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

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

where $T_{(k)}$ is the time of the k'th event, $\Delta_{(k)} \in \{\ell, a, y\}$ (stroke, visit, death), $A\big(T_{(k)}\big)$ is the treatment status at time $T_{(k)}$, and $L\big(T_{(k)}\big)$ 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 Exp(λ) 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

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

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

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

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:

$$
\begin{aligned}
S^{\ell}_{(2)} &\sim \text{Exp}\Big(\lambda^{\ell}_2 \exp\Big(\beta^{\ell}_{2,\text{age}}\ \text{age} + \beta^{\ell}_{2,A} A\Big(T_{(1)}\Big)\Big)\mathbb{1}\Big\{\Delta_{(1)} = a\Big\}\Big) \\
S^{y}_{(2)} &\sim \text{Exp}\Big(\lambda^{y}_2 \exp\Big(\beta^{y}_{2,\text{age}}\ \text{age} + \beta^{y}_{2,A} A\Big(T_{(1)}\Big) + \beta^{\ell}_{2,L} L\Big(T_{(1)}\Big)\Big)\Big) \\
S^{a}_{(2)} &\sim \text{Exp}\Big(\gamma_0 \exp(\gamma_{\text{age}}\ \text{age})\mathbb{1}\Big\{\Delta_{(1)} = \ell\Big\}\Big) \\
\Delta_{(2)} &= \text{argmin}_{x=a,\ell,y,} S^{x}_{(2)} \\
T_{(2)} &= T_{(1)} + S^{\Delta_{(2)}}_{(2)} \\
A\Big(T_{(2)}\Big) \mid T_{(2)} = t, \text{age} = x, A\Big(T_{(1)}\Big) = a_1, L\Big(T_{(1)}\Big) = l_1 & \\
&= \begin{cases} \sim \text{Bernoulli}(\text{expit}\big(\alpha_{20} + \alpha_{2,\ \text{age}} x + \alpha_{2,\text{L}} l_1\big)) \text{ if } \Delta_{(2)} = a \\ 1 \text{ otherwise} \end{cases} \\
L\Big(T_{(2)}\Big) &= 1.
\end{aligned}
$$

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:

$$
\begin{aligned}
S^{y}_{(3)} &\sim \text{Exp}\Big(\lambda^{y}_3 \exp\Big(\beta^{y}_{3,\text{age}}\ \text{age} + \beta^{y}_{3,A} A\Big(T_{(2)}\Big) + \beta^{y}_{3,L} L\Big(T_{(2)}\Big)\Big)\Big) \\
\Delta_{(3)} &= y \\
T_{(3)} &= T_{(2)} + S^{\Delta_{(3)}}_{(3)}.
\end{aligned}
$$

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.

2 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.

3 Table with example coefficients

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., $\lambda^{y}_1 = \lambda^{y}_2 = \lambda^{y}_3$ (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 $\alpha_{k,\text{L}}$ and the effect of treatment on stroke $\beta^{\ell}_{k,A}$.

We 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.
- $\alpha_{k,\text{L}}$: If negative: You will be less likely to be treated if you have had a stroke.
- $\beta^x_{k,\text{age}}$: 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 the table below. Each value is varied, holding the others fixed. The values with bold font correspond to the values used when fixed.

Parameters	Values
α_{k0}	0.3
$\alpha_{k,\ \text{age}}$	0.02
$\alpha_{k,\text{L}}$	**−0.1**, 0, 0.1
$\beta^{y}_{k,\ \text{age}}$	0.025
$\beta^{\ell}_{k,\ \text{age}}$	0.015
$\beta^{y}_{k,A}$	**−0.15**, 0, 0.15
$\beta^{\ell}_{k,A}$	**−0.2**, 0, 0.2
$\beta^{y}_{k,\text{L}}$	−0.25, 0, **0.25**
λ^{y}_{k}	0.0001
λ^{ℓ}_{k}	0.001
γ_{age}	0
γ_{0}	0.005

Strong confounding is considered in the following table in two different simulation settings.

Parameters	Values
α_{k0}	0.3
$\alpha_{k,\ \text{age}}$	0.02
$\alpha_{k,\text{L}}$	−0.3, 0.3
$\beta^{y}_{k,\ \text{age}}$	0.025
$\beta^{\ell}_{k,\ \text{age}}$	0.015
$\beta^{y}_{k,A}$	−0.4, 0.4
$\beta^{\ell}_{k,A}$	−0.2
$\beta^{y}_{k,\text{L}}$	0.5
λ^{y}_{k}	0.0001
λ^{ℓ}_{k}	0.001

Parameters	Values
γ_{age}	0
γ_0	0.005

Additionally, we consider the setting of no confounding effect on treatment and outcome, i.e., we set $\alpha_{k,\mathrm{L}} = 0$ and $\beta^y_{k,L} = 0$. In this case, the corresponding table is:

Parameters	Values
α_{k0}	0.3
$\alpha_{k,\ \text{age}}$	0.02
$\alpha_{k,\mathrm{L}}$	0
$\beta^y_{k,\ \text{age}}$	0.025
$\beta^\ell_{k,\ \text{age}}$	0.015
$\beta^y_{k,A}$	−0.15, 0, 0.15
$\beta^\ell_{k,A}$	−0.2
$\beta^y_{k,\mathrm{L}}$	0
λ^y_k	0.0001
λ^ℓ_k	0.001
γ_{age}	0
γ_0	0.005

4 Modeling

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

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

For $k = 3$, recall that

$$\mathcal{F}_{T_{(2)}} = \Big(\text{age}, A\Big(T_{(1)}\Big), L\Big(T_{(1)}\Big), T_{(1)}, \Delta_{(1)}, A\Big(T_{(2)}\Big), L\Big(T_{(2)}\Big), T_{(2)}, \Delta_{(2)}\Big).$$

We need to estimate

$$\bar{Q}^g_{2,\tau} = P\Big(T_{(3)} \leq \tau, \Delta_{(3)} = y \mid \mathcal{F}_{T_{(2)}}\Big)$$

Note that it is 0 if $T_{(2)} < \tau$ or $\Delta_{(2)} = y$, as in that case $T_{(3)} > \tau$, so we only need to estimate it for the individuals who are still at risk, i.e., those with $T_{(2)} < \tau$ and $\Delta_{(2)} \neq y$. This can be estimated by e.g., logistic regression. Importantly, in this first step, we do not impose any intervention, as you cannot visit the doctor for $k = 3$. Denote this estimator by $\hat{\nu}_2$.

For $k = 2$, we should model the conditional counterfactual probability $\bar{Q}^g_{1,\tau}$ of having the primary event within $\tau = 720$ days given the history $\mathcal{F}_{T_{(1)}}$, among the people who are still at risk $\Big(T_{(1)} < \tau \wedge \Delta_{(1)} \neq y\Big)$ using the model for $k = 3$, see e.g., the main note for elaboration on why $\bar{Q}^g_{1,\tau}$ has this interpretation. Here

$$\mathcal{F}_{T_{(1)}} = \Big(\text{age}, A\Big(T_{(1)}\Big), L\Big(T_{(1)}\Big), T_{(1)}, \Delta_{(1)}\Big).$$

As described in the section "Algorithm for ICE-IPCW Estimator" (set $\hat{S}^c = 1$ in the algorithm), this is done by calculating $\hat{Z}^a_1$ (outcome) for each individual at risk using $\hat{\nu}_2$, and regressing on $\mathcal{F}_{T_{(1)}}$ (covariates). We apply a generalized linear model (GLM) with the option `family = quasibinomial`. The resulting estimator is denoted $\hat{\nu}_1$ which can provide predictions for the conditional counterfactual probability of having a primary event within $\tau = 720$ days given the information that you have after one event.

For $k = 1$, we need to estimate the conditional probability of having a stroke within $\tau = 720$ days given the history $\mathcal{F}_0$ for all individuals.

For the treatment propensity, we can simply estimate this using logistic regression. For instance $\pi_k\Big(t, \mathcal{F}_{T_{(k-1)}}\Big)$ can be estimated by regressing $A\Big(T_{(k)}\Big)$ on $\mathcal{F}_{T_{(k-1)}}$ and $T_{(k)}$ among people with $\Delta_{(k-1)} = a, \ell$ and $\Delta_{(k)} = a$.

5 Extensions

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). Then, we might consider

$$S^y_{(3)} \sim \text{Exp}\Big(\lambda^y_3 \exp\Big(\beta^y_{3,\text{age}}\ \text{age} + \beta^y_{3,A} A\Big(T_{(2)}\Big) + \beta^y_{3,L} L\Big(T_{(2)}\Big) + \beta^y_{3,T^\ell} T^\ell\Big)\Big),$$

or we might consider a model in which the baseline hazard is not constant. It also might be easier to state a realistic model in terms of the intensities directly, in which case, we can then "transform" to the interevent scale. For example, a realistic intensity for the primary event,

$$\begin{aligned}\lambda^y(t) = \lambda^y_0(t) \exp\big(\beta^y_{\text{age}}\ \text{age}\big) \exp\Big(&\beta^y_L \exp\big(\beta^{y*}_L (t - T^\ell)\big) L(t-) + \beta^y_A \exp(\beta^{y*}_A (t - T^a))(1 - A(t-)) \\ &+ \beta^y_Z \mathbb{1}\Big\{T_{(2)} < t\Big\} \mathbb{1}\Big\{\Delta_{(1)} = a\Big\} \mathbb{1}\Big\{\Delta_{(2)} = \ell\Big\}\Big) \mathbb{1}\{t \leq T^y\}\end{aligned}$$

Here T^a denotes the time to the last treatment. Note that each term is zero if the corresponding event has not happened yet, so we do not condition on the future. Here, we can let each coefficient depend on event number, but for simplicity of notation, we do not do so. The last term corresponds to there being an effect of the order in which the events happened after two events. This is one way to include longe range dependencies. Simulating from this model is significantly more complicated, because we have to rely on numeric integration.

More generally, let $\Lambda^x_k\Big(t, \mathcal{F}_{T_{(k-1)}}\Big)$ denote the cumulative hazard function for the k'th event of type x at time t conditional on the history $\mathcal{F}_{T_{(k-1)}}$. The cumulative hazard-cause specific hazard of $S_{(k)}$ of the k'th event type x is given by

$$\Big[\tau_{\text{end}} - T_{(k-1)}, 0\Big) \ni t \mapsto \Lambda^x_k\Big(t + T_{(k-1)}, \mathcal{F}_{T_{(k-1)}}\Big) - \Lambda^x_k\Big(T_{(k-1)}, \mathcal{F}_{T_{(k-2)}}\Big)$$

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

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

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 $A\Big(T_{(k)}\Big) = 1$ for each $k = 1, ..., K$.

First, let us define the counting processes as

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

Using Theorem II.7.1 of Andersen et al. (1993), we find that N^y has the compensator

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

If $\lambda_1^y = \lambda_2^y = \lambda_3^y$, we can write the intensity as

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

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

Let $\pi_t(a) = \sum_{k=1}^{2} \mathbb{1}\Big\{T_{(k-1)} < t \leq T_{(k)}\Big\} \text{ expit } \Big(\alpha_{k0} + \alpha_{k,\text{ age}} \text{ age} + \alpha_{k,\text{L}} L\Big(T_{(k-1)}\Big)\Big)$ and $\pi_t^*(a) = \sum_{k=1}^{2} \mathbb{1}\Big\{T_{(k-1)} < t \leq T_{(k)}\Big\}\mathbb{1}\{a = 1\}$. 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

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

respectively. Simulating from the interventional mechanism corresponds to having the compensators

$$\Lambda^{a1}(t) = \int_0^t \pi^*_s(1)\lambda^a(s)\,\mathrm{d}s,$$
$$\Lambda^{a0}(t) = \int_0^t \pi^*_s(0)\lambda^a(s)\,\mathrm{d}s.$$

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