Contents

1. Simulating longitudinal data for time-to-event analysis in continuous time

We simulate a cohort of patients who initiate treatment at time $t = 0$, denoted by $A(0) = 1$ and who are initially stroke-free, $L(0) = 0$. All individuals are followed for up to $\tau_{\text{end}} = 900$ days or until death. Initially, we do not consider censoring or competing events. During follow-up, patients may experience (at most) one stroke, stop treatment (irreversibly), and die, that is $N^x(t) \leq 1$ for $x = a, \ell, y$. With these assumptions $K = 2$ in the main note (at most two non-terminal events). The primary outcome is the *risk of death within* $\tau = 720$ *days.*

Our observations consist of

$$O = \left(T_{(3)}, \Delta_{(3)}, A\left(T_{(2)}\right), L\left(T_{(2)}\right), T_{(2)}, \Delta_{(2)}, A\left(T_{(1)}\right), L\left(T_{(1)}\right), T_{(1)}, \Delta_{(1)}, A(0), L(0), \text{age}\right),$$

where $T_{(k)}$ is the time of the k'th event, $\Delta_{(k)} \in \{\ell, a, y\}$ (stroke, visit, death), $A\left(T_{(k)}\right)$ is the treatment status at time $T_{(k)}$, and $L\left(T_{(k)}\right)$ is the value of the covariate at time $T_{(k)}$. Note that we let $T_{(k)} = \infty$ if the k'th event cannot happen (because the previous event was a terminal event or the end of the study period was reached) or if the k'th event occurs after the end of the study period. Let

$\mathrm{Exp}(\lambda)$ denote the exponential distribution with rate $\lambda \geq 0$. We let $\lambda = 0$ correspond to the case that the event cannot happen, i.e., $T^x_{(1)} = \infty$.

Then, we generate the baseline variables as follows

$$
\begin{aligned}
&\text{age} \sim \mathrm{Unif}(40, 90), \\
&L = 0, \\
&A(0) = 1,
\end{aligned}
$$

Now we describe the simulation mechanism corresponding to the first event that can happen. We define $T^a_{(1)}$ such that the patient can be expected to go to the doctor within the first year, if the two other events have not occurred first. As the first event, we draw

$$
\begin{aligned}
&T^x_{(1)} \sim \mathrm{Exp}\left(\lambda^x_1 \exp\left(\beta^x_{1,\text{age}} \text{ age} + \beta^x_{1,\mathrm{A}} A\left(T_{(1)}\right) + \beta^x_{1,\mathrm{L}} L\left(T_{(1)}\right)\right)\right), x = \ell, y \\
&T^a_{(1)} \sim 1 + \mathcal{N}(0, \delta) \\
&\Delta_{(1)} = \operatorname{argmin}_{x=a,\ell,y} T^x_{(1)} \\
&T_{(1)} = T^{\Delta_{(1)}}_{(1)} \\
&A\left(T_{(1)}\right) \mid T_{(1)} = t, \text{age} = x \begin{cases} \sim \text{Bernoulli}(\text{expit}\left(\alpha_{10} + \alpha_{1,\text{ age}} x\right) \text{ if } \Delta_{(1)} = a \\ 1 \text{ otherwise} \end{cases} \\
&L\left(T_{(1)}\right) = 1,
\end{aligned}
$$

Note that we simulate from a "competing event setup" by defining latent variables $T^x_{(1)}$ for each possible event type x.

We now describe the second event that can happen. If the first event was a terminal event – either outcome or that the previous event happened after the end of the study period – we stop and do not generate more data for this patient. Now, we let $S_{(2)}$ denote the time between $T_{(1)}$ and the second event $T_{(2)}$, i.e., $S_{(2)} = T_{(2)} - T_{(1)}$. As we required that $N^x(t) \leq 1$, if the first event was a stroke, we cannot have a second stroke, and if the first event was a visit, we cannot have a second visit. If the first event was a stroke, the doctor visit is likely to happen soon after, so we simulate the corresponding latent time as an exponential random variable. We will then generate the second event time $T_{(2)}$ as follows:

$$
\begin{aligned}
&S^\ell_{(2)} \sim \mathrm{Exp}\left(\lambda^\ell_2 \exp\left(\beta^\ell_{2,\text{age}} \text{ age} + \beta^\ell_{2,A} A\left(T_{(1)}\right)\right) \mathbb{1}\left\{\Delta_{(1)} = a\right\}\right) \\
&S^y_{(2)} \sim \mathrm{Exp}\left(\lambda^y_2 \exp\left(\beta^y_{2,\text{age}} \text{ age} + \beta^y_{2,A} A\left(T_{(1)}\right) + \beta^\ell_{2,L} L\left(T_{(1)}\right)\right)\right) \\
&S^a_{(2)} \sim \mathrm{Exp}\left(\gamma_0 \exp\left(\gamma_{\text{age}} \text{ age}\right) \mathbb{1}\left\{\Delta_{(1)} = \ell\right\}\right) \\
&\Delta_{(2)} = \operatorname{argmin}_{x=a,\ell,y,} S^x_{(2)} \\
&T_{(2)} = T_{(1)} + S^{\Delta_{(2)}}_{(2)} \\
&A\left(T_{(2)}\right) \mid T_{(2)} = t, \text{age} = x, A\left(T_{(1)}\right) = a_1, L\left(T_{(1)}\right) = l_1 \\
&\qquad\qquad = \begin{cases} \sim \text{Bernoulli}(\text{expit}\left(\alpha_{20} + \alpha_{2,\text{ age}} x + \alpha_{2,\mathrm{L}} l_1\right)) \text{ if } \Delta_{(2)} = a \\ 1 \text{ otherwise} \end{cases} \\
&L\left(T_{(2)}\right) = 1.
\end{aligned}
$$

Finally, we let $T_{(3)} = S_{(3)} + T_{(2)}$ denote the time of the third event, if it can happen. We define the time $S_{(3)}$ as follows:

$$S^y_{(3)} \sim \mathrm{Exp}\Big(\lambda^y_3 \exp\Big(\beta^y_{3,\mathrm{age}}\ \mathrm{age} + \beta^y_{3,A} A\big(T_{(2)}\big) + \beta^y_{3,L} L\big(T_{(2)}\big)\Big)\Big)$$
$$\Delta_{(3)} = y$$
$$T_{(3)} = T_{(2)} + S^{\Delta_{(3)}}_{(3)}.$$

Here, we furthermore make the assumption that it does not matter whether the patient had a stroke first and then visited the doctor, or visited the doctor first and then had a stroke. Also, we assumed that the previous event times have no influence on anything, only the "marks". However, this may be unrealistic, as the effect of a stroke on mortality may naturally decrease over time.

When the static intervention is applied, we put $A\big(T_{(k)}\big) = 1$ for each $k = 1, ..., K$. Some explanation for this is (probably) warranted (see Section 6).

Section 2 details the target estimand and how to arrive at the iterative conditional expectation formula. Also, we discuss how to use the algorithm in practice for the simple data generating mechanism. In Section 3, we present a simulation study to illustrate the algorithm and the target estimand with time-varying confounding and censored data. Finally, the mechanism detailed here may be too simplistic. We discuss extensions in Section 5.

1.1. Plain Language Summary (for Clinical Audience)

We simulate patients who all begin treatment and are initially healthy. Over two years, they may have a stroke, stop treatment (only at doctor visits), or die. A routine doctor visit is scheduled about a year after treatment begins, unless a stroke happens first, in which case a visit is likely to occur soon after. Doctors are less likely to stop treatment after a stroke. The chance of dying depends on whether the patient has had a stroke and whether they are still on treatment.

2. Target estimand and example usage of algorithm

We explain here what the target estimand is and how to arrive at the iterative conditional expectation formula. Recall that

$$\begin{aligned}
\mathcal{F}_0 &= (A(0), L(0)),\\
\mathcal{F}_{T_{(1)}} &= \big(T_{(1)}, \Delta_{(1)}, A\big(T_{(1)}\big), L\big(T_{(1)}\big), A(0), L(0), \mathrm{age}\big),\\
\mathcal{F}_{T_{(2)}} &= \big(T_{(2)}, \Delta_{(2)}, A\big(T_{(2)}\big), L\big(T_{(2)}\big), T_{(1)}, \Delta_{(1)}, A\big(T_{(1)}\big), L\big(T_{(1)}\big), A(0), L(0), \mathrm{age}\big),
\end{aligned}$$

Let $\Lambda^x_k\Big(t, \mathcal{F}_{T_{(k-1)}}\Big)$ denote the cumulative cause-specific hazard function for $T_{(k)}$ and $\Delta_{(k)} = x$ at time t given the history $\mathcal{F}_{T_{(k-1)}}$. For instance, if $k = 2$ and $x = y$, then in the simulation mechanism, we have

$$\Lambda^y_2\Big(t, \mathcal{F}_{T_{(2-1)}}\Big) = \lambda^y_2 \exp\Big(\beta^y_{2,\ \mathrm{age}}\ \mathrm{age} + \beta^y_{2,A} A\big(T_{(1)}\big) + \beta^y_{2,L} L\big(T_{(1)}\big)\Big)$$

Furthermore, let $\pi_k\Big(t, \mathcal{F}_{T_{(k-1)}}\Big)$ denote the probability of being treated as the k'th event given that you go to the doctor at time t, i.e., $\Delta_{(k)} = a$, and your history $\mathcal{F}_{T_{(k-1)}}$. We let, for convenience of notation, "age" be included in $L(0)$.

Using the notation $f_{t_k} = (t_k, d_k, a_k, l_k, ..., a_0, l_0)$ with $f_0 = (a_0, l_0)$, we, analogously to (Rytgaard et al., 2022), define our target parameter $\Psi_\tau : \mathcal{P} \to \mathbb{R}$ for a non-parametric model $\mathcal{P}$ as

$$
\begin{aligned}
\Psi_\tau(P) = \int \Bigg(\int_{(0,\tau]} \Bigg(\int_{(t_1,\tau]} \Bigg(\int_{(t_2,\tau]} \prod_{w_3 \in (t_3,\tau)} \left(1 - \Lambda_3^y\left(\mathrm{d}w_3 \mid f_{t_2}\right)\right) \Lambda_3^y\left(\mathrm{d}t_3 \mid f_{t_2}\right) \Bigg) \\
\times \prod_{w_2 \in (t_2,\tau)} \Bigg(1 - \sum_{x=a,y} \Lambda_2^x\left(\mathrm{d}w_2 \mid f_{t_1}\right)\Bigg) \mathbb{1}\{a_2 = 1\} \times \Lambda_2^a\left(\mathrm{d}t_2 \mid f_{t_1}\right)\Bigg) \\
\times \prod_{w_1 \in (t_1,\tau)} \Bigg(1 - \sum_{x=a,y,\ell} \Lambda_1^x(\mathrm{d}w_1 \mid f_0)\Bigg) \Lambda_1^\ell(\mathrm{d}t_1 \mid f_0)\Bigg) P_{L(0)}(\mathrm{d}l_0) \\
+ \int \Bigg(\int_{(0,\tau]} \Bigg(\int_{(t_1,\tau]} \Bigg(\int_{(t_2,\tau]} \prod_{w_3 \in (t_3,\tau)} \left(1 - \Lambda_3^y\left(\mathrm{d}w_3 \mid f_{t_2}\right)\right) \Lambda_3^y\left(\mathrm{d}t_3 \mid f_{t_2}\right) \Bigg) \\
\times \prod_{w_2 \in (t_2,\tau)} \Bigg(1 - \sum_{x=y,\ell} \Lambda_2^x\left(\mathrm{d}w_2 \mid f_{t_1}\right)\Bigg) \Lambda_2^\ell\left(\mathrm{d}t_2 \mid f_{t_1}\right)\Bigg) \\
\times \prod_{w_1 \in (t_1,\tau)} \Bigg(1 - \sum_{x=a,y,\ell} \Lambda_1^x(\mathrm{d}w_1 \mid f_0)\Bigg) \mathbb{1}\{a_1 = 1\} \times \Lambda_1^a(\mathrm{d}t_1 \mid f_0)\Bigg) P_{L(0)}(\mathrm{d}l_0) \\
+ \int \Bigg(\int_{(0,\tau]} \Bigg(\int_{(t_1,\tau]} \prod_{w_2 \in (t_2,\tau)} \Bigg(1 - \sum_{x=a,y} \Lambda_2^x\left(\mathrm{d}w_2 \mid f_{t_1}\right)\Bigg) \Lambda_2^y\left(\mathrm{d}t_2 \mid f_{t_1}\right)\Bigg) \\
\times \prod_{w_1 \in (t_1,\tau)} \Bigg(1 - \sum_{x=a,y,\ell} \Lambda_1^x(\mathrm{d}w_1 \mid f_0)\Bigg) \Lambda_1^\ell(\mathrm{d}t_1 \mid f_0)\Bigg) P_{L(0)}(\mathrm{d}l_0) \\
+ \int \Bigg(\int_{(0,\tau]} \Bigg(\int_{(t_1,\tau]} \prod_{w_2 \in (t_2,\tau)} \Bigg(1 - \sum_{x=a,y} \Lambda_2^x\left(\mathrm{d}w_2 \mid f_{t_1}\right)\Bigg) \Lambda_2^y\left(\mathrm{d}t_2 \mid f_{t_1}\right)\Bigg) \\
\times \prod_{w_1 \in (t_1,\tau)} \Bigg(1 - \sum_{x=a,y,\ell} \Lambda_1^x(\mathrm{d}w_1 \mid f_0)\Bigg) \mathbb{1}\{a_1 = 1\} \times \Lambda_1^a(\mathrm{d}t_1 \mid f_0)\Bigg) P_{L(0)}(\mathrm{d}l_0) \\
+ \int \Bigg(\int_{(0,\tau]} \prod_{w_1 \in (t_1,\tau)} \Bigg(1 - \sum_{x=a,y,\ell} \Lambda_1^x(\mathrm{d}w_1 \mid f_0)\Bigg) \Lambda_1^y(\mathrm{d}t_1 \mid f_0)\Bigg) P_{L(0)}(\mathrm{d}l_0), \\
:= \zeta_1(P) + \zeta_2(P) + \zeta_3(P) + \zeta_4(P) + \zeta_5(P)
\end{aligned}
$$

corresponding to setting $\pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right) = 1$. This expression is fairly long and quite complicated.

We now explain how one goes from this apparently complicated expression to the iterative conditional expectation formula, which reduces the dimensionality of the problem significantly.

Example 2.1.: Let,

$$Z_3^a = \mathbb{1}\left\{T_{(3)} \leq \tau\right\},$$

$$\bar{Q}_{2,\tau}^g\left(f_{t_2}\right) \\ = \mathbb{E}_P\left[Z_3^a \mid \mathcal{F}_{T_{(2)}} = f_{t_2}\right],$$

$$Z_2^a = \mathbb{1}\left\{T_{(2)} \leq \tau, \Delta_{(2)} = y\right\} + \mathbb{1}\left\{T_{(2)} < \tau, \Delta_{(2)} = \ell\right\}\bar{Q}_{2,\tau}^g\left(\mathcal{F}_{T_{(2)}}\right) + \mathbb{1}\left\{T_{(2)} < \tau, \Delta_{(2)} = a\right\}\bar{Q}_{2,\tau}^g\left(\mathcal{F}_{T_{(2)}}^{\mathbf{1}}\right)$$

$$\bar{Q}_{1,\tau}^g\left(f_{t_1}\right) \\ = \mathbb{E}_P\left[Z_2^a \mid \mathcal{F}_{T_{(1)}} = f_{t_1}\right],$$

$$Z_1^a = \mathbb{1}\left\{T_{(1)} \leq \tau, \Delta_{(1)} = y\right\} + \mathbb{1}\left\{T_{(1)} < \tau, \Delta_{(1)} = \ell\right\}\bar{Q}_{1,\tau}^g\left(\mathcal{F}_{T_{(1)}}\right) + \mathbb{1}\left\{T_{(1)} < \tau, \Delta_{(1)} = a\right\}\bar{Q}_{1,\tau}^g\left(\mathcal{F}_{T_{(1)}}^{\mathbf{1}}\right)$$

$$\bar{Q}_{0,\tau}^g(f_0) \\ = \mathbb{E}_P[Z_1^a \mid \mathcal{F}_0 = f_0],$$

and

$$\mathcal{F}_{T_{(2)}}^{\mathbf{1}} = \left(T_{(2)}, \Delta_{(2)}, 1, L\left(T_{(2)}\right), T_{(1)}, \Delta_{(1)}, A\left(T_{(1)}\right), L\left(T_{(1)}\right), A(0), L(0), \text{age}\right),$$
$$\mathcal{F}_{T_{(1)}}^{\mathbf{1}} = \left(T_{(1)}, \Delta_{(1)}, 1, L\left(T_{(1)}\right), A(0), L(0), \text{age}\right),$$

which is the history where we set the *current* treatment to 1. Then, $\Psi_\tau(P) = \mathbb{E}_{P_{L(0)}}\left[\bar{Q}_{0,\tau}^g(\mathcal{F}_0)\right]$, where $P_{L(0)}$ is the distribution of the baseline confounders.

Proof:

First note that,

$$P\left(T_{(k)} \leq t, \Delta_{(k)} = x \mid \mathcal{F}_{T_{(k-1)}} = f_{t_{k-1}}\right) \\ = \int_{(t_{k-1}, s]} \prod_{u \in (t_{k-1}, s)} \left(1 - \sum_{x = a, \ell, y} \Lambda_k^x\left(\mathrm{d}u \mid f_{t_{k-1}}\right)\right) \Lambda_k^x\left(\mathrm{d}s \mid f_{t_{k-1}}\right), t < \tau_{\text{end}}$$

for $x = y, \ell$ by definition and

$$P\left(T_{(k)} \leq t, \Delta_{(k)} = a, A\left(T_{(k)}\right) = 1 \mid \mathcal{F}_{T_{(k-1)}} = f_{t_{k-1}}\right) \\ = \int_{(t_{k-1}, s]} \prod_{u \in (t_{k-1}, s)} \left(1 - \sum_{x = a, \ell, y} \Lambda_k^x\left(\mathrm{d}u \mid f_{t_{k-1}}\right)\right) \pi_k\left(T_{(k)}, \mathcal{F}_{T_{(k-1)}}\right) \Lambda_k^a\left(\mathrm{d}s \mid f_{t_{k-1}}\right)$$

Using this, we now see that

$$
\begin{aligned}
\zeta_1(P) = \mathbb{E}_{P_{L(0)}}\Bigg[\mathbb{E}_P\Bigg[\mathbb{1}\Big\{T_{(1)} < \tau, \Delta_{(1)} = \ell\Big\}\mathbb{E}_P\Bigg[\mathbb{1}\Big\{T_{(2)} < \tau, \Delta_{(2)} = a\Big\}\frac{\mathbb{1}\Big\{A\Big(T_{(2)}\Big) = 1\Big\}}{\pi_2\Big(T_{(2)}, \mathcal{F}_{T_{(1)}}\Big)} \\
\times \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(3)} \le \tau\Big\} \mid \mathcal{F}_{T_{(2)}}\Big] \mid \mathcal{F}_{T_{(1)}}\Bigg] \mid \mathcal{F}_0\Bigg]\Bigg] \\
= \mathbb{E}_{P_{L(0)}}\Bigg[\mathbb{E}_P\Bigg[\Bigg(\mathbb{1}\Big\{T_{(1)} < \tau, \Delta_{(1)} = \ell\Big\} + \mathbb{1}\Big\{T_{(1)} < \tau, \Delta_{(1)} = a\Big\}\frac{\mathbb{1}\Big\{A\Big(T_{(1)}\Big) = 1\Big\}}{\pi_1\Big(T_{(1)}, \mathcal{F}_0\Big)}\Bigg) \\
\times \mathbb{E}_P\Bigg[\mathbb{1}\Big\{T_{(2)} < \tau, \Delta_{(2)} = a\Big\}\frac{\mathbb{1}\Big\{A\Big(T_{(2)}\Big) = 1\Big\}}{\pi_2\Big(T_{(2)}, \mathcal{F}_{T_{(1)}}\Big)} \\
\times \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(3)} \le \tau\Big\} \mid \mathcal{F}_{T_{(2)}}\Big] \mid \mathcal{F}_{T_{(1)}}\Bigg] \mid \mathcal{F}_0\Bigg]\Bigg]
\end{aligned}
$$

and

$$
\begin{aligned}
\zeta_2(P) = \mathbb{E}_{P_{L(0)}}\Bigg[\mathbb{E}_P\Bigg[\mathbb{1}\Big\{T_{(1)} < \tau, \Delta_{(1)} = a\Big\}\frac{\mathbb{1}\Big\{A\Big(T_{(1)}\Big) = 1\Big\}}{\pi_1\Big(T_{(1)}, \mathcal{F}_0\Big)}\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(2)} < \tau, \Delta_{(2)} = \ell\Big\} \\
\times \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(3)} \le \tau\Big\} \mid \mathcal{F}_{T_{(2)}}\Big] \mid \mathcal{F}_{T_{(1)}}\Big] \mid \mathcal{F}_0\Bigg]\Bigg] \\
= \mathbb{E}_{P_{L(0)}}\Bigg[\mathbb{E}_P\Bigg[\Bigg(\mathbb{1}\Big\{T_{(1)} < \tau, \Delta_{(1)} = \ell\Big\} + \mathbb{1}\Big\{T_{(1)} < \tau, \Delta_{(1)} = a\Big\}\frac{\mathbb{1}\Big\{A\Big(T_{(1)}\Big) = 1\Big\}}{\pi_1\Big(T_{(1)}, \mathcal{F}_0\Big)}\Bigg) \\
\times \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(2)} < \tau, \Delta_{(2)} = \ell\Big\} \times \mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(3)} \le \tau\Big\} \mid \mathcal{F}_{T_{(2)}}\Big] \mid \mathcal{F}_{T_{(1)}}\Big] \mid \mathcal{F}_0\Bigg]\Bigg]
\end{aligned}
$$

Here, we simply add terms which are zero corresponding to two treatments and two strokes. Similarly, we find

$$
\begin{aligned}
\zeta_3(P) &= \mathbb{E}_{P_{L(0)}}\Big[\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(1)} < \tau, \Delta_{(1)} = \ell\Big\}\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(2)} \le \tau, \Delta_{(2)} = y\Big\} \mid \mathcal{F}_{T_{(1)}}\Big] \mid \mathcal{F}_0\Big]\Big] \\
\zeta_4(P) &= \mathbb{E}_{P_{L(0)}}\Bigg[\mathbb{E}_P\Bigg[\frac{\mathbb{1}\Big\{T_{(1)} < \tau, \Delta_{(1)} = a\Big\}}{\pi_1\Big(T_{(1)}, \mathcal{F}_0\Big)}\mathbb{1}\Big\{A\Big(T_{(1)}\Big) = 1\Big\}\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(2)} \le \tau, \Delta_{(2)} = y\Big\} \mid \mathcal{F}_{T_{(1)}}\Big] \mid \mathcal{F}_0\Bigg]\Bigg] \\
\zeta_5(P) &= \mathbb{E}_{P_{L(0)}}\Big[\mathbb{E}_P\Big[\mathbb{1}\Big\{T_{(1)} \le \tau, \Delta_{(1)} = y\Big\} \mid \mathcal{F}_0\Big]\Big]
\end{aligned}
$$

Now we note that

$$
\begin{aligned}
&\mathbb{E}_P\left[\mathbb{1}\left\{T_{(2)}<\tau, \Delta_{(2)}=a\right\} \frac{\mathbb{1}\left\{A\left(T_{(2)}\right)=1\right\}}{\pi_2\left(T_{(2)}, \mathcal{F}_{T_{(1)}}\right)} \bar{Q}_{2, \tau}^g\left(\mathcal{F}_{T_{(2)}}\right) \mid \mathcal{F}_{T_{(1)}}\right] \\
&=\mathbb{E}_P\left[\mathbb{1}\left\{T_{(2)}<\tau, \Delta_{(2)}=a\right\} \frac{\mathbb{1}\left\{A\left(T_{(2)}\right)=1\right\}}{\pi_2\left(T_{(2)}, \mathcal{F}_{T_{(1)}}\right)} \bar{Q}_{2, \tau}^g\left(\mathcal{F}_{T_{(2)}}^{\mathbf{1}}\right) \mid \mathcal{F}_{T_{(1)}}\right] \\
&=\mathbb{E}_P\left[\mathbb{1}\left\{T_{(2)}<\tau, \Delta_{(2)}=a\right\} \frac{\mathbb{E}_P\left[\mathbb{1}\left\{A\left(T_{(2)}\right)=1\right\} \mid \mathcal{F}_{T_{(1)}}, \Delta_{(2)}, T_{(2)}\right]}{\pi_2\left(T_{(2)}, \mathcal{F}_{T_{(1)}}\right)} \bar{Q}_{2, \tau}^g\left(\mathcal{F}_{T_{(2)}}^{\mathbf{1}}\right) \mid \mathcal{F}_{T_{(1)}}\right] \\
&=\mathbb{E}_P\left[\mathbb{1}\left\{T_{(2)}<\tau, \Delta_{(2)}=a\right\} \bar{Q}_{2, \tau}^g\left(\mathcal{F}_{T_{(2)}}^{\mathbf{1}}\right) \mid \mathcal{F}_{T_{(1)}}\right],
\end{aligned}
$$

Here, we use the iterated law of conditional expectations and that $\pi_2\left(T_{(2)}, \mathcal{F}_{T_{(1)}}\right)$ is the probability of being treated at time $T_{(2)}$ given the history $\mathcal{F}_{T_{(1)}}$ and that you visit the doctor as the second event. Using this, we can rewrite $\zeta_1(P)$ as follows:

$$
\begin{aligned}
\zeta_1(P)=\mathbb{E}_{P_{L(0)}}\Bigg[\mathbb{E}_P\Bigg[&\left(\mathbb{1}\left\{T_{(1)}<\tau, \Delta_{(1)}=\ell\right\}+\mathbb{1}\left\{T_{(1)}<\tau, \Delta_{(1)}=a\right\} \frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_1\left(T_{(1)}, \mathcal{F}_0\right)}\right) \\
&\times \mathbb{E}_P\left[\mathbb{1}\left\{T_{(2)}<\tau, \Delta_{(2)}=a\right\} \bar{Q}_{2, \tau}^g\left(\mathcal{F}_{T_{(2)}}^{\mathbf{1}}\right) \mid \mathcal{F}_{T_{(1)}}\right] \mid \mathcal{F}_0\Bigg]\Bigg]
\end{aligned}
$$

Adding $\zeta_1(P), \zeta_2(P), \zeta_3(P), \zeta_4(P)$, together, we find

$$
\begin{aligned}
\sum_{j=1}^4 \zeta_j(P)=\mathbb{E}_{P_{L(0)}}\Bigg[\mathbb{E}_P\Bigg[&\left(\mathbb{1}\left\{T_{(1)}<\tau, \Delta_{(1)}=\ell\right\}+\mathbb{1}\left\{T_{(1)}<\tau, \Delta_{(1)}=a\right\} \frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=1\right\}}{\pi_1\left(T_{(1)}, \mathcal{F}_0\right)}\right) \\
&\times \bar{Q}_{1, \tau}^g\left(\mathcal{F}_{T_{(1)}}\right) \mid \mathcal{F}_0\Bigg]\Bigg]
\end{aligned}
$$

Similarly, we have that

$$
\begin{aligned}
&\mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)}<\tau, \Delta_{(1)}=a\right\} \frac{\mathbb{1}\left\{A\left(T_{(1)}\right)=a\right\}}{\pi_1\left(T_{(1)}, \mathcal{F}_0\right)} \bar{Q}_{1, \tau}^g\left(\mathcal{F}_{T_{(1)}}\right) \mid \mathcal{F}_0\right] \\
&=\mathbb{E}_P\left[\mathbb{1}\left\{T_{(1)}<\tau, \Delta_{(1)}=a\right\} \bar{Q}_{1, \tau}^g\left(\mathcal{F}_{T_{(1)}}^{\mathbf{1}}\right) \mid \mathcal{F}_0\right],
\end{aligned}
$$

so that

$$
\begin{aligned}
\sum_{j=1}^4 \zeta_j(P)=\mathbb{E}_{P_{L(0)}}\Big[\mathbb{E}_P\Big[&\mathbb{1}\left\{T_{(1)}<\tau, \Delta_{(1)}=\ell\right\} \bar{Q}_{1, \tau}^g\left(\mathcal{F}_{T_{(1)}}\right)+\mathbb{1}\left\{T_{(1)}<\tau, \Delta_{(1)}=a\right\} \bar{Q}_{1, \tau}^g\left(\mathcal{F}_{T_{(1)}}^{\mathbf{1}}\right) \\
&+\mathbb{1}\left\{T_{(1)} \leq \tau, \Delta_{(1)}=y\right\} \mid \mathcal{F}_0\Big]\Big].
\end{aligned}
$$

Now, we finally see that

$$
\sum_{j=1}^4 \zeta_j(P)+\zeta_5(P)=\mathbb{E}_{P_{L(0)}}\left[\bar{Q}_{0, \tau}^g\left(\mathcal{F}_0\right)\right],
$$

and we are done. $\square$

Thus, regression techniques can be used to estimate the target parameter $\Psi_\tau(P)$. Not only can it be used for the estimates of the target parameter, but it turns that the terms $\bar{Q}^g_{0,\tau}$, $\bar{Q}^g_{1,\tau}$, and $\bar{Q}^g_{2,\tau}$, $\Psi_\tau(P)$, as well as $\pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ are precisely the terms we encounter in the efficient influence function. Therefore, inference can be obtained as part of the procedure where we estimate $\bar{Q}^g_{0,\tau}$, $\bar{Q}^g_{1,\tau}$, and $\bar{Q}^g_{2,\tau}$ by considering a one-step estimator. Furthermore, the resulting estimator will be doubly robust. We also give the ICE formula in case of censoring. Let $\Lambda^c_k\left(\mathrm{d}t \mid \mathcal{F}_{T_{(k-1)}}\right)$ denote the cumulative cause-specific hazard function for the censoring for the k'th event. Also, let $\Lambda^c(t)$ denote the compensator for the censoring process with respect to the natural filtration (e.g., the observed filtration) of all processes involved.

Example 2.2.: Let $S^c_{(k)}\left(T_{(k)} - \mid \mathcal{F}_{T_{(k-1)}}\right) = \prod_{s \in \left(T_{(k-1)}, T_{(k)}\right)} \left(1 - \Lambda^c_k\left(\mathrm{d}s \mid \mathcal{F}_{T_{(k-1)}}\right)\right) = \prod_{s \in \left(T_{(k-1)}, T_{(k)}\right)} (1 - \Lambda^c(\mathrm{d}s))$ denote the survival function for the censoring process. Defining,

$$
\begin{aligned}
&Z^a_3 = \frac{\mathbb{1}\left\{T_{(3)} \leq \tau\right\}}{S^c_{(3)}\left(T_{(3)} - \mid \mathcal{F}_{T_{(2)}}\right)}, \\
&\bar{Q}^g_{2,\tau}\left(f_{t_2}\right) \\
&= \mathbb{E}_P\left[Z^a_3 \mid \mathcal{F}_{T_{(2)}} = f_{t_2}\right], \\
&Z^a_2 = \frac{1}{S^c_{(2)}\left(T_{(1)} - \mid \mathcal{F}_{T_{(1)}}\right)} \times \Big(\mathbb{1}\left\{T_{(2)} \leq \tau, \Delta_{(2)} = y\right\} + \mathbb{1}\left\{T_{(2)} < \tau, \Delta_{(2)} = \ell\right\}\bar{Q}^g_{2,\tau}\left(\mathcal{F}_{T_{(2)}}\right) + \\
&\qquad \mathbb{1}\left\{T_{(2)} < \tau, \Delta_{(2)} = a\right\}\bar{Q}^g_{2,\tau}\left(\mathcal{F}^{\mathbf{1}}_{T_{(2)}}\right)\Big) \\
&\bar{Q}^g_{1,\tau}\left(f_{t_1}\right) \\
&= \mathbb{E}_P\left[Z^a_2 \mid \mathcal{F}_{T_{(1)}} = f_{t_1}\right], \\
&Z^a_1 = \frac{1}{S^c_{(1)}\left(T_{(1)} - \mid \mathcal{F}_0\right)} \times \Big(\mathbb{1}\left\{T_{(1)} \leq \tau, \Delta_{(1)} = y\right\} + \mathbb{1}\left\{T_{(1)} < \tau, \Delta_{(1)} = \ell\right\}\bar{Q}^g_{1,\tau}\left(\mathcal{F}_{T_{(1)}}\right) + \\
&\qquad \mathbb{1}\left\{T_{(1)} < \tau, \Delta_{(1)} = a\right\}\bar{Q}^g_{1,\tau}\left(\mathcal{F}^{\mathbf{1}}_{T_{(1)}}\right)\Big) \\
&\bar{Q}^g_{0,\tau}(f_0) \\
&= \mathbb{E}_P[Z^a_1 \mid \mathcal{F}_0 = f_0],
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{F}^{\mathbf{1}}_{T_{(2)}} &= \left(T_{(2)}, \Delta_{(2)}, 1, L\left(T_{(2)}\right), T_{(1)}, \Delta_{(1)}, A\left(T_{(1)}\right), L\left(T_{(1)}\right), A(0), L(0), \mathrm{age}\right), \\
\mathcal{F}^{\mathbf{1}}_{T_{(1)}} &= \left(T_{(1)}, \Delta_{(1)}, 1, L\left(T_{(1)}\right), A(0), L(0), \mathrm{age}\right),
\end{aligned}
$$

which is the history where we set the *current* treatment to 1. Then, $\Psi_\tau(P) = \mathbb{E}_{P_{L(0)}}\left[\bar{Q}^g_{0,\tau}(\mathcal{F}_0)\right]$, where $P_{L(0)}$ is the distribution of the baseline confounders.

2.1. Example usage of the Algorithm

To help illustrate the algorithm, we present a simple example in Table 1 in the case where $\tau = 5$. We start at $k = 3$. As in the rest of the paper, we suppose that the time horizon is $\tau = 720$. We apply the definitions given in Example 2.1..

Iteration $k = 3$

Denote by $R_{3,\tau}$ the set of people for which it is possible for them to die as their third event before time τ, that is people with $T_{(2)} < \tau$ and $T_{(2)} \in \{a, \ell\}$ (otherwise the probability we are trying to determine $\bar{Q}^g_{2,\tau}$ is zero). We find that $R_{3,\tau} = \{6, 7\}$. For each of these people find Z^a_3 and regress on $\mathcal{F}_{T_{(2)}}$ to obtain a prediction function $\hat{\nu}_2$. In **R**, this can be done as follows via e.g., `glm` assuming the data is given as `data`:

```
D_3 <- data[data$status_2 %in% c("a", "l") & data$time_2 < tau, ] ## data set that
consists of ids from R_3
data$Z_3 <- 1*(D_3$time_3 <= tau)
fit <- glm(Z_3 ~ time_2+status_2+A_2+L_2+ time_1+status_1+A_1+L_1+A_0+L_0+age,
data = D_3, family = "binomial") ## example; can use ML methods or in principle
also include interactions
hat_nu_2 <- function(data) {predict(fit, newdata=data, type = "response")}
```

Iteration $k = 2$

As in the case $k = 3$, we find $R_{2,\tau} = \{3, 4, 6, 7\}$.

- For $i = 3$, we produce the altered history, where $\mathcal{F}^1_{\bar{T}_{2,3}} = (55, 0, 1, 1, 1, 62, \ell, 1, 1)$ to $\hat{\nu}_2$ and find $\hat{\nu}_2\left(\mathcal{F}^1_{\bar{T}_{2,3}}\right)$. Based on this we calculate $\hat{Z}^a_{2,3} = 0 \times \hat{\nu}_2\left(\mathcal{F}^1_{\bar{T}_{2,3}}\right)$.
- For $i = 4$, we apply the actual history $\mathcal{F}_{\bar{T}_{2,4}}$ to $\hat{\nu}_2$. Again, we calculate $\hat{Z}^a_{2,4} = 1 \times \hat{\nu}_2\left(\mathcal{F}_{\bar{T}_{2,4}}\right)$.
- For $i = 6$, we simply find $\hat{Z}^a_{2,6} = 1$.
- For $i = 7$, similarly calculate $\hat{Z}^a_{2,7} = 1 \times \hat{\nu}_2\left(\mathcal{F}_{\bar{T}_{2,7}}\right)$.

Regress the predicted values $\hat{Z}^a_2$ on $\mathcal{F}_{T_{(1)}}$ to obtain a prediction function $\hat{\nu}_1$. In **R**, this can be done as follows via e.g., `lm` assuming the data is given as `data`:

```
D_2 <- data[data$status_1 %in% c("a", "l") & data$time_1 < tau, ] ## data set that
consists of ids from R_2
D_2a <- copy(data)
D_2a[data$status_2 == "a", "treat_2"] <- 1 ## alter treat_2 to 1 for those with
status_2 = a
data$Z_2 <- 1*(D_2$time_2 <= tau & D_2$status_2 == "y") +
    1*(D_2$time_2 < tau & D_2$status_2 == "l") * hat_nu_2(D_2a) +
    1*(D_2$time_2 < tau & D_2$status_2 == "a") * hat_nu_2(D_2a)
fit <- lm(Z_2 ~ time_1+status_1+A_1+L_1+A_0+L_0+age, data = D_2)
hat_nu_1 <- function(data) {predict(fit, newdata=data, type = "response")}
```

Iteration $k = 1$

Same procedure as $k = 2$. Note that we find $R_{1,\tau} = \{1, 2, 3, 4, 5, 6, 7\}$.

Iteration $k = 0$

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

id	age	$L(T_{(1)})$	$A(T_{(1)})$	$T_{(1)}$	$\Delta_{(1)}$	$L(T_{(2)})$	$A(T_{(2)})$	$T_{(2)}$	$\Delta_{(2)}$	$T_{(3)}$	$\Delta_{(3)}$
1	60	0	1	745	a	Ø	Ø	770	y	∞	Ø
2	50	Ø	Ø	140	y	Ø	Ø	∞	Ø	∞	Ø
3	55	1	1	62	ℓ	1	0	850	a	870	y
4	46	1	1	70	ℓ	1	1	170	a	182	y
5	67	Ø	Ø	60	y	Ø	Ø	∞	Ø	∞	Ø
6	52	1	1	120	ℓ	Ø	Ø	175	y	∞	Ø
7	56	0	0	40	a	1	0	80	ℓ	645	y

Table 1: Example data for illustration of the ICE algorithm. $L(0)$ and $A(0)$ are not shown as they are constant.

3. Simulation study

Consider the following example coefficients for the simulation mechanism, corresponding to a simulation mechanism, which is compatible with the time-varying Cox model, e.g., $\lambda^y := \lambda^y_1 = \lambda^y_2 = \lambda^y_3$ (see e.g., Section 6). We vary the treatment effect on the outcome $\beta^y_{k,A}$, corresponding to $\beta^y_{k,A} > 0$, $\beta^y_{k,A} = 0$, and $\beta^y_{k,A} < 0$, and the effect of a stroke on the outcome $\beta^y_{k,L}$, corresponding to $\beta^y_{k,L} > 0$, $\beta^y_{k,L} = 0$, and $\beta^y_{k,L} < 0$. We also vary the effect of a stroke on the treatment propensity $\alpha_{k,\mathrm{L}}$ and the effect of treatment on stroke $\beta^\ell_{k,A}$. The overall goal is to assess the impact of baseline and time-varying confounding and if our method is a viable method of estimating the target parameter $\Psi_\tau(P)$. We compare our method to the naive method using the Cox model, which treats deviation from protocol as censoring. Furthermore, we discretize time into 8 intervals (Section 3.1), enabling comparisons with Longitudinal Targeted Maximum Likelihood Estimation (LTMLE) (Laan & Gruber, 2012). We consider both the debiased ICE estimator and the simple ICE estimator; the difference between them being that we add the efficient influence function to the first, which allows us to obtain doubly robust inference. Finally, we also compare with a continuous-time inverse probability weighting estimator which cannot be misspecified if one knows the true data generating mechanism.

Additionally, we vary sample size $n \in \{100, 2000, 500, 1000\}$. In all other cases, we fix $n = 1000$.

We thus consider three overall scenarios:
- **No baseline and time-varying confounding**.
- **No time-varying confounding but baseline confounding**.
- **Time-varying confounding**
- **Strong confounding**.

We highlight the interpretation of the most important parameters in the simulation mechanism:
- $\alpha_{k,\text{ age}}$: If positive: You will more likely be treated if you are older.
- $\alpha_{k,\mathrm{L}}$: If negative: You will be less likely to be treated if you have had a stroke.
- $\beta^x_{k,\mathrm{age}}$: If positive: The risk of having a stroke or primary event increases with age.
- $\beta^\ell_{k,A}$: If negative: The risk of having a stroke is lower if you are treated.
- $\beta^y_{k,L}$: If positive: The risk of having a primary event is higher if you have had a stroke.

Proposed values for the parameters are shown in Table 2. Strong confounding is considered in Table 3 in two different simulation settings.

Parameters	**Values**
α_{k0}	0.3
$\alpha_{k,\text{ age}}$	0.02
$\alpha_{k,\text{L}}$	$\mathbf{-0.2}$, $\overline{0}$, 0.2
$\beta^{y}_{k,\text{ age}}$	0.025
$\beta^{\ell}_{k,\text{ age}}$	0.015
$\beta^{y}_{k,A}$	$\mathbf{-0.3}$, $\underline{0}$, 0.3
$\beta^{\ell}_{k,A}$	$\mathbf{-0.2}$, $\underline{0}$, 0.2
$\beta^{y}_{k,\text{L}}$	−0.5, $\overline{0}$, **0.5**
λ^{y}_{k}	0.0001
λ^{ℓ}_{k}	0.001
γ_{age}	0
γ_{0}	0.005

Table 2: Simulation parameters for the simulation study. Each value is varied, holding the others fixed. The values with bold font correspond to the values used when fixed. The corresponding cases corresponding to no effect of baseline confounders are marked with an overline, and the cases corresponding to no effect of time-varying confounders are marked with an underline.

Parameters	Values
α_{k0}	0.3
$\alpha_{k,\text{ age}}$	0.02
$\alpha_{k,\text{L}}$	−0.6, 0.6
$\beta^{y}_{k,\text{ age}}$	0.025
$\beta^{\ell}_{k,\text{ age}}$	0.015
$\beta^{y}_{k,A}$	−0.8, 0.8
$\beta^{\ell}_{k,A}$	−0.2
$\beta^{y}_{k,\text{L}}$	1
λ^{y}_{k}	0.0001
λ^{ℓ}_{k}	0.001
γ_{age}	0
γ_{0}	0.005

Table 3: Simulation parameters for the strong confounding simulation study. Each value is varied, holding the others fixed.

3.1. Discretizing time

We briefly illustrate how to discretize the time horizon into K intervals, with time horizon τ, representing the usual longitudinal setting. Let $t_k = k \times \frac{\tau}{K}$ for $k = 1, ..., K$.

Put

$$\begin{aligned} Y_k &= N^y(t_k), \\ L_k &= L(t_k), \\ A_k &= A(t_k). \end{aligned}$$

Our data set then consists of

$$O = (\text{age}, L(0), A(0), Y_1, L_1, A_1, ..., Y_{K-1}, L_{K-1}, A_{K-1}, Y_K)$$

3.2. Nuisance parameter modeling

To apply our debiased ICE estimator in the uncensored situation, we need to estimate two types of nuisance parameters:

1. The treatment propensity $\pi_k\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ for each k.
2. The conditional counterfactual probabilities $\bar{Q}^g_{k,\tau}$ for each k.

For the treatment propensities, we consider correctly specified models using logistic regression.

For modeling the conditional counterfactual probabilities $\bar{Q}^g_{k,\tau}$, we a generalized linear model (GLM) with the option `family = quasibinomial()`, using no interactions in the history, as discussed in Section 2.1.

For the LTMLE procedure, we use an undersmoothed LASSO (Tibshirani, 1996) estimator.

3.3. Censoring

We consider a simulation involving *completely* independent censoring. Concretely, the censoring variable is simply generated as $C \sim \text{Exp}(\lambda_c)$. The processes under considerations are then observed up to this censoring time. We vary $\lambda_c \in \{0.0002, 0.0005, 0.0008\}$. The overall purpose is to determine:

1. If the ICE algorithm can be used to estimate the target parameter $\Psi_\tau(P)$ in the case of censoring.
2. What model should be used to estimate the conditional counterfactual probabilities $\bar{Q}^g_{k,\tau}$?
3. Figure out whether or not we need also to debias the *so-called* censoring martingale.

Here, we consider a linear model, the scaled quasibinomial `glm` (this means that we divide with the largest value of $\hat{Z}^k_a$ in the sample so that it is scaled down to $[0, 1]$; afterwards the predictions are rescaled with the largest value), a tweedie `glm` (the pseudo-outcomes Z^a_k may appear marginally as a mixture of a continuous random variable and a point mass at 0) as estimators of the conditional counterfactual probabilities $\bar{Q}^g_{k,\tau}$.

We consider only two parameter settings for the censoring martingale as outlined in Table 4.

Parameters	**Values**
α_{k0}	0.3
$\alpha_{k,\text{ age}}$	0.02
$\alpha_{k,\text{L}}$	0, −0.2
$\beta^y_{k,\text{ age}}$	0.025
$\beta^\ell_{k,\text{ age}}$	0.015
$\beta^y_{k,A}$	0, −0.3
$\beta^\ell_{k,A}$	−0.2
$\beta^y_{k,\text{L}}$	0, 0.5
λ^y_k	0.0001
λ^ℓ_k	0.001
γ_{age}	0
γ_0	0.005

Table 4: Simulation parameters for the censoring simulation study. Each value is varied, holding the others fixed.

4. Results

Here, the results are presented in a fairly unstructured format. In the tables, we report the mean squared error (MSE), mean bias, standard deviation of the estimates, and the mean of the estimated standard error, as well as coverage of 95% confidence intervals.

Overall conclusions for the uncensored case:

1. The debiased ICE-IPCW estimator works well in all cases considered with respect to bias. Moreover, it is unbiased in cases with substantial time-varying confounding and performs better than both the naive Cox model and the LTMLE estimator.
2. Standard errors are generally good, but appear less unbiased. Is this due to misspecification of the nuisance parameters $\bar{Q}^g_{k,\tau}$?
3. LTMLE standard errors are generally a little bit smaller than the debiased ICE-IPCW standard errors.
4. In the case with no confounding and $\beta^y_A = 0.3$, no method appears to be unbiased. Possible reasons: Either the simulation mechanism is too extreme or the risk of the event is too low (within the time horizon).

Overall conclusions for the censored case:

1. All choices of λ^c and all choices of nuisance parameter models seem to give unbiased estimates.
2. Standard errors appear to be (slightly) conservative. Standard errors for the tweedie model appear to be slightly higher than for the linear model or the scaled quasibinomial model.
3. The linear model appears to give the most unstable estimates. This can be seen in the boxplots for the (simple) ICE-IPCW estimator (Figure 19).

4.1. Tables

4.1.1. Uncensored

4.1.1.1. No confounding

β^y_A	Estimator	Coverage	MSE	Bias	sd(Est)	Mean($\widehat{\text{SE}}$)
-0.3	Debiased ICE-IPCW	0.942	0.0000681	0.0000409	0.00825	0.00822
-0.3	LTMLE (grid size = 8)					
-0.3	Naive Cox		0.0000661	0.0000378	0.00813	
-0.3	Inverse Probability Weighting		0.000068	0.0000407	0.00825	
-0.3	ICE-IPCW		0.0000681	0.0000487	0.00825	
0	Debiased ICE-IPCW	0.949	0.0000886	0.000172	0.00941	0.00944
0	LTMLE (grid size = 8)					
0	Naive Cox		0.0000857	0.000112	0.00926	
0	Inverse Probability Weighting		0.0000885	0.000171	0.00941	
0	ICE-IPCW		0.0000884	0.000177	0.0094	
0.3	Debiased ICE-IPCW	0.95	0.000116	0.0000599	0.0108	0.0108
0.3	LTMLE (grid size = 8)	0.945	0.000109	-0.0007	0.0104	0.0104
0.3	Naive Cox		0.000112	0.0000182	0.0106	
0.3	Inverse Probability Weighting		0.000116	0.0000591	0.0108	
0.3	ICE-IPCW		0.000116	0.0000643	0.0108	

4.1.1.2. No time-varying confounding

β^y_A	Estimator	Coverage	MSE	Bias	sd(Est)	Mean($\widehat{\text{SE}}$)
-0.3	Debiased ICE-IPCW	0.947	0.000252	0.0000732	0.0159	0.0158
-0.3	LTMLE (grid size = 8)	0.946	0.000245	0.00217	0.0155	0.0154
-0.3	Naive Cox		0.000244	-0.0000436	0.0156	
-0.3	Inverse Probability Weighting		0.000252	0.0000752	0.0159	
-0.3	ICE-IPCW		0.000252	0.0000597	0.0159	
0	Debiased ICE-IPCW	0.946	0.000296	0.00055	0.0172	0.0171
0	LTMLE (grid size = 8)	0.947	0.000284	0.000553	0.0168	0.0167
0	Naive Cox		0.000287	0.000419	0.0169	
0	Inverse Probability Weighting		0.000296	0.00056	0.0172	
0	ICE-IPCW		0.000295	0.000555	0.0172	
0.3	Debiased ICE-IPCW	0.949	0.000323	0.0000806	0.018	0.018
0.3	LTMLE (grid size = 8)	0.946	0.000314	-0.00221	0.0176	0.0176
0.3	Naive Cox		0.000314	-0.000105	0.0177	
0.3	Inverse Probability Weighting		0.000324	0.00009	0.018	
0.3	ICE-IPCW		0.000323	0.0000917	0.018	

4.1.1.3. Strong time-varying confounding

β^y_A	α_L	Estimator	Coverage	MSE	Bias	sd(Est)	Mean($\widehat{\text{SE}}$)
-0.8	-0.6	Debiased ICE-IPCW	0.947	0.000315	-0.000168	0.0178	0.0177
-0.8	-0.6	LTMLE (grid size = 8)	0.898	0.00042	0.0115	0.017	0.0167
-0.8	-0.6	Naive Cox		0.000323	-0.00599	0.017	
-0.8	-0.6	Inverse Probability Weighting		0.000315	-0.000168	0.0178	
-0.8	-0.6	ICE-IPCW		0.000314	-0.000202	0.0177	
0.8	0.6	Debiased ICE-IPCW	0.95	0.000256	0.00014	0.016	0.016
0.8	0.6	LTMLE (grid size = 8)	0.951	0.000261	-0.00315	0.0159	0.0161
0.8	0.6	Naive Cox		0.000273	0.00476	0.0158	
0.8	0.6	Inverse Probability Weighting		0.000256	0.000129	0.016	
0.8	0.6	ICE-IPCW		0.000255	0.000124	0.016	

4.1.1.4. Varying effects (A on Y, L on Y, A on L, L on A)

β^y_A	Estimator	Coverage	MSE	Bias	sd(Est)	Mean($\widehat{\text{SE}}$)
-0.3	Debiased ICE-IPCW	0.951	0.0003	-0.00000649	0.0173	0.0174
-0.3	LTMLE (grid size = 8)	0.948	0.00029	0.00296	0.0168	0.0169
-0.3	Naive Cox		0.000291	-0.00158	0.017	
-0.3	Inverse Probability Weighting		0.0003	-0.00000138	0.0173	
-0.3	ICE-IPCW		0.0003	-0.0000142	0.0173	
0	Debiased ICE-IPCW	0.951	0.000336	-0.000497	0.0183	0.0184

0	LTMLE (grid size = 8)	0.952	0.000316	-0.000303	0.0178	0.0178
0	Naive Cox		0.000327	-0.00231	0.0179	
0	Inverse Probability Weighting		0.000336	-0.000487	0.0183	
0	ICE-IPCW		0.000336	-0.000491	0.0183	
0.3	Debiased ICE-IPCW	0.95	0.000355	0.000395	0.0188	0.0188
0.3	LTMLE (grid size = 8)	0.948	0.00034	-0.00228	0.0183	0.0183
0.3	Naive Cox		0.000347	-0.00166	0.0185	
0.3	Inverse Probability Weighting		0.000355	0.000401	0.0188	
0.3	ICE-IPCW		0.000354	0.000417	0.0188	

β^y_L	**Estimator**	**Coverage**	**MSE**	**Bias**	**sd(Est)**	**Mean($\widehat{\text{SE}}$)**
-0.5	Debiased ICE-IPCW	0.948	0.000221	0.000198	0.0149	0.0147
-0.5	LTMLE (grid size = 8)	0.952	0.000212	0.00185	0.0144	0.0144
-0.5	Naive Cox		0.000216	0.00119	0.0147	
-0.5	Inverse Probability Weighting		0.000221	0.000199	0.0149	
-0.5	ICE-IPCW		0.00022	0.000194	0.0148	
0	Debiased ICE-IPCW	0.948	0.000257	-0.0000468	0.016	0.016
0	LTMLE (grid size = 8)	0.948	0.000249	0.00224	0.0156	0.0156
0	Naive Cox		0.000248	-0.000156	0.0157	
0	Inverse Probability Weighting		0.000257	-0.0000423	0.016	
0	ICE-IPCW		0.000257	-0.000044	0.016	
0.5	Debiased ICE-IPCW	0.951	0.0003	-0.00000649	0.0173	0.0174
0.5	LTMLE (grid size = 8)	0.948	0.00029	0.00296	0.0168	0.0169
0.5	Naive Cox		0.000291	-0.00158	0.017	
0.5	Inverse Probability Weighting		0.0003	-0.00000138	0.0173	
0.5	ICE-IPCW		0.0003	-0.0000142	0.0173	

β^L_A	**Estimator**	**Coverage**	**MSE**	**Bias**	**sd(Est)**	**Mean($\widehat{\text{SE}}$)**
-0.2	Debiased ICE-IPCW	0.951	0.0003	-0.00000649	0.0173	0.0174
-0.2	LTMLE (grid size = 8)	0.948	0.00029	0.00296	0.0168	0.0169
-0.2	Naive Cox		0.000291	-0.00158	0.017	
-0.2	Inverse Probability Weighting		0.0003	-0.00000138	0.0173	
-0.2	ICE-IPCW		0.0003	-0.0000142	0.0173	
0	Debiased ICE-IPCW	0.951	0.000304	0.000162	0.0174	0.0176
0	LTMLE (grid size = 8)	0.948	0.000298	0.00309	0.017	0.017
0	Naive Cox		0.000297	-0.00165	0.0171	
0	Inverse Probability Weighting		0.000304	0.000169	0.0174	
0	ICE-IPCW		0.000303	0.000153	0.0174	
0.2	Debiased ICE-IPCW	0.949	0.000317	0.000196	0.0178	0.0178
0.2	LTMLE (grid size = 8)	0.95	0.000305	0.00316	0.0172	0.0171
0.2	Naive Cox		0.000305	-0.00171	0.0174	

0.2	Inverse Probability Weighting		0.000317	0.000209	0.0178	
0.2	ICE-IPCW		0.000317	0.000198	0.0178	

α_L	Estimator	Coverage	MSE	Bias	sd(Est)	Mean($\widehat{\text{SE}}$)
-0.2	Debiased ICE-IPCW	0.951	0.0003	-0.00000649	0.0173	0.0174
-0.2	LTMLE (grid size = 8)	0.948	0.00029	0.00296	0.0168	0.0169
-0.2	Naive Cox		0.000291	-0.00158	0.017	
-0.2	Inverse Probability Weighting		0.0003	-0.00000138	0.0173	
-0.2	ICE-IPCW		0.0003	-0.0000142	0.0173	
0	Debiased ICE-IPCW	0.95	0.000292	-0.000207	0.0171	0.0171
0	LTMLE (grid size = 8)	0.947	0.000283	0.00257	0.0166	0.0167
0	Naive Cox		0.000284	-0.000892	0.0168	
0	Inverse Probability Weighting		0.000292	-0.000207	0.0171	
0	ICE-IPCW		0.000292	-0.000207	0.0171	
0.2	Debiased ICE-IPCW	0.951	0.000281	-0.000305	0.0168	0.0169
0.2	LTMLE (grid size = 8)	0.949	0.000276	0.00224	0.0164	0.0165
0.2	Naive Cox		0.000275	-0.000257	0.0166	
0.2	Inverse Probability Weighting		0.000281	-0.000289	0.0168	
0.2	ICE-IPCW		0.000281	-0.000311	0.0168	

4.1.1.5. Sample size

n	Estimator	Coverage	MSE	Bias	sd(Est)	Mean($\widehat{\text{SE}}$)
100	Debiased ICE-IPCW	0.934	0.00325	0.000382	0.0571	0.0555
100	Inverse Probability Weighting		0.00326	0.000592	0.0571	
100	ICE-IPCW		0.00316	0.000521	0.0562	
200	Debiased ICE-IPCW	0.947	0.00152	-0.000339	0.039	0.039
200	Inverse Probability Weighting		0.00152	-0.000304	0.039	
200	ICE-IPCW		0.00151	-0.000265	0.0388	
500	Debiased ICE-IPCW	0.95	0.000599	0.000325	0.0245	0.0246
500	Inverse Probability Weighting		0.000599	0.000354	0.0245	
500	ICE-IPCW		0.000598	0.000344	0.0245	
1000	Debiased ICE-IPCW	0.952	0.000303	0.000142	0.0174	0.0174
1000	Inverse Probability Weighting		0.000303	0.000139	0.0174	
1000	ICE-IPCW		0.000302	0.000136	0.0174	

4.1.2. Censored

β^y_A	β^y_L	α_L	λ_c	Model type	Estimator	Coverage	MSE	Bias	sd(Est)	Mean($\widehat{\text{SE}}$)
-0.3	0.5	-0.2	0.0002	Scaled Quasi-binomial	Debiased ICE-IPCW	0.952	0.000331	0.000188	0.0182	0.0184

-0.3	0.5	-0.2	0.0002	Scaled Quasi-binomial	Inverse Probability Weighting		0.000331	0.000197	0.0182	
-0.3	0.5	-0.2	0.0002	Scaled Quasi-binomial	ICE-IPCW		0.00033	0.000171	0.0182	
-0.3	0.5	-0.2	0.0002	Tweedie	Debiased ICE-IPCW	0.953	0.00033	0.0000904	0.0182	0.0184
-0.3	0.5	-0.2	0.0002	Tweedie	Inverse Probability Weighting		0.000329	0.0000815	0.0181	
-0.3	0.5	-0.2	0.0002	Tweedie	ICE-IPCW		0.000328	0.000288	0.0181	
-0.3	0.5	-0.2	0.0002	Linear model	Debiased ICE-IPCW	0.95	0.000335	-0.000294	0.0183	0.0184
-0.3	0.5	-0.2	0.0002	Linear model	Inverse Probability Weighting		0.000335	-0.000296	0.0183	
-0.3	0.5	-0.2	0.0002	Linear model	ICE-IPCW		0.000327	0.00113	0.018	
-0.3	0.5	-0.2	0.0005	Scaled Quasi-binomial	Debiased ICE-IPCW	0.95	0.000388	-0.0000739	0.0197	0.0199
-0.3	0.5	-0.2	0.0005	Scaled Quasi-binomial	Inverse Probability Weighting		0.000388	-0.0000679	0.0197	
-0.3	0.5	-0.2	0.0005	Scaled Quasi-binomial	ICE-IPCW		0.000387	-0.0000568	0.0197	
-0.3	0.5	-0.2	0.0005	Tweedie	Debiased ICE-IPCW	0.955	0.000381	0.000187	0.0195	0.0199
-0.3	0.5	-0.2	0.0005	Tweedie	Inverse Probability Weighting		0.00038	0.000192	0.0195	
-0.3	0.5	-0.2	0.0005	Tweedie	ICE-IPCW		0.000377	0.000596	0.0194	
-0.3	0.5	-0.2	0.0005	Linear model	Debiased ICE-IPCW	0.951	0.000382	-0.0000409	0.0195	0.0199
-0.3	0.5	-0.2	0.0005	Linear model	Inverse Probability Weighting		0.000382	-0.0000387	0.0195	

-0.3	0.5	-0.2	0.0005	Linear model	ICE-IPCW		0.000375	0.00143	0.0193	
-0.3	0.5	-0.2	0.0008	Scaled Quasi-binomial	Debiased ICE-IPCW	0.953	0.000435	-0.0000717	0.0209	0.0215
-0.3	0.5	-0.2	0.0008	Scaled Quasi-binomial	Inverse Probability Weighting		0.000436	-0.0000649	0.0209	
-0.3	0.5	-0.2	0.0008	Scaled Quasi-binomial	ICE-IPCW		0.000434	0.0000358	0.0208	
-0.3	0.5	-0.2	0.0008	Tweedie	Debiased ICE-IPCW	0.953	0.000447	0.00000225	0.0212	0.0215
-0.3	0.5	-0.2	0.0008	Tweedie	Inverse Probability Weighting		0.000446	-0.0000101	0.0211	
-0.3	0.5	-0.2	0.0008	Tweedie	ICE-IPCW		0.000441	0.000636	0.021	
-0.3	0.5	-0.2	0.0008	Linear model	Debiased ICE-IPCW	0.956	0.000431	0.000211	0.0208	0.0215
-0.3	0.5	-0.2	0.0008	Linear model	Inverse Probability Weighting		0.000431	0.00022	0.0208	
-0.3	0.5	-0.2	0.0008	Linear model	ICE-IPCW		0.000425	0.00176	0.0205	
0	0	0	0.0002	Scaled Quasi-binomial	Debiased ICE-IPCW	0.952	0.000316	0.000438	0.0178	0.0179
0	0	0	0.0002	Scaled Quasi-binomial	Inverse Probability Weighting		0.000316	0.000445	0.0178	
0	0	0	0.0002	Scaled Quasi-binomial	ICE-IPCW		0.000316	0.000438	0.0178	
0	0	0	0.0002	Tweedie	Debiased ICE-IPCW	0.951	0.000317	0.000434	0.0178	0.018
0	0	0	0.0002	Tweedie	Inverse Probability Weighting		0.000316	0.000425	0.0178	
0	0	0	0.0002	Tweedie	ICE-IPCW		0.000315	0.000464	0.0178	

0	0	0	0.0002	Linear model	Debiased ICE-IPCW	0.949	0.000323	0.00048	0.018	0.0179
0	0	0	0.0002	Linear model	Inverse Probability Weighting		0.000323	0.000479	0.018	
0	0	0	0.0002	Linear model	ICE-IPCW		0.000321	0.00154	0.0179	
0	0	0	0.0005	Scaled Quasi-binomial	Debiased ICE-IPCW	0.953	0.000357	0.000352	0.0189	0.0193
0	0	0	0.0005	Scaled Quasi-binomial	Inverse Probability Weighting		0.000357	0.000359	0.0189	
0	0	0	0.0005	Scaled Quasi-binomial	ICE-IPCW		0.000356	0.000385	0.0189	
0	0	0	0.0005	Tweedie	Debiased ICE-IPCW	0.954	0.000364	0.00027	0.0191	0.0194
0	0	0	0.0005	Tweedie	Inverse Probability Weighting		0.000363	0.000252	0.0191	
0	0	0	0.0005	Tweedie	ICE-IPCW		0.000361	0.000368	0.019	
0	0	0	0.0005	Linear model	Debiased ICE-IPCW	0.955	0.00036	0.000512	0.019	0.0194
0	0	0	0.0005	Linear model	Inverse Probability Weighting		0.00036	0.000509	0.019	
0	0	0	0.0005	Linear model	ICE-IPCW		0.000359	0.00173	0.0189	
0	0	0	0.0008	Scaled Quasi-binomial	Debiased ICE-IPCW	0.955	0.000413	0.000149	0.0203	0.0209
0	0	0	0.0008	Scaled Quasi-binomial	Inverse Probability Weighting		0.000413	0.000155	0.0203	
0	0	0	0.0008	Scaled Quasi-binomial	ICE-IPCW		0.000411	0.000241	0.0203	
0	0	0	0.0008	Tweedie	Debiased ICE-IPCW	0.958	0.000412	0.0000457	0.0203	0.0209

0	0	0	0.0008	Tweedie	Inverse Probability Weighting		0.000411	0.0000306	0.0203	
0	0	0	0.0008	Tweedie	ICE-IPCW		0.000409	0.00025	0.0202	
0	0	0	0.0008	Linear model	Debiased ICE-IPCW	0.953	0.000421	0.00021	0.0205	0.0209
0	0	0	0.0008	Linear model	Inverse Probability Weighting		0.000421	0.000212	0.0205	
0	0	0	0.0008	Linear model	ICE-IPCW		0.000421	0.00159	0.0205	

4.2. Boxplots

4.2.1. Uncensored

4.2.1.1. No confounding

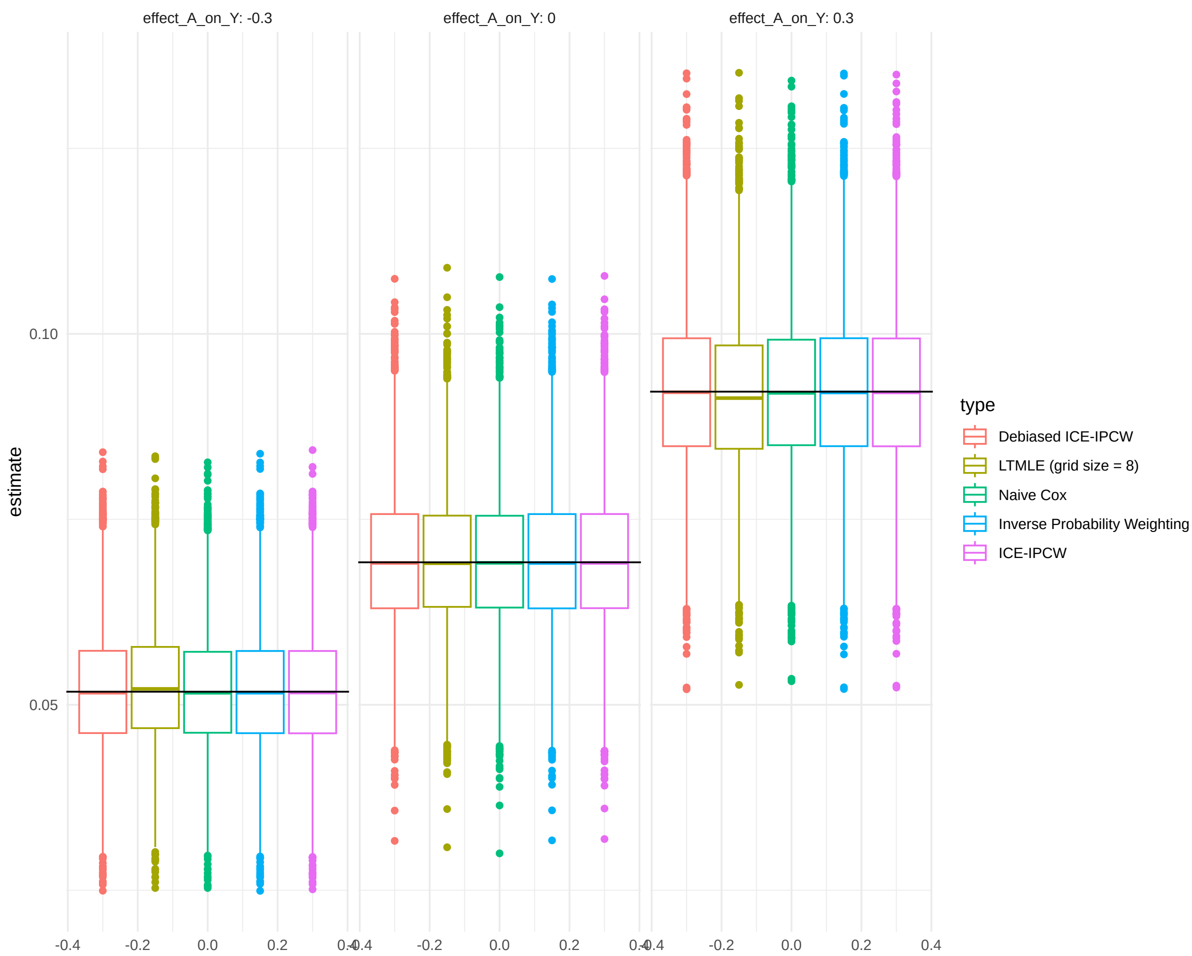

Figure 1: Boxplots of the results for the case without time confounding. The lines indicates the true value of the parameter.

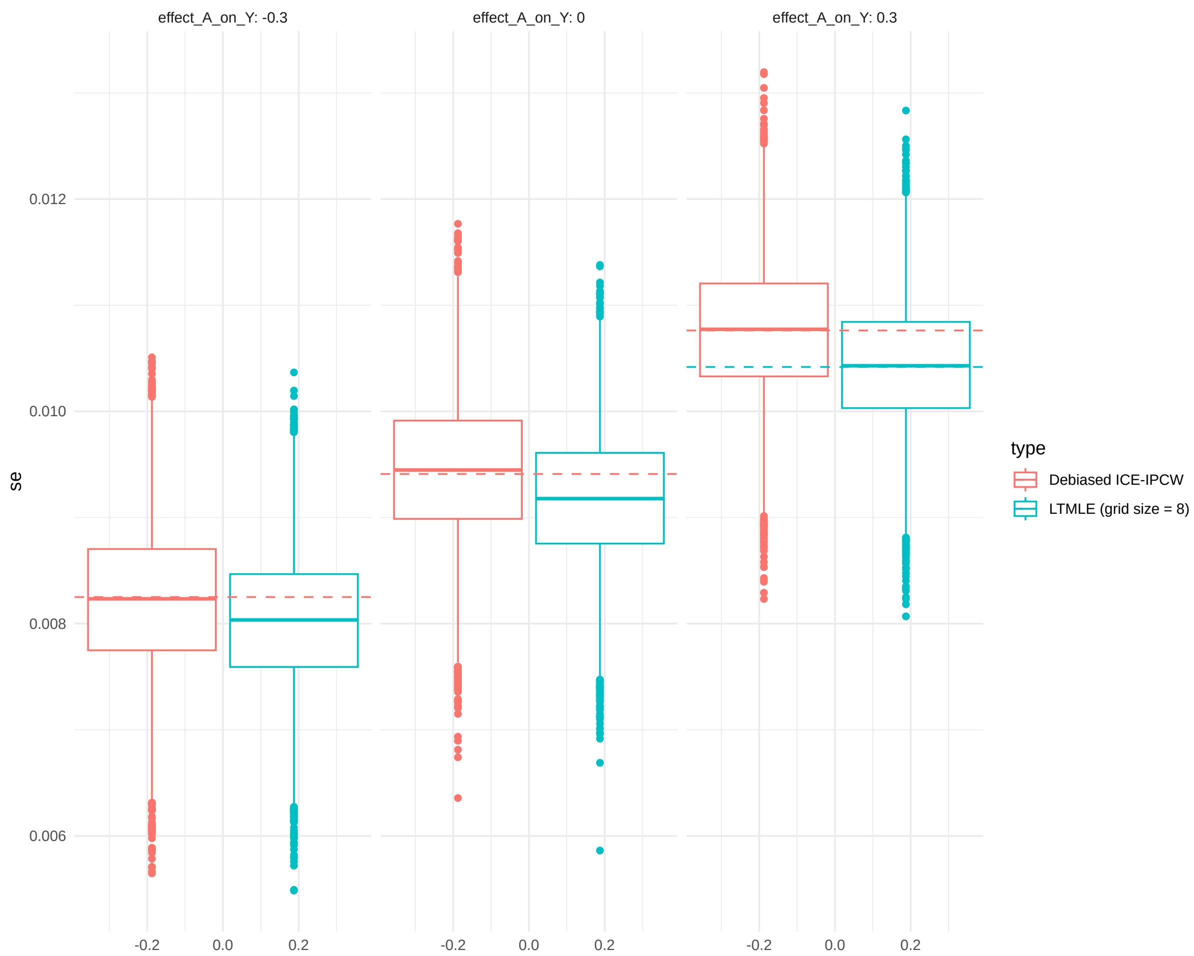

Figure 2: Boxplots of the standard errors for the case without time confounding. The red line indicates the empirical standard error of the estimates for each estimator.

4.2.1.2. No time-varying confounding

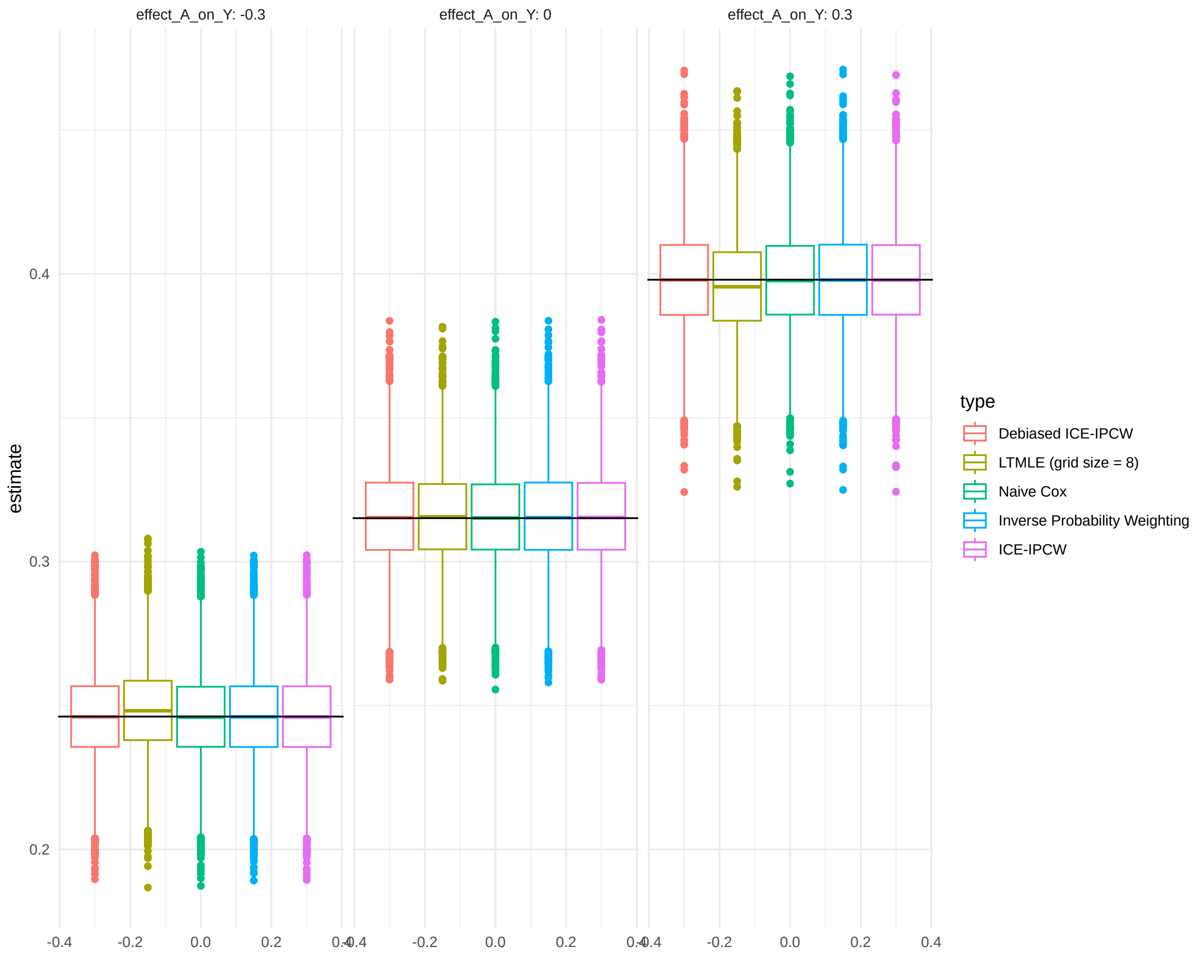

Figure 3: Boxplots of the results for the case without time confounding. The lines indicates the true value of the parameter.

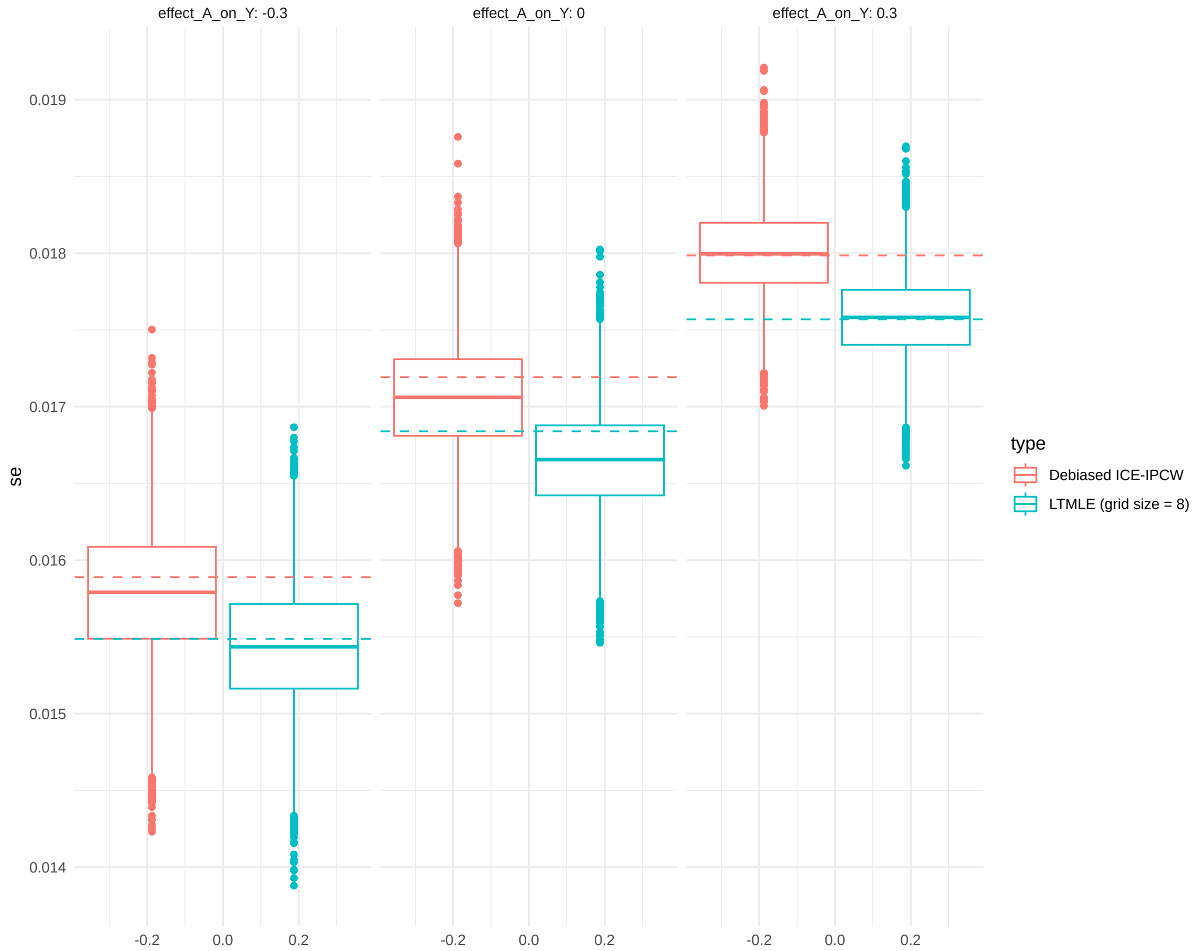

Figure 4: Boxplots of the standard errors for the case without time confounding. The red line indicates the empirical standard error of the estimates for each estimator.

4.2.1.3. Strong time-varying confounding

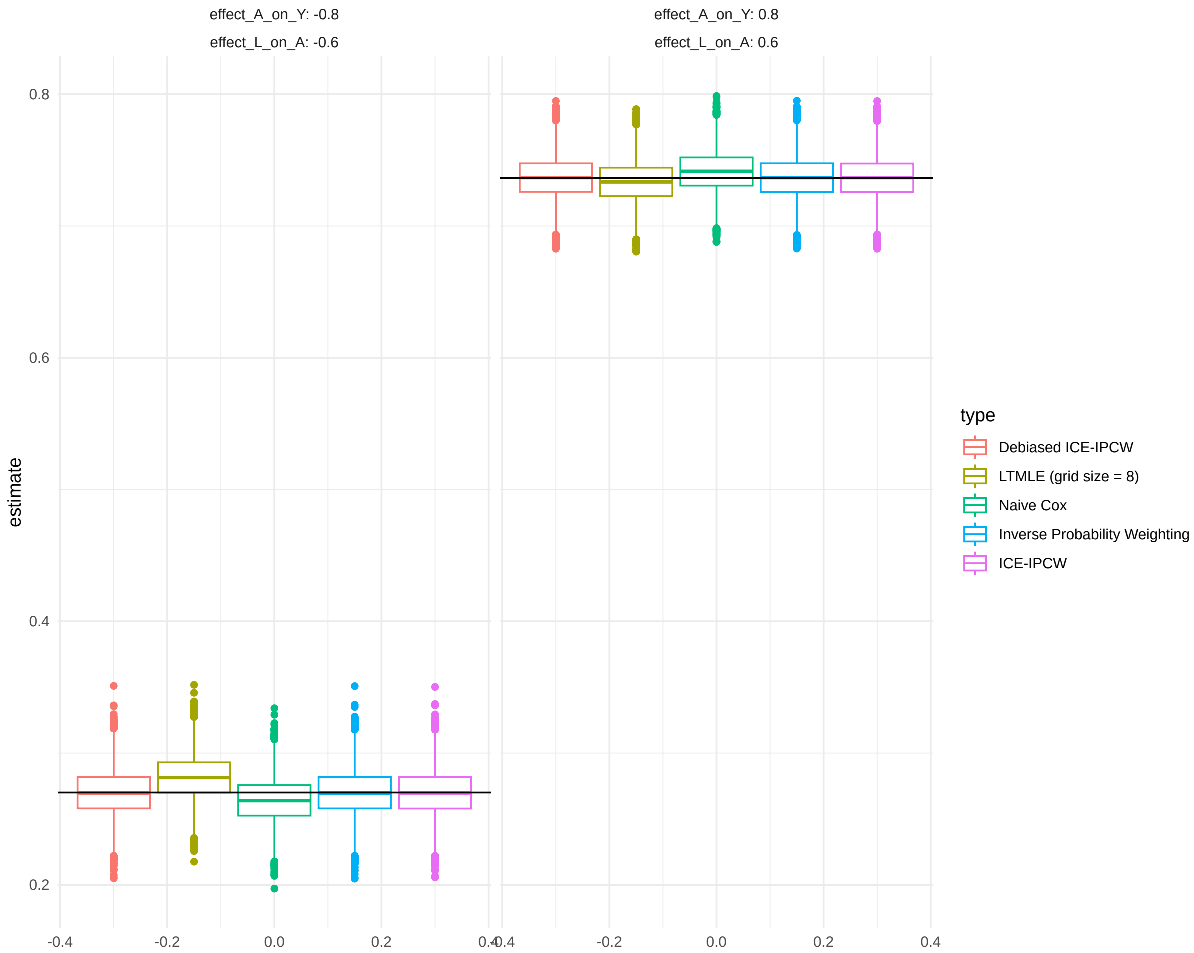

Figure 5: Boxplots of the results for the case with strong time confounding. The lines indicates the true value of the parameter.

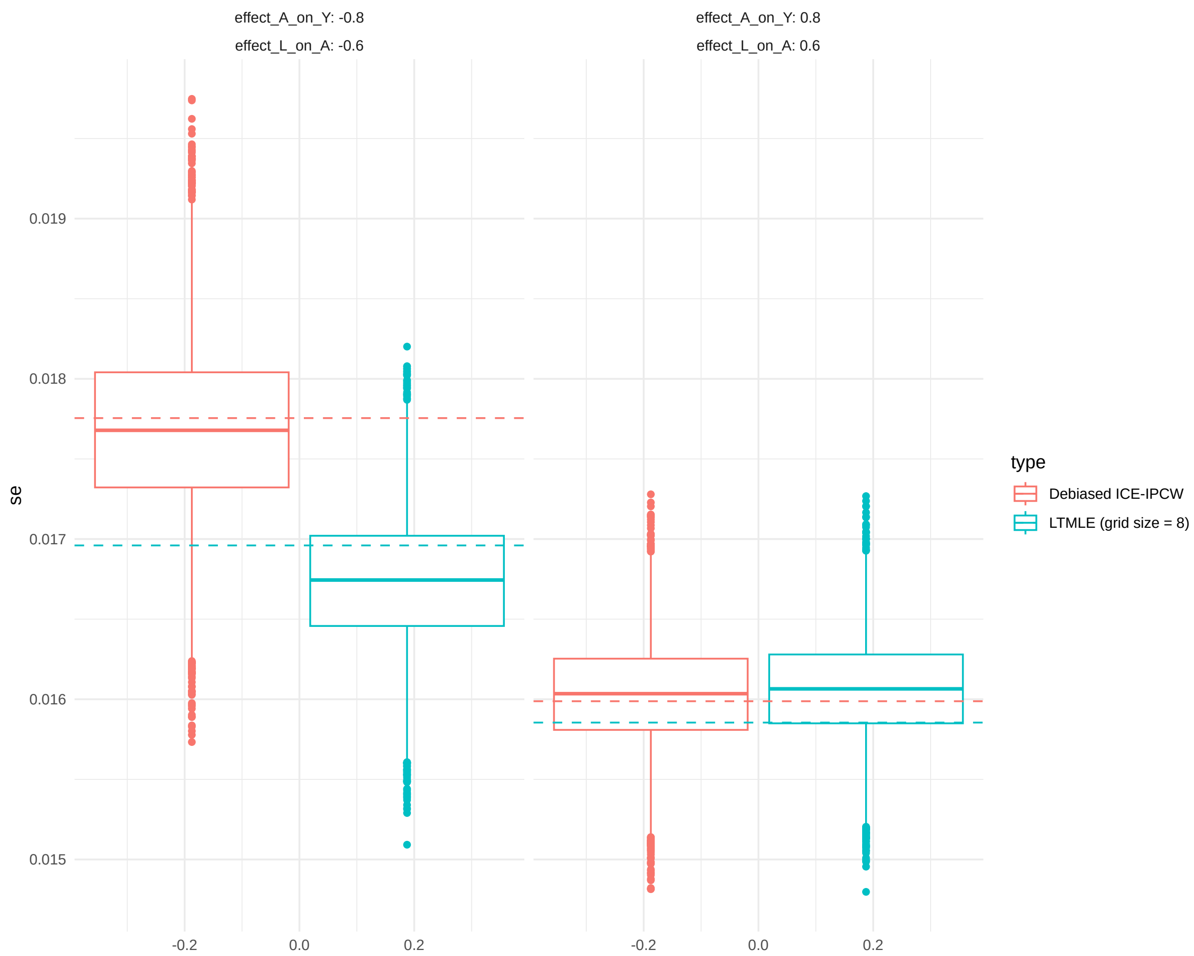

Figure 6: Boxplots of the standard errors for the case with strong time confounding. The red line indicates the empirical standard error of the estimates for each estimator.

4.2.1.4. Varying effects (A on Y, L on Y, A on L, L on A)

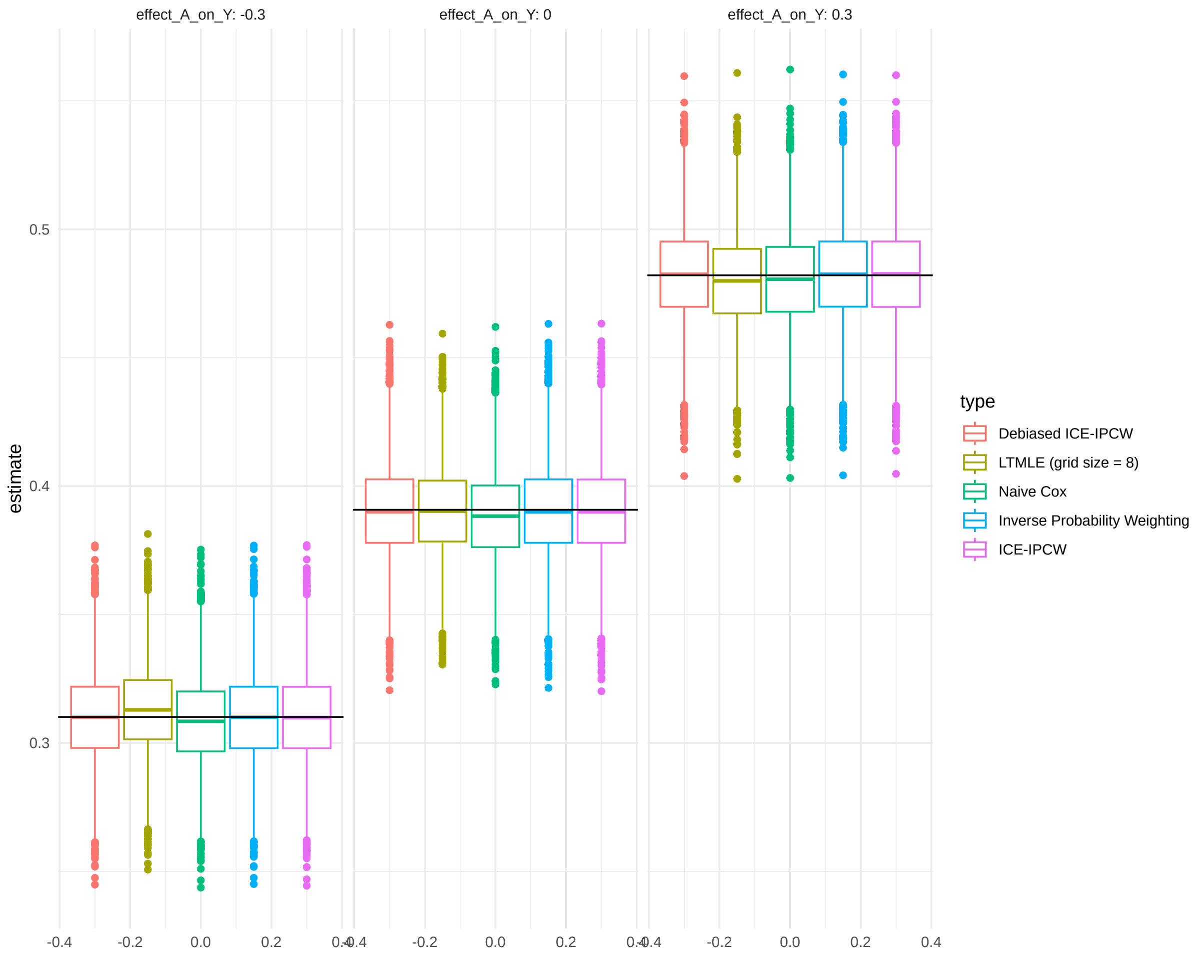

Figure 7: Boxplots of the results for the case with varying effect of A on Y. The lines indicates the true value of the parameter.

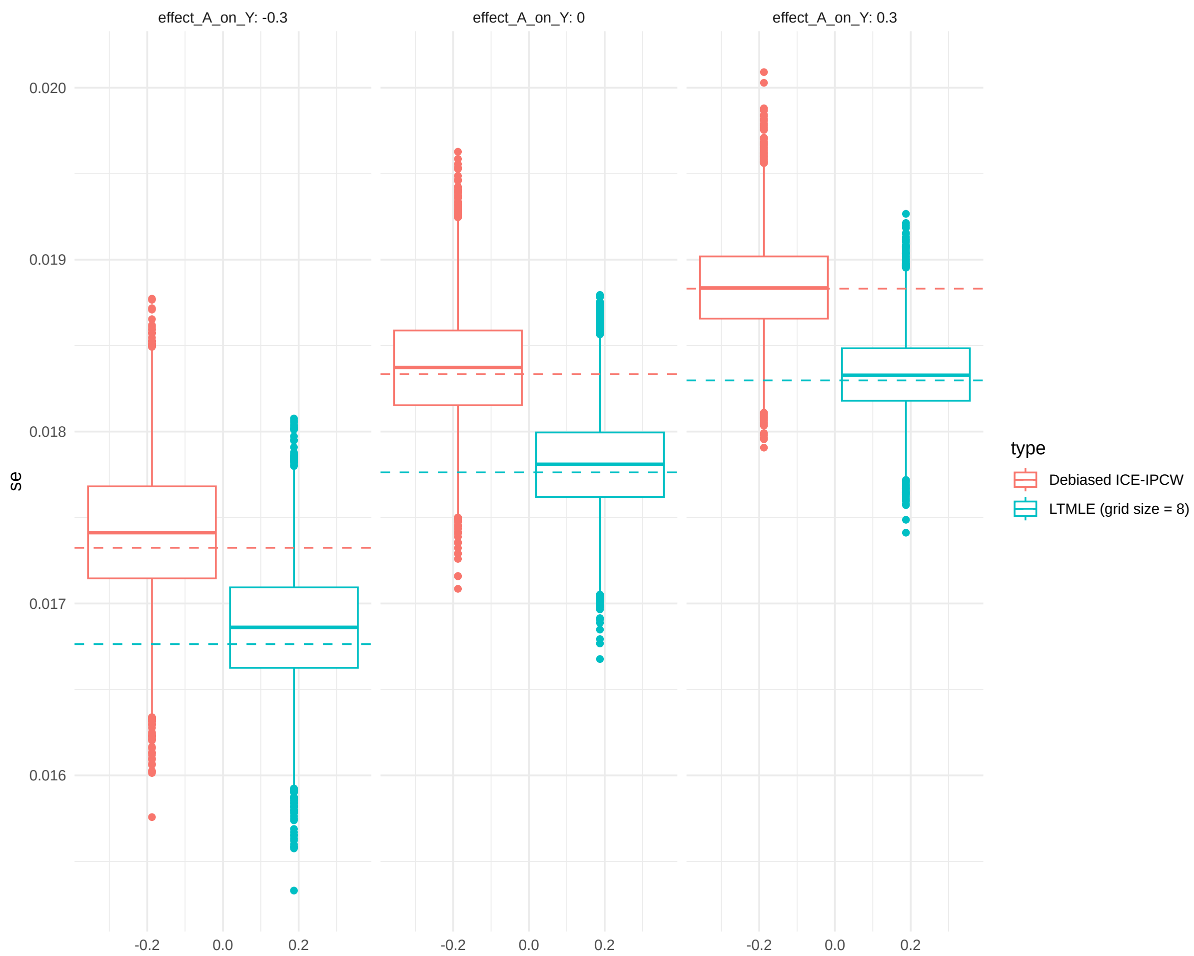

Figure 8: Boxplots of the standard errors for the case with varying effect of A on Y. The red line indicates the empirical standard error of the estimates for each estimator.

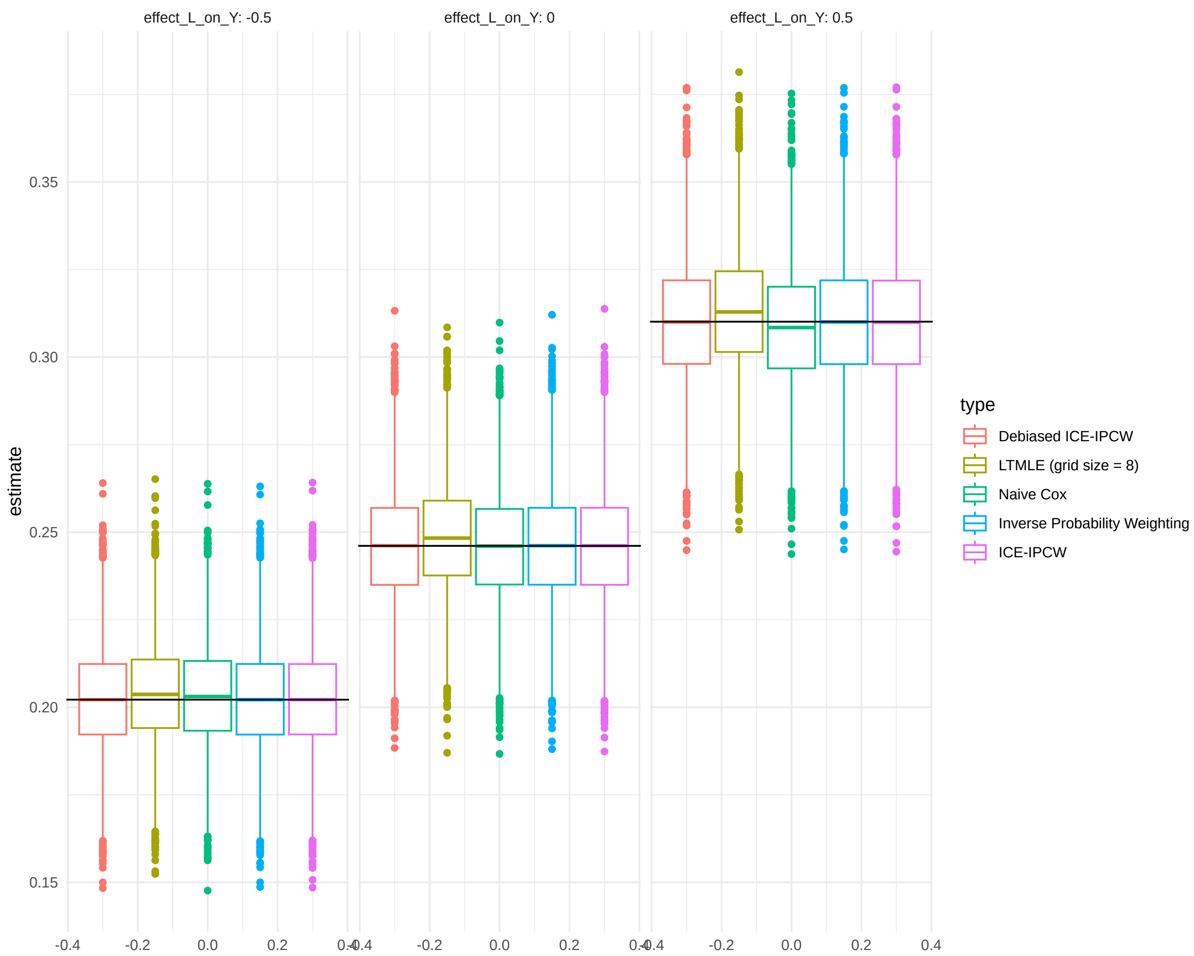

Figure 9: Boxplots of the results for the case with varying effect of L on Y. The lines indicates the true value of the parameter.

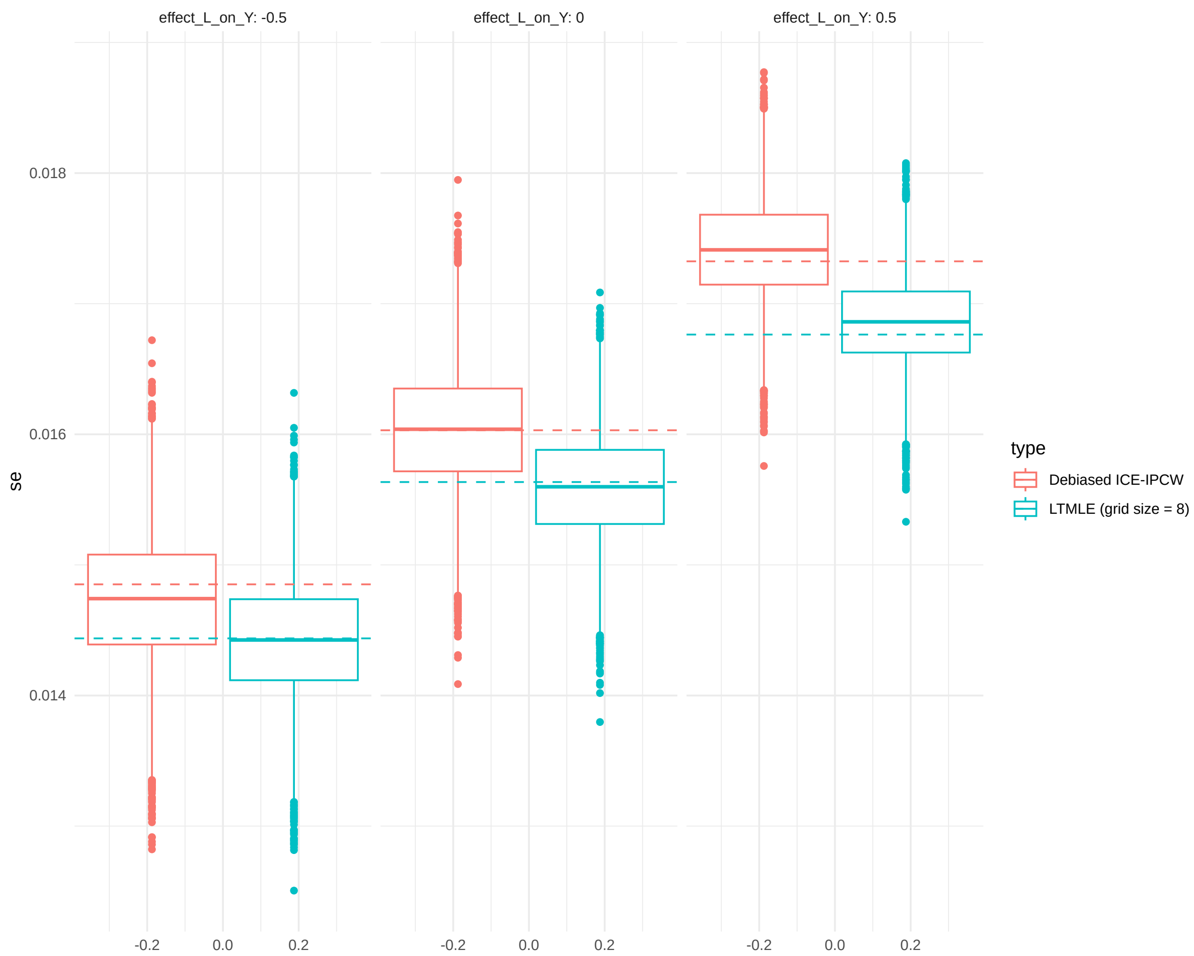

Figure 10: Boxplots of the standard errors for the case with varying effect of L on Y. The red line indicates the empirical standard error of the estimates for each estimator.

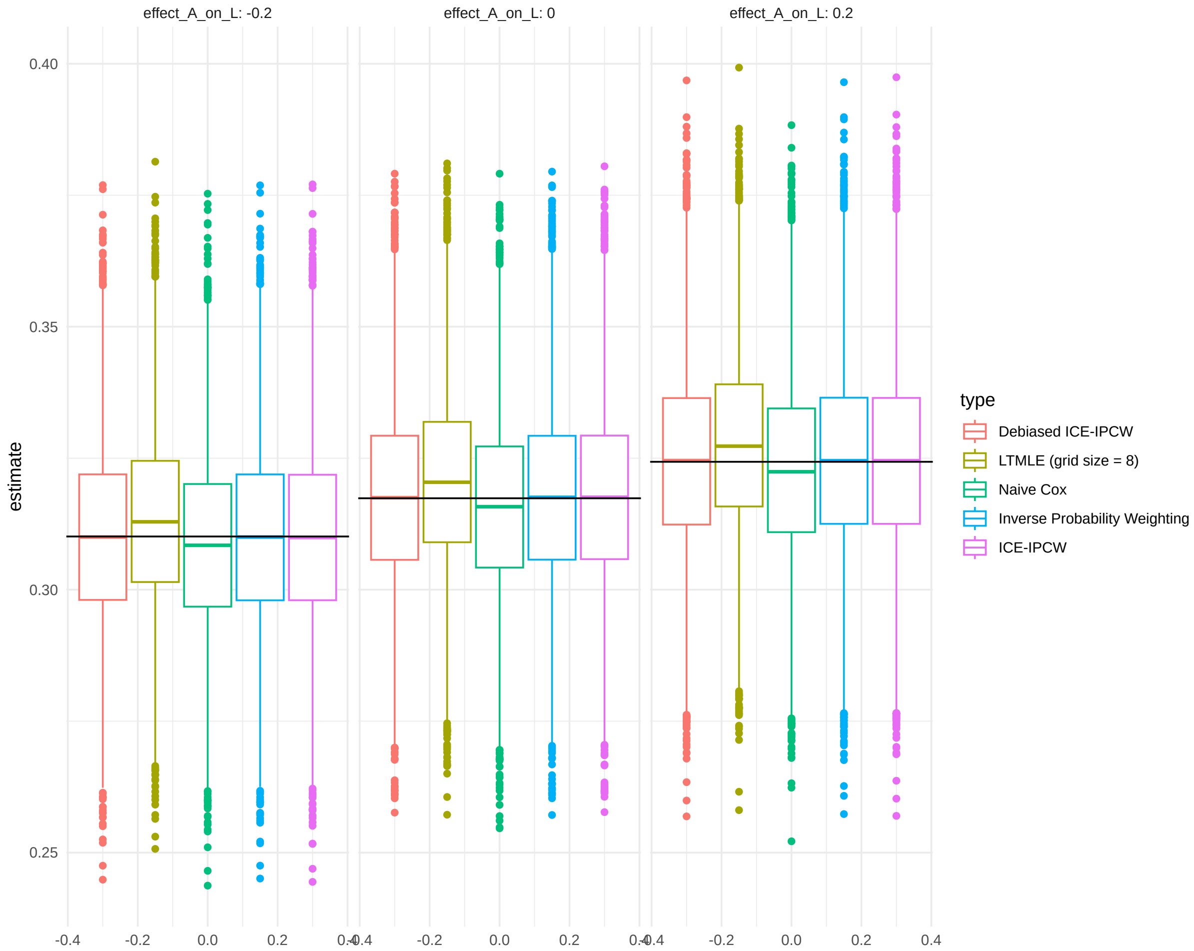

Figure 11: Boxplots of the results for the case with varying effect of A on L. The lines indicates the true value of the parameter.

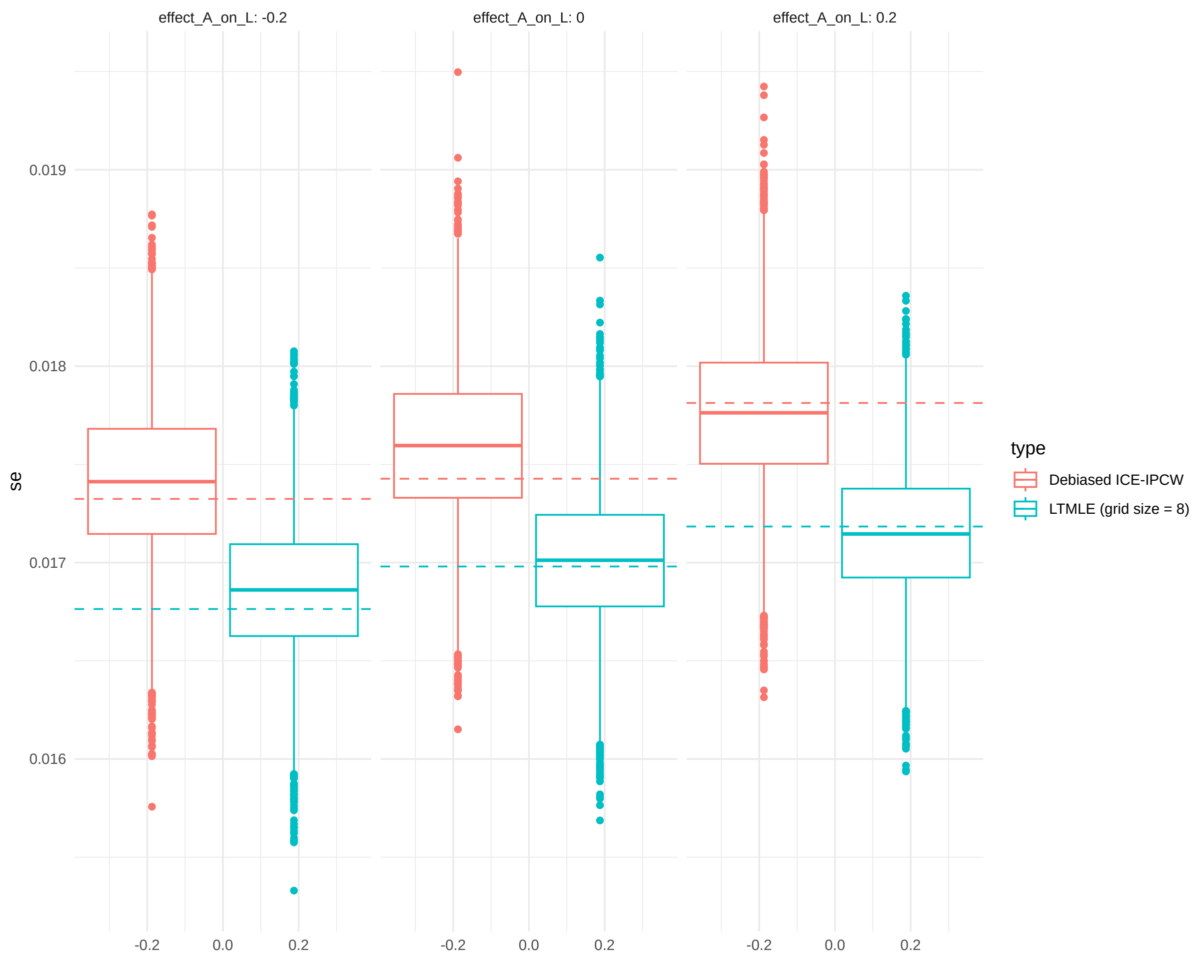

Figure 12: Boxplots of the standard errors for the case with varying effect of A on L. The red line indicates the empirical standard error of the estimates for each estimator.

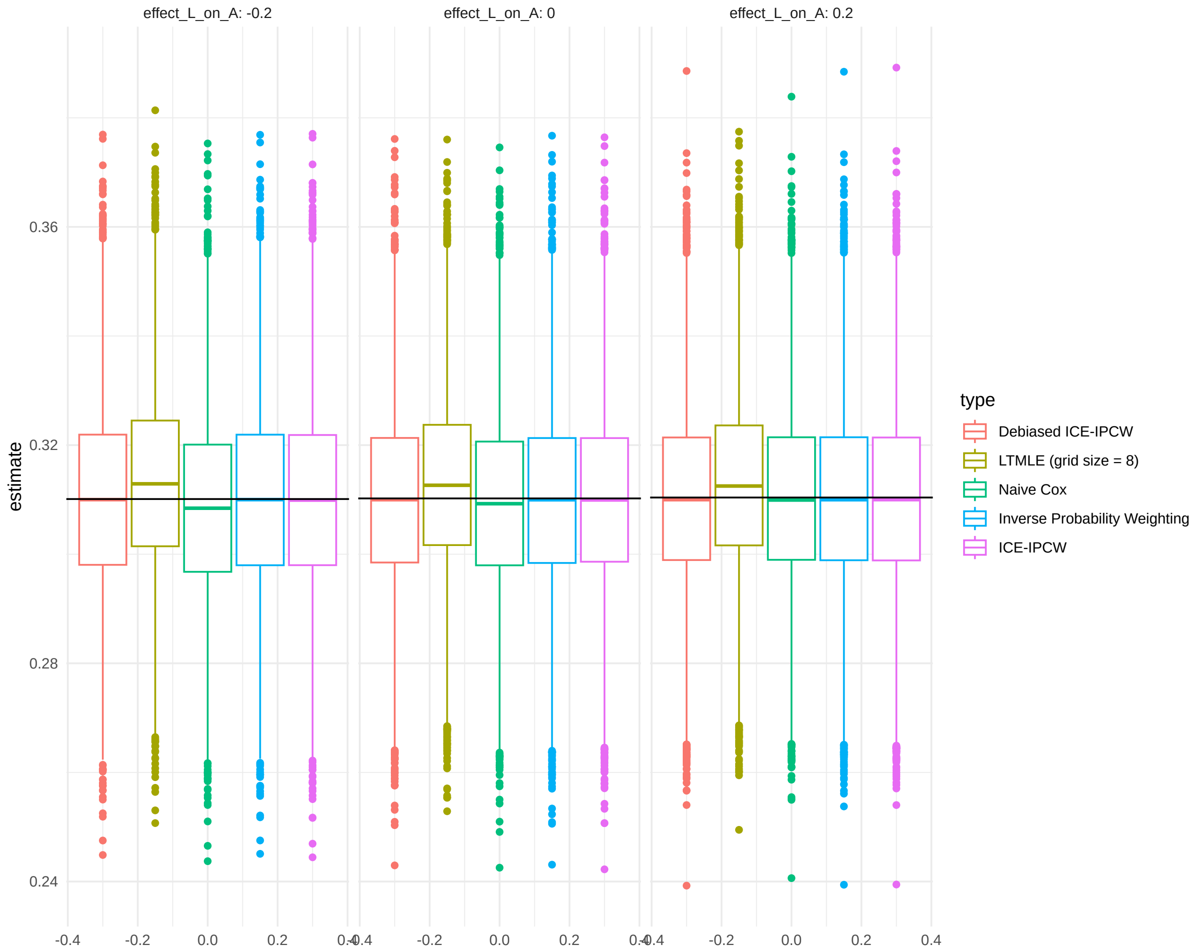

Figure 13: Boxplots of the results for the case with varying effect of L on A. The lines indicates the true value of the parameter.

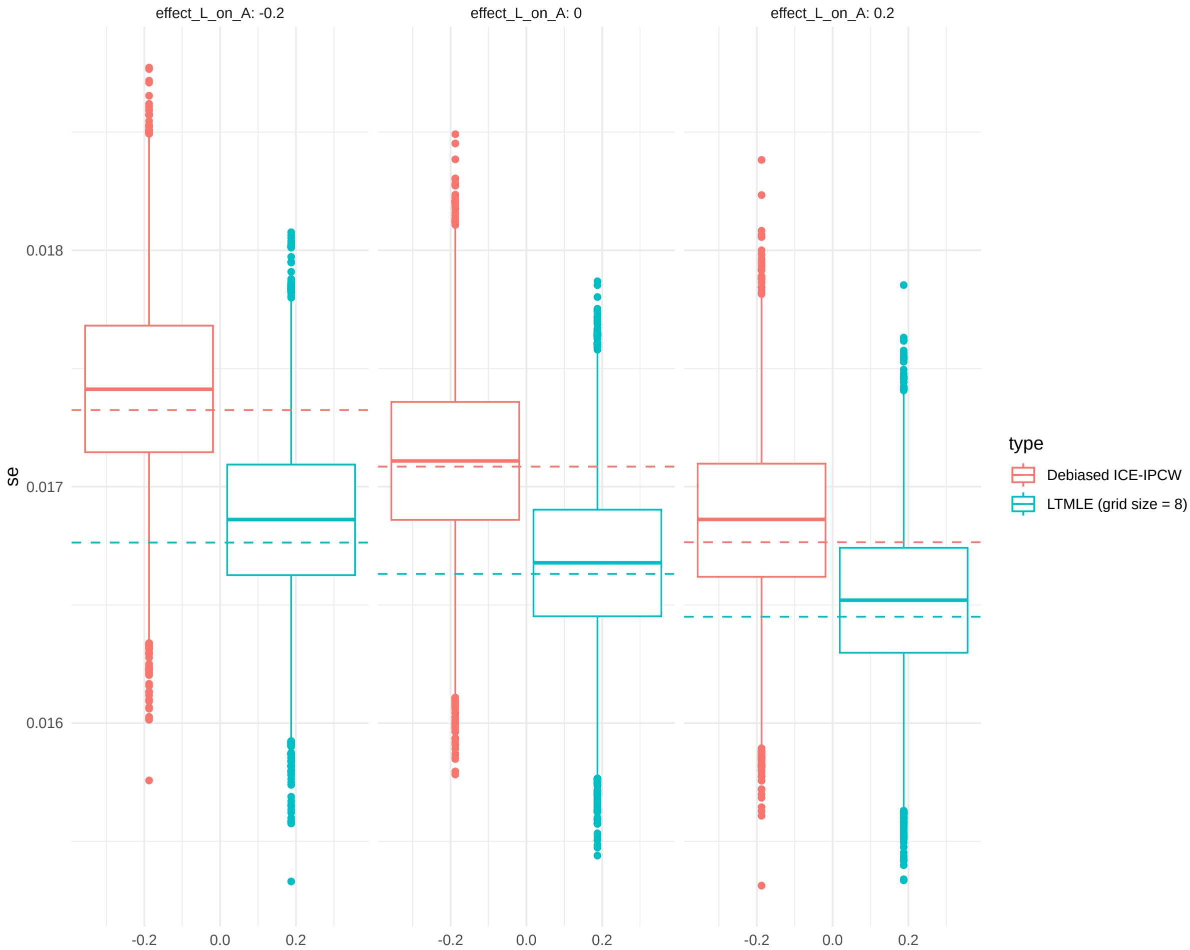

Figure 14: Boxplots of the standard errors for the case with varying effect of L on A. The red line indicates the empirical standard error of the estimates for each estimator.

4.2.1.5. Sample size

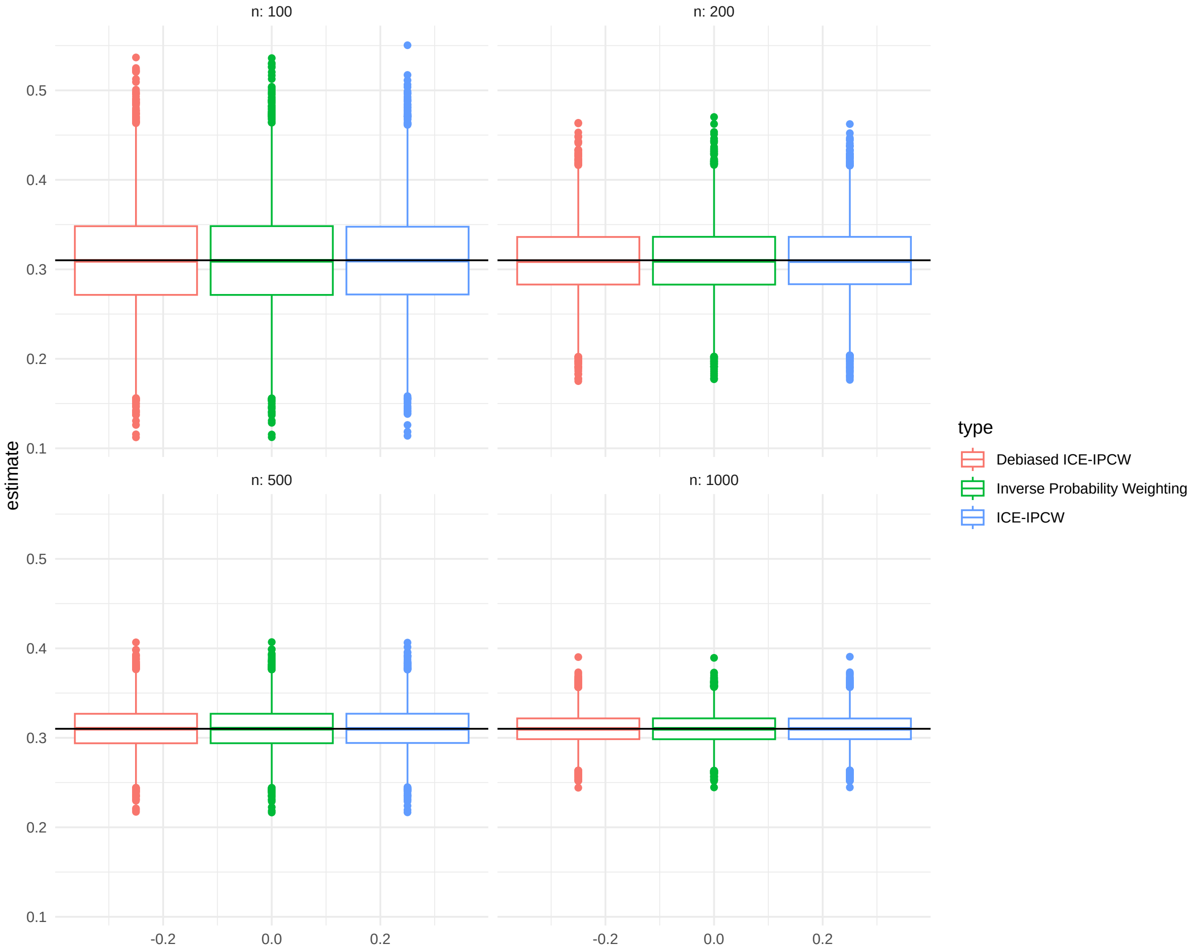

Figure 15: Boxplots of the results for the case with varying sample size. The lines indicates the true value of the parameter.

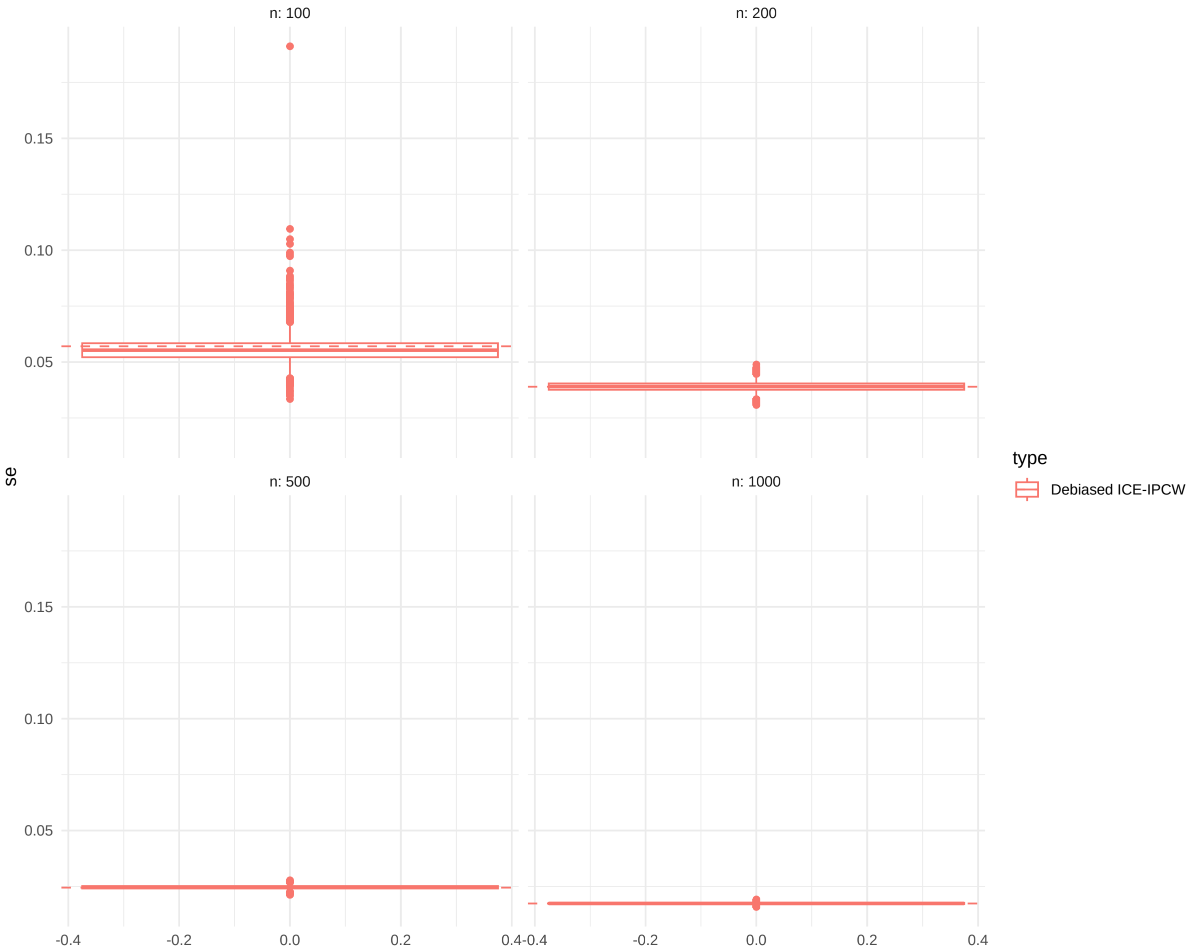

Figure 16: Boxplots of the standard errors for the case with varying sample size. The red line indicates the empirical standard error of the estimates for each estimator.

4.2.2. Censored

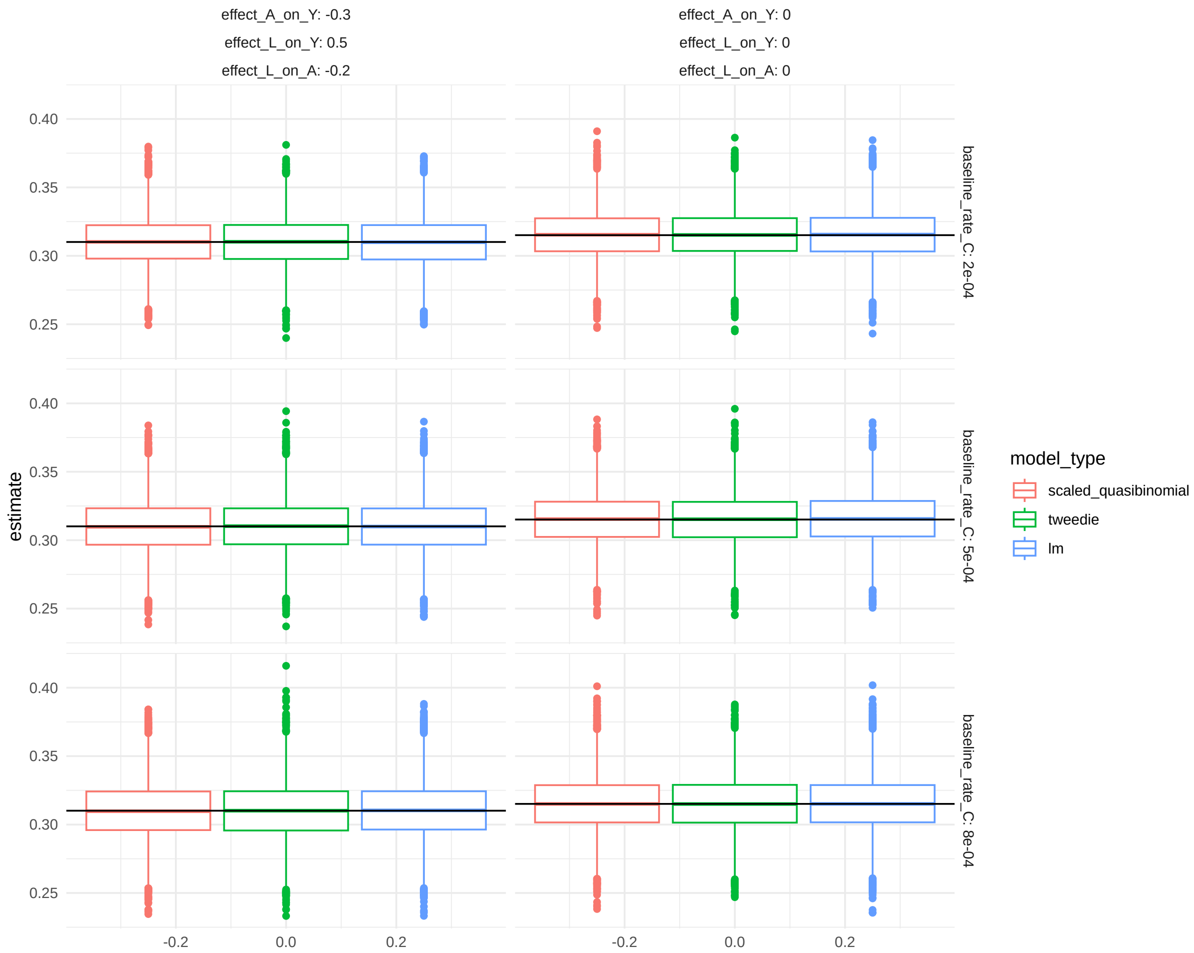

Figure 17: Boxplots of the results for the case with censoring. Different degrees of censoring are considered as well different model types for the pseudo-outcomes. Only the debiased ICE-IPCW estimator is shown.

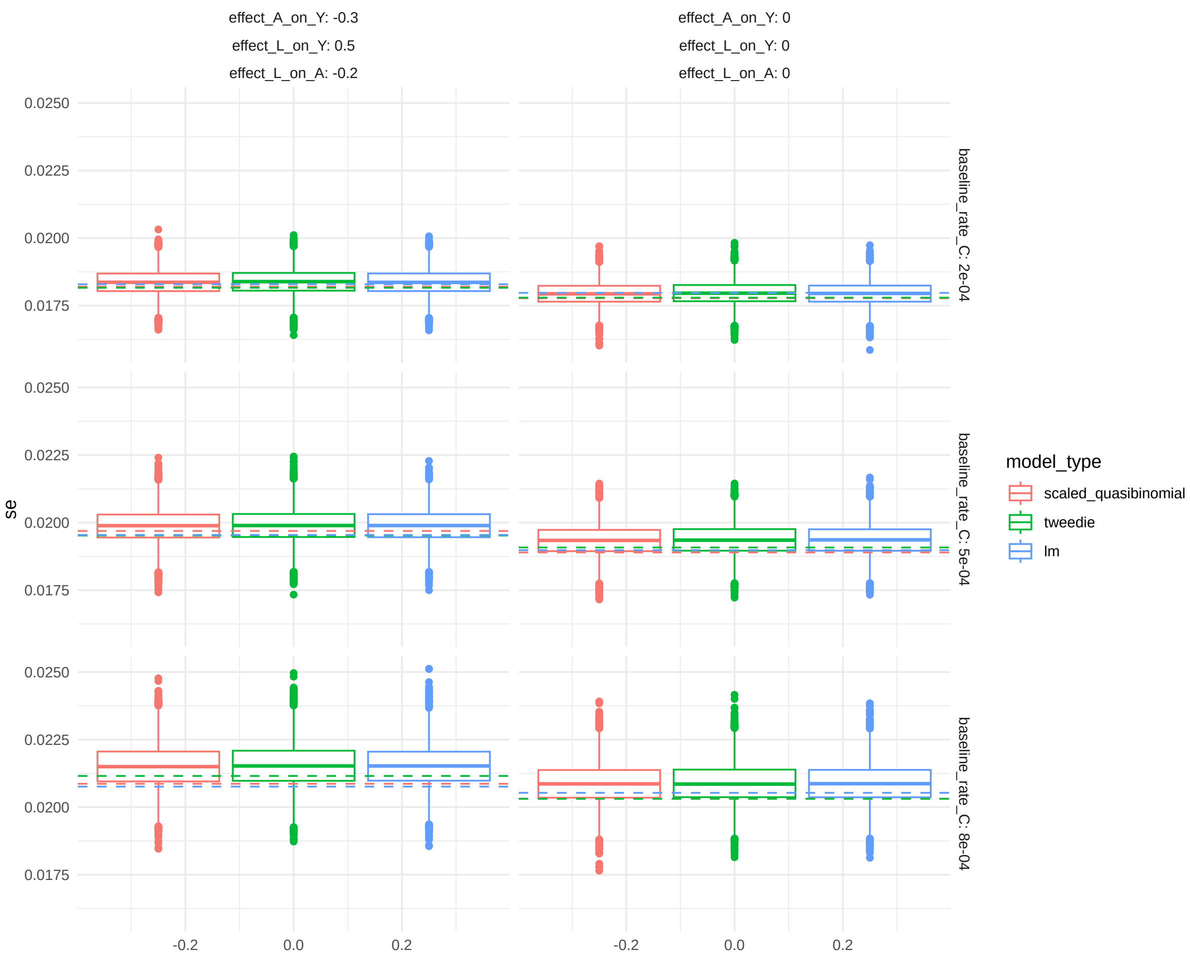

Figure 18: Boxplots of the standard errors for the case with censoring. The red line indicates the empirical standard error of the estimates for each estimator.

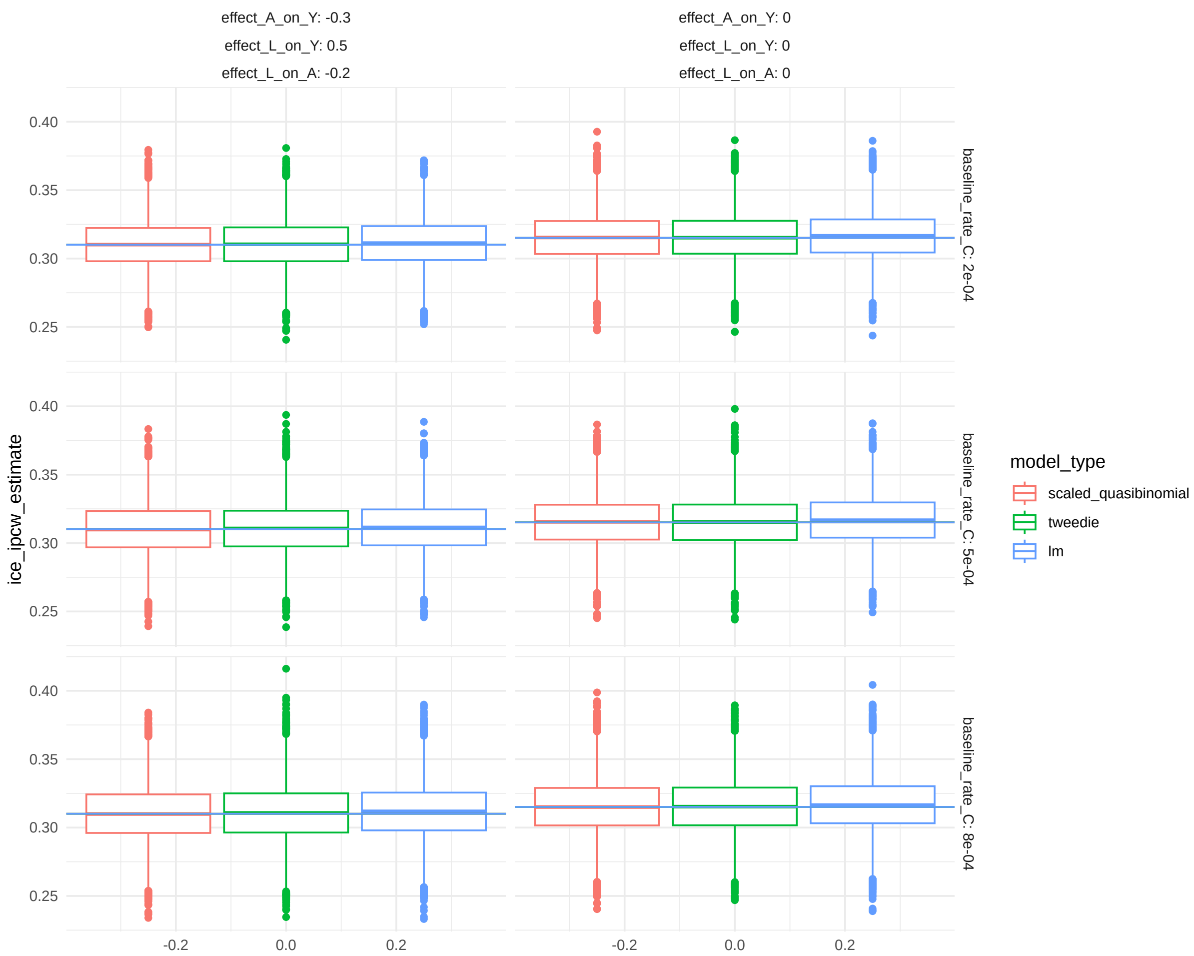

Figure 19: Boxplots of the results for the case with censoring. Different degrees of censoring are considered as well different model types for the pseudo-outcomes. Here, the (not debiased) ICE-IPCW estimator is shown.

5. Extensions of simulation procedure

In this section, we briefly outline some extensions of the simulation procedure, which can be used to simulate from more complex, realistic models. For instance, the simulation procedure outlined above does not use the timing of the events. It may therefore be interesting to see if the discrete time methods become even more biased in this case. We might also use a model in which the baseline hazard is not constant. We now consider an example.

Let T^ℓ be the time since the last stroke (i.e., 0 if stroke occurred as the previous event and $T_{(2)} - T_{(1)}$ if it happened as the first event and T^a denotes the time to the last treatment.

A realistic intensity for the primary event might be

$$\lambda^y(t) = \lambda_0^y(t) \exp\left(\beta_{\text{age}}^y \text{ age}\right) \exp\Big(\beta_L^y \exp\left(\beta_L^{y*}(t - T^\ell)\right) L(t-) + \beta_A^y \exp\left(\beta_A^{y*}(t - T^a)\right)(1 - A(t-)) \\ + \beta_Z^y \mathbb{1}\Big\{T_{(2)} < t\Big\} \mathbb{1}\Big\{\Delta_{(1)} = a\Big\} \mathbb{1}\Big\{\Delta_{(2)} = \ell\Big\}\Big) \mathbb{1}\{t \le T^y\}$$

Note that each term is zero if the corresponding event has not happened yet, so we do not condition on the future. For stroke, we may have $\beta_L^y > 0$ and $\beta_L^{y*} < 0$. The term for the stroke can now be interpreted: The intensity for death when a stroke occurs spikes at the time of the stroke, and then decays exponentially on the intensity scale, corresponding to the stroke having less of an impact on death as time passes. A similar interpretation can be given for treatment. The last term corresponds to there being an effect of the order in which the events happened after two events.

Simulating from this model is significantly more complicated, because we have to rely on numeric integration if we are to use a similar simulation procedure as we outlined previously (Section 5.1). Otherwise, we can use thinning algorithms such as Ogata's thinning algorithm (Ogata, 1981).

5.1. Simulating from the interevent time scale

The cumulative hazard-cause specific hazard of $S_{(k)}$ of the k'th event type x is given by

$$\left[\tau_{\text{end}} - T_{(k-1)}, 0\right) \ni t \mapsto \Lambda_k^x\left(t + T_{(k-1)}, \mathcal{F}_{T_{(k-1)}}\right) - \Lambda_k^x\left(T_{(k-1)}, \mathcal{F}_{T_{(k-2)}}\right)$$

If we suppose for simplicity that $\Lambda_k^x\left(t, \mathcal{F}_{T_{(k-1)}}\right)$ is invertible on $\left(T_{(k-1)}, \tau_{\text{end}}\right]$ with say inverse $\Lambda_k^{-1,x}\left(\cdot, \mathcal{F}_{T_{(k-1)}}\right)$ letting $E \sim \text{Exp}(1)$ be an exponential random variable with mean 1, we can simulate $S_{(k)}^x$ as follows

$$S_{(k)}^x = \Lambda_k^{-1,x}\left(E + \Lambda_k^x\left(T_{(k-1)}, \mathcal{F}_{T_{(k-2)}}\right), \mathcal{F}_{T_{(k-2)}}\right) - T_{(k-1)},$$

This can be seen by using the fact that if Λ is a cumulative hazard function for the random variable T, then $\Lambda^{-1}(E)$ is a random variable with the same distribution as T.

6. Intensities

It is illustrative to compare the simulation mechanism with a model for the intensities. Furthermore, we argue that observations from a counterfactual distribution can be simulated by setting $A\left(T_{(k)}\right) = 1$ for each $k = 1, ..., K$.

First, let us define the counting processes as

$$\begin{aligned}
N^y(t) &= \sum_{k=1}^{3} \mathbb{1}\left\{\Delta_{(k)} = y, T_{(k)} \leq t\right\}, \\
N^\ell(t) &= \sum_{k=1}^{2} \mathbb{1}\left\{\Delta_{(k)} = \ell, T_{(k)} \leq t\right\}, \\
N^{a1}(t) &= \sum_{k=1}^{2} \mathbb{1}\left\{\Delta_{(k)} = a, T_{(k)} \leq t, A\left(T_{(k)}\right) = 1\right\}, \\
N^{a0}(t) &= \sum_{k=1}^{2} \mathbb{1}\left\{\Delta_{(k)} = a, T_{(k)} \leq t, A\left(T_{(k)}\right) = 0\right\}.
\end{aligned}$$

Using Theorem II.7.1 of Andersen et al. (1993), we find that N^y has the P-$\mathcal{F}_t$ compensator

$$\Lambda^y(t) = \int_0^t \sum_{k=1}^{3} \mathbb{1}\left\{T_{(k-1)} < s \leq T_{(k)}\right\} \lambda_k^y \exp\left(\beta_{k,\text{ age}}^y \text{ age} + \beta_{k,A}^y A\left(T_{(k-1)}\right) + \beta_{k,L}^y L\left(T_{(k-1)}\right)\right) \mathrm{d}s,$$

with respect to the natural filtration $\mathcal{F}_t$. If $\lambda_1^y = \lambda_2^y = \lambda_3^y$, we can write the intensity as

$$
\begin{aligned}
\lambda^y(t) = \lambda_1^y & \\
& \exp\left(\beta^y_{k,\text{ age}} \text{ age}\right) \exp\Big(\beta^y_{1,A} A(t-) + \left(\beta^y_{2,A} - \beta^y_{1,A}\right) \mathbb{1}\left\{T_{(1)} < t \leq T_{(2)}\right\} A(t-) \\
& \quad + \left(\beta^y_{3,A} - \beta^y_{2,A}\right) \mathbb{1}\left\{T_{(2)} < t \leq T_{(3)}\right\} A(t-) + \beta^y_{1,\text{L}} L(t-) \\
& \quad + \left(\beta^y_{2,\text{L}} - \beta^y_{1,\text{L}}\right) \mathbb{1}\left\{T_{(1)} < t \leq T_{(2)}\right\} L(t-) \\
& \quad + \left(\beta^y_{3,\text{L}} - \beta^y_{2,\text{L}}\right) \mathbb{1}\left\{T_{(2)} < t \leq T_{(3)}\right\} L(t-)\Big) \mathbb{1}\left\{t \leq T_{(3)} \wedge \tau_{\text{end}}\right\}
\end{aligned}
$$

which shows that the model is compatible with the time-varying Cox model. We may find a similar expression for N^ℓ.

Let $\pi_t(a) = \sum_{k=1}^2 \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \text{ expit } \left(\alpha_{k0} + \alpha_{k,\text{ age}} \text{ age} + \alpha_{k,\text{L}} L\left(T_{(k-1)}\right)\right)$ and $\pi_t^*(a) = \sum_{k=1}^2 \mathbb{1}\left\{T_{(k-1)} < t \leq T_{(k)}\right\} \mathbb{1}\{a = 1\}$. Let λ^a be defined analogously to λ^y, then we find via Theorem II.7.1 of Andersen et al. (1993), that N^{a1} and N^{a0} have the compensators

$$
\begin{aligned}
\Lambda^{a1}(t) &= \int_0^t \pi_s(1) \lambda^a(s) \, \mathrm{d}s, \\
\Lambda^{a0}(t) &= \int_0^t \pi_s(0) \lambda^a(s) \, \mathrm{d}s.
\end{aligned}
$$

respectively. Simulating from the interventional mechanism corresponds to using the compensators for N^{ai},

$$
\begin{aligned}
\Lambda^{a1}(t) &= \int_0^t \pi_s^*(1) \lambda^a(s) \, \mathrm{d}s, \\
\Lambda^{a0}(t) &= \int_0^t \pi_s^*(0) \lambda^a(s) \, \mathrm{d}s.
\end{aligned}
$$

with the other compensators unchanged, which is the continuous time g-formula.

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

Laan, M. J. van der, & Gruber, S. (2012). *The International Journal of Biostatistics*, *8*(1). https://doi.org/doi:10.1515/1557-4679.1370

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, *27*(1), 23–31.

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

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. http://www.jstor.org/stable/2346178