)\frac{\tilde{S}_0^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S_0\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\tilde{\Lambda}_{k,0}^c\left(du\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)$$

where the last equality follows from the same argument as in Equation 13. □

By an exchange of integrals, it follows that

$$\int_{\bar{T}_{(k-1)}}^{\tau}\left(\mu_{k-1}\left(\tau\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\mu_{k-1,\tau}\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)\frac{\tilde{S}_0^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S_0\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}_{k,0}^c\left(du\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}_k^c\left(du\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\int_u^{\tau}\mu_{k-1}\left(\mathrm{d}s\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\frac{\tilde{S}_0^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S_0\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}_{k,0}^c\left(du\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}_k^c\left(du\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\int_{(\bar{T}_{(k-1)},s)}\frac{\tilde{S}_0^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S_0\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)S\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}_{k,0}^c\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}_k^c\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\int_{(\bar{T}_{(k-1)},s)}\left(\frac{S_0\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{S\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}-1\right)\frac{\tilde{S}_0^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}_{k,0}^c\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}_k^c\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)$$

$$+\int_{\bar{T}_{(k-1)}}^{\tau}\int_{(\bar{T}_{(k-1)},s)}\frac{\tilde{S}_0^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}_{k,0}^c\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}_k^c\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)$$

$$=\int_{\bar{T}_{(k-1)}}^{\tau}\int_{(\bar{T}_{(k-1)},s)}\left(\frac{S_0\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{S\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}-1\right)\frac{\tilde{S}_0^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}_{k,0}^c\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)-\tilde{\Lambda}_k^c\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)$$

$$-\int_{\bar{T}_{(k-1)}}^{\tau}\left(\frac{\tilde{S}_0^c\left(s-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(s-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}-1\right)\mu_{k-1,\tau}\left(\mathrm{d}s\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)$$

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

$$\mathbb{E}_{P_0}\left[\bar{Z}_{k,\tau}^a\left(\tilde{S}^c,\bar{Q}_{k,\tau}^g\right)\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]-\mathbb{E}_{P_0}\left[\bar{Z}_{k,\tau}^a\left(\tilde{S}_0^c,\bar{Q}_{k,\tau}^g\right)\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right]$$
$$=\int_{\bar{T}_{(k-1)}}^{\tau}\left(\frac{\tilde{S}_0^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u-\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right)}-1\right)\bar{Q}_{k-1,\tau}^g\left(\mathrm{d}u\mid\bar{\mathcal{F}}_{\bar{T}_{(k-1)}}\right),$$

it follows from Equation 12 and Lemma 1 that

$$
\begin{aligned}
&\mathbb{E}_{P_0}\left[h_k\left(\mathcal{F}_{\bar{T}_{(k)}}\right) \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right] \\
&= \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c_0, \bar{Q}^g_{k,\tau}\right)\right) \mid \mathcal{F}_{\bar{T}_{(k-1)}}] - \nu_{k-1,\tau}\left(\mathcal{F}_{\bar{T}_{(k-1)}}\right) \\
&\quad + \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \nu_k\right) \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right] - \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right] \\
&\quad + \int_{\bar{T}_{(k-1)}}^{\tau} \left(\frac{\tilde{S}^c_0\left(u- \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(u- \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} - 1\right)\left(\bar{Q}^g_{k-1,\tau}\left(du \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) - \mu_{k-1,\tau}\left(du \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)\right) \\
&\quad + \int_{\bar{T}_{(k-1)}}^{\tau} \int_{(\bar{T}_{(k-1)},u)} \left(\frac{S_0\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)}{S\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)} - 1\right)\frac{\tilde{S}^c_0\left(s- \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)}{\tilde{S}^c\left(s \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)}\left(\tilde{\Lambda}^c_{k,0}\left(ds \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right) - \tilde{\Lambda}^c_k\left(ds \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)\right)\mu_{k-1,\tau}\left(du \mid \mathcal{F}_{\bar{T}_{(k-1)}}\right)
\end{aligned} \tag{14}
$$

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

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

we have that

$$
\begin{aligned}
&\int \mathbb{1}\{t_1 < \dots < t_k < \tau\}\mathbb{1}\{a_0 = \dots = a_k = 1\}\frac{\pi_{0,0}(l_0)}{\pi_0(l_0)}\prod_{j=1}^{k}\left(\frac{\pi_{0,j}(t_j, f_{j-1})}{\pi_j(t_j, f_{j-1})}\right)^{\mathbb{1}\{d_j=a\}}\frac{1}{\prod_{j=1}^{k}\tilde{S}^c(t_j- \mid f_{j-1})}\mathbb{1}\{d_1 \in \{\ell,a\},\dots,d_k \in \{\ell,a\}\} \\
&\qquad \times \left(\mathbb{E}_{P_0}\left[\bar{Z}^a_{k+1,\tau}\left(\tilde{S}^c_0, \bar{Q}^g_{k+1,\tau}\right) \mid \mathcal{F}_{\bar{T}_{(k)}} = f_k\right] - \nu_{k,\tau}(f_k)\right)P_{0,\mathcal{F}_{\bar{T}_{(k)}}}(df_k) \\
&+ \int \mathbb{1}\{t_1 < \dots < t_{k-1} < \tau\}\mathbb{1}\{a_0 = \dots = a_{k-1} = 1\}\frac{\pi_{0,0}(l_0)}{\pi_0(l_0)}\prod_{j=1}^{k-1}\left(\frac{\pi_{0,j}(t_k, f_{j-1})}{\pi_j(t_k, f_{j-1})}\right)^{\mathbb{1}\{d_j=a\}}\frac{1}{\prod_{j=1}^{k-1}\tilde{S}^c(t_j- \mid f_{j-1})}\mathbb{1}\{d_1 \in \{\ell,a\},\dots,d_{k-1} \in \{\ell,a\}\} \\
&\qquad \times \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \nu_k\right) \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right] - \mathbb{E}_{P_0}\left[\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \bar{Q}^g_{k,\tau}\right) \mid \mathcal{F}_{\bar{T}_{(k-1)}} = f_{k-1}\right]P_{0,\mathcal{F}_{\bar{T}_{(k-1)}}}(df_{k-1}) \\
&= \int \mathbb{1}\{t_1 < \dots < t_k < \tau\}\mathbb{1}\{a_0 = \dots = a_k = 1\}\frac{\pi_{0,0}(l_0)}{\pi_0(l_0)} \\
&\qquad \prod_{j=1}^{k-1}\left(\frac{\pi_{0,j}(t_j, f_{k-1})}{\pi_j(t_j, f_{j-1})}\right)^{\mathbb{1}\{d_k=a\}}\left(\left(\frac{\pi_{0,k}(k_j, f_{k-1})}{\pi_k(t_k, f_{k-1})}\right)^{\mathbb{1}\{d_k=a\}} - 1\right)\frac{1}{\prod_{j=1}^{k}\tilde{S}^c(t_j- \mid f_{j-1})}\mathbb{1}\{d_1 \in \{\ell,a\},\dots,d_k \in \{\ell,a\}\} \\
&\qquad \times \left(\bar{Q}^g_{k,\tau}(f_k) - \nu_{k,\tau}(f_k)\right)P_{0,\mathcal{F}_{\bar{T}_{(k)}}}(df_k).
\end{aligned} \tag{15}
$$

By combining Equation 11, Equation 14 and Equation 15, we are done.

2.3 Algorithm for the calculation of censoring martingale

In this subsection, we present an algorithm for computing the martingale term in Equation 7 along a specified time grid $\{t_1, \dots, t_m\}$ at iteration k of the influence function estimation procedure. In Steps 6, 8, 10, and 11 of the algorithm, we may use coarse approximations for the survival function and the associated integrals. For example, one may approximate the survival function using the exponential function or apply numerical integration techniques such as Simpson's rule to simplify computation. Note that we integrate over time on the interarrival scale. This means that we usually select $t_1 = 0$ and $t_m \leq \tau - \min_i \bar{T}_{k+1,i}$.

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

for $i = 1, ..., n$:
  $j_{\max} \leftarrow \max\left\{v \mid t_v \leq \tau - \bar{T}_{k,i}\right\}$
  $\hat{\nu}^y_\tau(0) \leftarrow \hat{\nu}^a_\tau(0) \leftarrow \hat{\nu}^\ell_\tau(0) \leftarrow t_0 \leftarrow \widehat{M}^c(0) \leftarrow 0$
  $\hat{S}_0 \leftarrow 1$
  for $j = 1, ..., j_{\max}$
    $\hat{S}(s-) \leftarrow \prod_{v \in [t_{j-1}, s)} \left(1 - \sum_{x=a,l,d,y} \hat{\Lambda}^x_{k+1}\left(\mathrm{d}v \mid \mathcal{F}_{\bar{T}_{k,i}}\right)\right)$
    $\hat{S}_j \leftarrow \hat{S}_{j-1} \cdot S(t_j)$
    $\hat{\nu}^y_\tau(t_j) \leftarrow \hat{\nu}^y_\tau(t_{j-1}) + \hat{S}_{j-1} \int_{(t_{j-1}, t_j]} \hat{S}(s-) \hat{\Lambda}^y_{k+1}\left(\mathrm{d}s \mid \mathcal{F}_{\bar{T}_{k,i}}\right)$
    if $k < K_\tau$:
      $\hat{\nu}^a_\tau(t_j) \leftarrow \hat{\nu}^a_\tau(t_{j-1}) + \hat{S}_{j-1} \int_{(t_{j-1}, t_j)} \hat{S}(s-) \tilde{\nu}_{k+1}\left(1, s + \bar{T}_{k+1,i}, a, \mathcal{F}_{\bar{T}_{k,i}}\right) \hat{\Lambda}^a_{k+1}\left(\mathrm{d}s \mid \mathcal{F}_{\bar{T}_{k,i}}\right)$
      $\hat{\nu}^\ell_\tau(t_j) \leftarrow \hat{\nu}^\ell_\tau(t_{j-1}) + \hat{S}_{j-1} \int_{(t_{j-1}, t_j)} \hat{S}(s-) \tilde{\nu}_{k+1}\left(A(\bar{T}_{k,i}), s + \bar{T}_{k+1,i}, \ell, \mathcal{F}_{\bar{T}_{k,i}}\right) \hat{\Lambda}^\ell_{k+1}\left(\mathrm{d}s \mid \mathcal{F}_{\bar{T}_{k,i}}\right)$
    else:
      $\hat{\nu}^a_\tau(t_j) \leftarrow \hat{\nu}^a_\tau(t_j) \leftarrow 0$
    $\hat{\nu}_\tau(t_j) \leftarrow \hat{\nu}^y_\tau(t_j) + \hat{\nu}^a_\tau(t_j) + \hat{\nu}^\ell_\tau(t_j)$
    $\widehat{M}^c(t_j) \leftarrow \mathbb{1}\left\{\bar{\Delta}_i = c, \bar{T}_{k+1,i} - \bar{T}_{k,i} \leq t_j\right\} - \hat{\Lambda}^c_{k+1}\left(t_j \mid \mathcal{F}_{\bar{T}_{k,i}}\right)$
    $\hat{S}^c(t_j) \leftarrow \prod_{v \in (0, t_j]} \left(1 - \hat{\Lambda}^c_{k+1}\left(\mathrm{d}v \mid \mathcal{F}_{\bar{T}_{k,i}}\right)\right)$
  $k_i \leftarrow \{v \mid t_v \leq \tau \wedge \bar{T}_{k+1,i} - \bar{T}_{k,i}\}$
  $\widehat{\mathrm{MG}}_i \leftarrow \sum_{j=1}^{k_i} \left(\hat{\nu}_\tau\left(t_{j_{\max}} \mid \mathcal{F}_{\bar{T}_{k,i}}\right) - \hat{\nu}_\tau\left(t_j \mid \mathcal{F}_{\bar{T}_{k,i}}\right)\right) \frac{1}{\hat{S}^c(t_j) \hat{S}_j} \left(\widehat{M}^c(t_j) - \widehat{M}^c(t_{j-1})\right)$
return $\widehat{\mathrm{MG}}$

3 Rates for the ICE-IPCW procedure

In this section, we discuss the rates of convergence for the ICE-IPCW estimator and variations thereof. Consider i.i.d. observations from $P_0 \in \mathcal{M}$ and a learner $\mathcal{L}_o$ that maps $\mathcal{L}_o : (O_i)_{i=1}^n \mapsto m_n$. Let $\mathcal{J}_{nk}$ be a collection of measurable functions $f_k : [0, \tau_{\text{end}}] \times \mathcal{F}_k \to [0, 1]$. Furthermore, let

$$\mathcal{A}_{ku} := \left\{h_u(O) : h_u \in L_2(P), \mathbb{E}_P\left[h_u \mid \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right] = \bar{Q}^g_{k+1,\tau}\left(u, \bar{\mathcal{F}}_{\bar{T}_{(k)}}\right)\right\}$$

for $u \leq \tau$. Consider the squared error loss $L : \mathcal{Z} \times \mathcal{F}_k$ where $L(m, f) := (m(f) - f)^2$.

We consider the least squares regression estimator given by

$$m_{k,n}(u) := \operatorname{argmin}_{f \in \mathcal{J}_{nk}} \frac{1}{n} \sum_{i=1}^n \left(\bar{Z}^a_{k,\tau}(u)\left(\hat{S}^c, m_{k+1,n}\right) - f\left(\bar{\mathcal{F}}_{\bar{T}_{(k)}}\right)\right)^2$$

x

Recall the definitions $\bar{Q}^g_{K,\tau} : (a_k, h_k) \mapsto 0$ and define recursively, for $k = K, ..., 0$,

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

and

$$\bar{Q}^g_{k,\tau} : (u, a_k, h_k) \mapsto \mathbb{E}_P\left[\bar{Z}^a_{k+1,\tau}(u) \mid A(\bar{T}_k) = a_k, \bar{H}_k = h_k\right], \quad u \le \tau$$

where $h_k = (a_k, l_k, t_k, d_k, ..., a_0, l_0)$. However, in practice, we need to estimate $\bar{Q}^g_{k,\tau}$ and $\tilde{S}^c$ since these are unknown quantities. We may denote these by $m_{k+1,n}$ and $\hat{S}^c$; whence estimates of the pseudo-outcomes $\bar{Z}^a_{k,\tau}\left(\hat{S}^c, m_{k+1,n}\right)$ can be obtained. After this we can apply our regression estimator of choice. Let $m_{k,n}$ be an estimator of $\bar{Q}^g_{k,\tau}$ obtained in this way and let $m^*_{k,n}$ denote the estimate obtained by regressing $\bar{Z}^a_{k,\tau}$ directly on the history $\bar{\mathcal{F}}_{\bar{T}_{(k)}}$. This is the oracle estimator in which we use the true values of $\tilde{S}^c$ and $\bar{Q}^g_{k+1,\tau}$. An elementary bound gives,

$$\int (\bar{Q}^g_{k,\tau}(u; f_k) - m_{k,n}(u, f_k)^2 P(\mathrm{d}f_k) \le 2 \int \left(\bar{Q}^g_{k,\tau}(u; f_k) - m^*_{k,n}(u, f_k)\right)^2 P(\mathrm{d}f_k)$$
$$+ 2 \int \left(m^*_{k,n}(u, f_k) - m_{k,n}(u, f_k)\right)^2 P(\mathrm{d}f_k)$$

The first term on the right-hand side can be known if one knows the convergence rate of the regression estimator used. However, the second term is evidently more difficult. To relax this problem, we consider specifically the case when the regression estimator is a least squares regression estimator, i.e., $m_{k,n}$ minimizes

$$m_{k,n} := \mathrm{argmin}_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n \left(\bar{Z}^a_{k,\tau}\left(\hat{S}^c, m_{k+1,n}\right) - f\left(\bar{\mathcal{F}}_{\bar{T}_{(k)}}\right)\right)^2$$

Similarly, $m^*_{k,n}$ minimizes

$$m^*_{k,n} := \mathrm{argmin}_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n \left(\bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \bar{Q}^g_{k+1,\tau}\right) - f\left(\bar{\mathcal{F}}_{\bar{T}_{(k)}}\right)\right)^2$$

The estimators are related in the following way.

$$m_{k,n} \le \frac{2}{n} \sum_{i=1}^n \left(\bar{Z}^a_{k,\tau}\left(\hat{S}^c, m_{k+1,n}\right) - \bar{Z}^a_{k,\tau}\left(\tilde{S}^c, \bar{Q}^g_{k+1,\tau}\right)\right)^2 + 2 m^*_{k,n}$$

We can write

$$\int \left(m^*_{k,n}(u, f_k) - m_{k,n}(u, f_k)\right)^2 P(\mathrm{d}f_k)$$
$$=$$

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

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

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

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

and that

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

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

One other direction is to use Bayesian methods. Bayesian methods may be particular useful for this problem since they do not have issues with finite sample size. They are also an excellent alternative to frequentist Monte Carlo methods for estimating the target parameter with Equation 5 because they offer uncertainty quantification directly through simulating the posterior distribution whereas frequentist simulation methods do not.

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

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

Bibliography

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

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

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

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

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

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

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

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

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

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

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

Rytgaard, H. C., Gerds, T. A., & van der Laan, M. J. (2022). Continuous-Time Targeted Minimum Loss-Based Estimation of Intervention-Specific Mean Outcomes. *The Annals of Stat Istics*, *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., Li, Y., & Laan, M. van der. (2023,). *Lassoed Tree Boosting*. https://arxiv.org/abs/2205.10697

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

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.

4 Appendix

4.1 Simulating the data

One classical causal inference perspective requires that we know how the data was generated up to unknown parameters (NPSEM) (Pearl, 2009). This approach has only to a small degree been discussed in continous-time causal inference literature (Røysland et al. (2024)). This is initially considered in the uncensored case, but is later extended to the censored case. For a moment, we ponder how the data was generated given a DAG. The DAG given in the section provides a useful tool for simulating continuous-time data, but not for drawing causal inference conclusions. For the event times, we can draw a figure representing the data generating mechanism which is shown in Figure 1. 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 2). 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 1, 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).

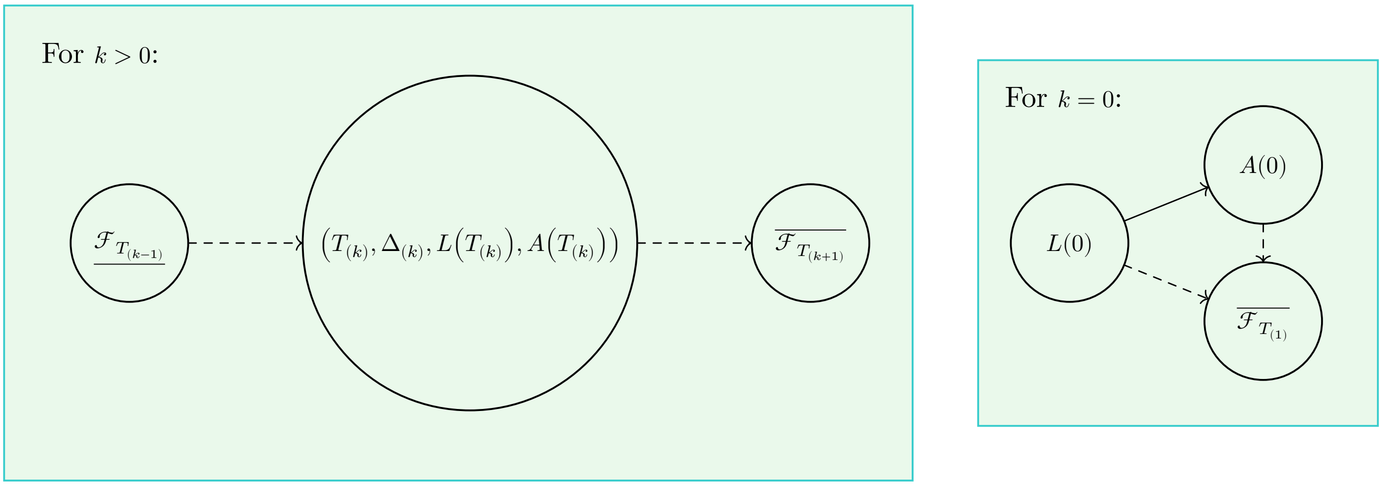

Figure 1: A DAG representing the relationships between the variables of O. The dashed lines indicate multiple edges from the dependencies in the past and into the future.

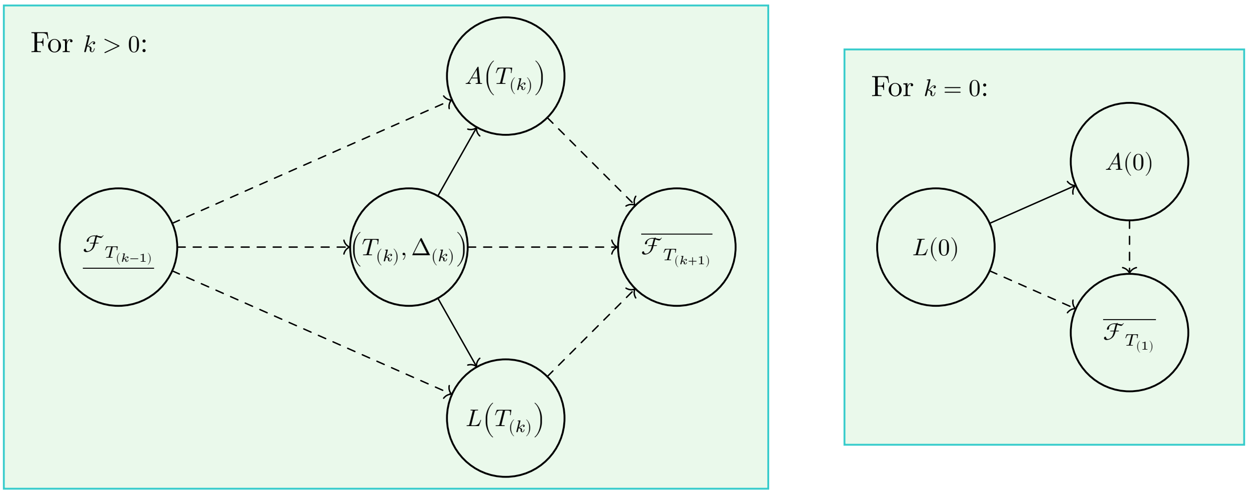

Figure 2: A DAG for simulating the data generating mechanism. The dashed lines indicate multiple edges from the dependencies in the past and into the future. Here $\underline{\mathcal{F}_{T_{(k)}}}$ is the history up to and including the k'th event and $\overline{\mathcal{F}_{T_{(k)}}}$ is the history after and including the k'th event.