Supplementary Materials for:
"Causal inference for semi-competing risks data"

Daniel Nevo and Malka Gorfine, Tel Aviv University

Description of contents in the Supplementary Materials

- Section A presents proofs and identification results when covariates are present.
- Section B describes a sensitivity analysis approach without using a frailty assumption.
- Section C provides additional information about the derivation of the nonparametric estimators (Section C.1), the asymptotic properties of these estimators (Section C.2) and details of the EM algorithm for fitting the semi-parametric model (Section C.3).
- Section D includes details on the simulation studies, including the data-generating mechanism (DGM, Section D.1), the different analyses (Section D.2), as well as simulation results (Section D.3).
- Section E gives more detailed information on the illustrative data analysis and presents additional results.

A Proofs and additional theory on identifiability

A.1 Proof of Proposition 1

Each of the causal contrasts in (3.8)–(3.10) have one component identified by our assumptions and one that is not. We start with the former:

$$\begin{aligned}
\Pr[T_2(0) \le t|ad] &= \Pr[T_2(0) \le t|T_1(0) \le T_2(0)] = \Pr[T_2(0) \le t|A=0, T_1 \le T_2] \\
&= F_{2|A=0,T_1 \le T_2}(t), \\
\Pr[T_2(1) \le t|nd] &= \Pr[T_2(1) \le t|T_1(1) > T_2(1)] = \Pr[T_2(1) \le t|A=1, T_1 > T_2] \\
&= F_{2|A=1,T_1 > T_2}(t), \\
\Pr[T_1(0) \le t|ad] &= \Pr[T_1(0) \le t|T_1(0) \le T_2(0)] = \Pr[T_1(0) \le t|A=0, T_1 \le T_2] \\
&= F_{1|A=0,T_1 \le T_2}(t).
\end{aligned}$$

In each of the equations, the first equality is by Assumption 3, the second by Assumption 2 and the third by Assumption 1. Turning to the partially identified components, lower bounds are obtained by:

$$\begin{aligned}
\Pr[T_2(1) \le t|ad] &= \Pr[T_2(1) \le t|T_1(0) \le T_2(0)] \\
&\ge \frac{\Pr[T_2(1) \le t] + \Pr[T_1(0) \le T_2(0)] - 1}{\Pr[T_1(0) \le T_2(0)]} \\
&= 1 - \frac{S_{2|A=1}(t)}{\eta_{A=0}}, \\
\Pr[T_2(0) \le t|nd] &= \Pr[T_2(0) \le t|T_1(1) > T_2(1)] \\
&\ge \frac{\Pr[T_2(0) \le t] + \Pr[T_1(1) > T_2(1)] - 1}{\Pr[T_1(1) > T_2(1)]} \\
&= 1 - \frac{S_{2|A=0}(t)}{1 - \eta_{A=1}}, \\
\Pr[T_1(1) \le t|ad] &= \Pr[T_1(1) \le t|T_1(0) \le T_2(0)] \\
&\ge \frac{\Pr[T_1(1) \le t] + \Pr[T_1(0) \le T_2(0)] - 1}{\Pr[T_1(0) \le T_2(0)]} \\
&= 1 - \frac{S_{1|A=1}(t)}{\eta_{A=0}}.
\end{aligned}$$

In each of the equations, the first equality is by Assumption 3, the second line is by the inequality $\Pr(B_1 \cap B_2) \geq \Pr(B_1) + \Pr(B_2) - 1$ for any two events B_1, B_2, and the third line is by Assumptions 1 and 2.

The upper bounds for the components are obtained by:

$$\begin{aligned}
\Pr[T_2(1) \leq t|ad] &= \Pr[T_2(1) \leq t|T_1(0) \leq T_2(0), T_1(1) \leq T_2(1)] \\
&= \Pr[T_1(0) \leq T_2(0), T_1(1) \leq T_2(1)|T_2(1) \leq t] \frac{F_{2|A=1}(t)}{\eta_{A=0}} \\
&\leq \Pr[T_1(1) \leq T_2(1)|T_2(1) \leq t] \frac{F_{2|A=1}(t)}{\eta_{A=0}} \\
&= \eta_{A=1,T_2 \leq t} \frac{F_{2|A=1}(t)}{\eta_{A=0}}, \\
\Pr[T_2(0) \leq t|nd] &= \Pr[T_2(0) \leq t|T_1(0) > T_2(0), T_1(1) > T_2(1)] \\
&= \Pr[T_1(0) > T_2(0), T_1(1) > T_2(1)|T_2(0) \leq t] \frac{F_{2|A=0}(t)}{1 - \eta_{A=1}} \\
&\leq \Pr[T_1(0) > T_2(0)|T_2(0) \leq t] \frac{F_{2|A=0}(t)}{1 - \eta_{A=1}} \\
&= (1 - \eta_{A=0,T_2 \leq t}) \frac{F_{2|A=0}(t)}{1 - \eta_{A=1}}, \\
\Pr[T_1(1) \leq t|ad] &= \Pr[T_1(1) \leq t|T_1(0) \leq T_2(0), T_1(1) \leq T_2(1)] \\
&= \Pr[T_1(0) \leq T_2(0), T_1(1) \leq T_2(1)|T_1(1) \leq t] \frac{F_{1|A=1}(t)}{\eta_{A=0}} \\
&\leq \Pr[T_1(1) \leq T_2(1)|T_1(1) \leq t] \frac{F_{1|A=1}(t)}{\eta_{A=0}} \\
&= \eta_{A=1,T_1 \leq t} \frac{F_{1|A=1}(t)}{\eta_{A=0}}.
\end{aligned}$$

In each of the equations, the second line is by Bayes' Theorem and Assumptions 1–3, the inequality is by increasing the event probability by omission, and the final equality is again by Assumptions 1 and 2.

A.2 Proof of Proposition 2

First, using the same arguments as in the proof of Proposition 3

$$\begin{aligned}\Pr[T_1(0) \leq t|ad] &= \Pr[T_1(0) \leq t|T_1(0) \leq T_2(0)] = \Pr[T_1(0) \leq t|A = 0, T_1 \leq T_2] \\ &= F_{1|A=0,T_1 \leq T_2}(t).\end{aligned}$$

Now,

$$\begin{aligned}&\Pr[T_1(1) \leq t|ad] \\ &\quad = \Pr[T_1(1) \leq t|ad]\Pr[ad|T_1(1) \leq T_2(1)] + \Pr[T_1(1) \leq t|ad]\Pr[dh|T_1(1) \leq T_2(1)] \\ &\quad \geq \Pr[T_1(1) \leq t|ad]\Pr[ad|T_1(1) \leq T_2(1)] + \Pr[T_1(1) \leq t|dh]\Pr[dh|T_1(1) \leq T_2(1)] \\ &\quad = \Pr[T_1(1) \leq t|T_1(1) \leq T_2(1)] \\ &\quad = \Pr[T_1(1) \leq t|A = 1, T_1 \leq T_2] \\ &\quad = F_{1|A=1,T_1 \leq T_2}(t),\end{aligned}$$

where the first equality is justified by $\Pr[ad|T_1(1) \leq T_2(1)] + \Pr[dh|T_1(1) \leq T_2(1)] = 1$, the inequality is by Assumption 4, and the rest is by the law of total probability and Assumptions 1 and 2.

A.3 Adjusted bounds

We prove below that under Assumptions 1, 3 and 5, adjusted bounds for (2.1) in the main text are given by (3.12). First, observe that under the above assumptions, $\Pr(Z = z|ad)$ is identifiable from the data by

$$\Pr(Z = z|ad) = \Pr[Z = z|T_1(0) < T_2(0)] = \Pr[Z = z|A = 0, T_1 \leq T_2] = \nu(z), \quad \text{(A.1)}$$

where the first equality is by Assumption 3, and the second equality is by Assumptions 1 and 5, and the third is again by Assumption 5. The second term in (1), $Pr[T_2(0) \leq$

$t \mid ad]$, is identified from the data under our assumptions by

$$\begin{aligned} Pr[T_2(0) \leq t \mid ad] &= \sum_z \Pr(Z = z|ad) \Pr[T_2(0) \leq t \mid ad, Z = z] \\ &= \sum_z \nu(z) \Pr[T_2(0) \leq t \mid T_1(0) < T_2(0), Z = z] \\ &= \sum_z \nu(z) F_{2|A=0,T_1 \leq T_2, Z=z}(t), \end{aligned}$$

where the second line is by (A.1) and Assumption 3, and the third line is by Assumptions 1 and 5. Turning to $\Pr[T_2(1) \leq t|ad]$, the upper bound is obtained by

$$\begin{aligned} &\Pr[T_2(1) \leq t|ad] = \sum_z \Pr(Z = z|ad) \Pr[T_2(1) \leq t \mid ad, Z = z] \\ &= \sum_z \nu(z) \Pr[T_2(1) \leq t | T_1(0) \leq T_2(0), T_1(1) \leq T_2(1), Z = z] \\ &= \sum_z \nu(z) \Pr[T_1(0) \leq T_2(0), T_1(1) \leq T_2(1) | T_2(1) \leq t, Z = z] \frac{F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}} \\ &\leq \sum_z \nu(z) \min\left\{1,\ \Pr[T_1(1) \leq T_2(1)|T_2(1) \leq t, Z = z] \frac{F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\} \\ &= \sum_z \nu(z) \min\left\{1,\ \eta_{A=1,T_2 \leq t, Z=z} \frac{F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\}. \end{aligned}$$

The second line is by (A.1); the third line is by Bayes' Theorem and Assumptions 1, 3 and 5; in the fourth line the probability was increased by omitting $T_1(0) \leq T_2(0)$; the last line is again by Assumptions 1 and 5. The lower bound is obtained by

$$\begin{aligned} &\Pr[T_2(1) \leq t|ad] = \sum_z \Pr(Z = z|ad) \Pr[T_2(1) \leq t \mid ad, Z = z] \\ &= \sum_z \nu(z) \Pr[T_2(1) \leq t \mid T_1(0) < T_2(0), Z = z] \\ &\geq \sum_z \nu(z) \max\left\{0\ ,\ \frac{\Pr[T_2(1) \leq t|Z = z] + \Pr[T_1(0) \leq T_2(0)|Z = z] - 1}{\Pr[T_1(0) \leq T_2(0)|Z = z]}\right\} \\ &= \sum_z \nu(z) \max\left\{0\ , 1 - \frac{1 - F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\}, \end{aligned}$$

where the second line is by (A.1) and Assumption 3, the third line is by the inequality $\Pr(B_1 \cap B_2) \geq \Pr(B_1) + \Pr(B_2) - 1$ for any two events B_1, B_2, and the fourth line is by Assumptions 1 and 5.

A.4 Proof that adjusted bounds are contained within the unadjusted bounds

The proof is similar to the proof of Proposition 1 of Long and Hudgens (2013). First, observe that

$$\sum_z \nu(z) F_{2|A=0,T_1 \leq T_2, Z=z}(t) = F_{2|A=0,T_1 \leq T_2}(t).$$

Thus, we can focus on the first component of $U^Z_{2,ad}(t)$ and $\mathcal{L}^Z_{2,ad}(t)$. Starting from the upper bound, we have

$$
\begin{aligned}
&\sum_z \nu(z) \min\left\{1, \eta_{A=1,T_2\le t,Z=z}\frac{F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\} \\
&\le \min\left\{1\ , \sum_z \Pr(Z=z|T_1\le T_2)\eta_{A=1,T_2\le t,Z=z}\frac{F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\} \\
&= \min\left\{1\ , \sum_z \Pr(Z=z|A=0,T_1\le T_2)\eta_{A=1,T_2\le t,Z=z}\frac{F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\} \\
&= \min\left\{1\ , \sum_z \frac{\Pr(Z=z|A=0)}{\eta_{A=0}}\eta_{A=1,T_2\le t,Z=z}F_{2|A=1,Z=z}(t)\right\} \\
&= \min\left\{1\ , \frac{1}{\eta_{A=0}}\sum_z \Pr(Z=z)\eta_{A=1,T_2\le t,Z=z}F_{2|A=1,Z=z}(t)\right\} \\
&= \min\left\{1\ , \frac{1}{\eta_{A=0}}\sum_z \Pr(Z=z)\frac{\Pr(A=1,T_1\le T_2,T_2\le t,Z=z)}{\Pr(A=1,T_2\le t,Z=z)}\frac{\Pr(A=1,T_2\le t,Z=z)}{\Pr(A=1)\Pr(Z=z)}\right\} \\
&= \min\left\{1\ , \frac{1}{\eta_{A=0}}\sum_z \frac{\Pr(A=1,T_1\le T_2,T_2\le t,Z=z)}{\Pr(A=1)}\right\} \\
&= \min\left\{1\ , \frac{1}{\eta_{A=0}}\frac{\Pr(A=1,T_1\le T_2,T_2\le t)}{\Pr(A=1)}\right\} \\
&= \min\left\{1\ , \frac{1}{\eta_{A=0}}\Pr(T_1\le T_2|A=1,T_2\le t)\Pr(T_2\le t|A=1)\right\} \\
&= \min\left\{1\ , F_{2|A=1}(t)\frac{\eta_{A=1,T_2\le t}}{\eta_{A=0}}\right\}.
\end{aligned}
$$

The second line is obtained by recalling that the sum of minimizers is lower or equal to the minimizer of the sum. In the third line, we could add $A=0$ to the condition because of Assumption 5. The fourth line is by Bayes' theorem, and in the fifth line we used Assumption 5 again, this time to remove $A=0$ from the condition. In the sixth line, we just wrote conditional probabilities as joint probabilities divided by the conditional probabilities, and utilized Assumption 5 to write $\Pr(A=1,Z=z)=\Pr(A=1)\Pr(Z=z)$. The eighth line is by the law of total probability, and in the ninth line we wrote a joint probability as the product of conditional probabilities and the marginal (which was canceled with the denominator). Turning to the lower bound,

$$
\begin{aligned}
&\sum_z \nu(z) \max\left\{0 \ , 1 - \frac{1 - F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\} \\
&\geq \max\left\{1 \ , 1 - \sum_z \Pr(Z = z|T_1 \leq T_2)\frac{1 - F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\} \\
&= \max\left\{1 \ , 1 - \sum_z \Pr(Z = z|A = 0, T_1 \leq T_2)\frac{1 - F_{2|A=1,Z=z}(t)}{\eta_{A=0,Z=z}}\right\} \\
&= \max\left\{1 \ , 1 - \sum_z \frac{\Pr(Z = z|A = 0)}{\eta_{A=0}} \Pr(T_2 > t|A = 1, Z = z)\right\} \\
&= \max\left\{1 \ , 1 - \frac{1}{\eta_{A=0}} \sum_z \Pr(Z = z) \Pr(T_2 > t|A = 1, Z = z)\right\} \\
&= \max\left\{1 \ , 1 - \frac{S_{2|A=1}(t)}{\eta_{A=0}}\right\}.
\end{aligned}
$$

The second line is obtained by recalling that the sum of maximizers is larger than the maximum of sums. In the third line, we could add $A = 0$ to the condition because of Assumption 5. The fourth line is by Bayes' theorem, and in the fifth line we used Assumption 5 again, this time to remove $A = 0$ from the condition. The last line is obtained by the law of total probability.

A.5 Proof of Proposition 4

For $j = 1, 2$ and $a = 0, 1$ we may write

$$
\begin{aligned}
\Pr[T_j(a) \leq t|ad] &= E_{\boldsymbol{\gamma}|ad}\{\Pr[T_j(a) \leq t \mid T_1(0) \leq T_2(0), T_1(1) \leq T_2(1), \boldsymbol{\gamma}]\} \\
&= E_{\boldsymbol{\gamma}|ad}\{\Pr[T_j(a) \leq t|T_1(a) \leq T_2(a), \gamma_a]\} \\
&= E_{\boldsymbol{\gamma}|ad}\{\Pr[T_j \leq t|T_1 \leq T_2, A = a, \gamma_a]\} \\
&= \int_0^\infty \int_0^\infty \Pr[T_j \leq t|T_1 \leq T_2, A = a, \gamma_a] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}|ad) d\boldsymbol{\gamma},
\end{aligned}
$$

where the second line is by Part (i) of Assumption 6 and the third line is by Part (ii) of Assumption 6. Now, by Bayes' Theorem

$$\begin{aligned} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}|ad) &= \frac{\Pr[T_1(0) \leq T_2(0), T_1(1) \leq T_2(1)|\boldsymbol{\gamma}] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma})}{\int_0^\infty \int_0^\infty Pr[T_1(0) \leq T_2(0), T_1(1) \leq T_2(1)|\boldsymbol{\gamma}'] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') d\boldsymbol{\gamma}'} \\ &= \frac{\Pr[T_1(0) \leq T_2(0)|\gamma_0] \Pr[T_1(1) \leq T_2(1)|\gamma_1] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma})}{\int_0^\infty \int_0^\infty \Pr[T_1(0) \leq T_2(0)|\gamma_0'] \Pr[T_1(1) \leq T_2(1)|\gamma_1'] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') d\boldsymbol{\gamma}'} \\ &= \frac{\Pr[T_1 \leq T_2|A=0, \gamma_0] \Pr[T_1 \leq T_2|A=1, \gamma_1] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma})}{\int_0^\infty \int_0^\infty \Pr[T_1 \leq T_2|A=0, \gamma_0'] \Pr[T_1 \leq T_2|A=1, \gamma_1'] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') d\boldsymbol{\gamma}'}, \\ &= \frac{\eta_{A=0,\gamma_0} \eta_{A=1,\gamma_1} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) d\boldsymbol{\gamma}}{\int_0^\infty \int_0^\infty \eta_{A=0,\gamma_0'} \eta_{A=1,\gamma_1'} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') d\boldsymbol{\gamma}'}, \end{aligned}$$

where the second line is by Part (i) of Assumption 6, the third line is by Part (ii) of Assumption 6 and by Assumption 1. Next, by Part (iv) of Assumption 6, $f_{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ is identifiable from the data. Therefore, estimation of $\Pr[T_j \leq t|T_1 \leq T_2, A=a, \gamma_a]$ for $j = 1, 2$ and $a = 0, 1$ as well as estimation of $\eta_{A=0,\gamma_0}\eta_{A=1,\gamma_1}$ and $\boldsymbol{\theta}$ is possible using, for example, a covariate-free version of the EM algorithm we proposed in Section C.3. The proof for identification of $\Pr[T_2(a) \leq t|nd]$ under Assumptions 1 and 6 is similar.

A.6 The analogue of Propositions 1 and 4 under conditional version of the assumptions

First, for simplicity of presentation assume $\boldsymbol{X}$ is continuous and let $f_{\boldsymbol{X}}(\boldsymbol{x})$ denote its joint density function and let $\mathcal{X}$ denote its support. If some components of $\boldsymbol{X}$ are discrete, the generalization of the results below are straightforward. For any event $\mathcal{Q}$ (including the empty set $\emptyset$), let also $F^{\boldsymbol{X}}_{j|\mathcal{Q}}(t) = \int_{\boldsymbol{x}\in\mathcal{X}} \Pr(T_j \leq t|\mathcal{Q}, \boldsymbol{X} = \boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}$, $S^{\boldsymbol{X}}_{j|\mathcal{Q}}(t) = 1 - F^{\boldsymbol{X}}_{j|\mathcal{Q}}(t)$ and $\eta^{\boldsymbol{X}}_{\mathcal{Q}} = \int_{\boldsymbol{x}\in\mathcal{X}} \Pr(T_1 \leq T_2|\mathcal{Q}, \boldsymbol{X} = \boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}$. We replace Assumption 2 with the following weaker assumption,

Assumption A.1. *Weak Ignorability* $A \perp\!\!\!\perp \{T_1(a), T_2(a)\}|\boldsymbol{X}$ *for* $a = 0, 1$.

A.6.1 Analogue of Proposition 1

Proposition A.1. *Under Assumptions 1, 3 and A.1, the causal effects of interest are partially identified by*

$$\mathcal{L}^{\boldsymbol{X}}_{2,ad}(t) \leq \Pr[T_2(1) \leq t|ad] - \Pr[T_2(0) \leq t|ad] \leq \mathcal{U}^{\boldsymbol{X}}_{2,ad}(t)$$
$$\mathcal{L}^{\boldsymbol{X}}_{2,nd}(t) \leq \Pr[T_2(1) \leq t|nd] - \Pr[T_2(0) \leq t|nd] \leq \mathcal{U}^{\boldsymbol{X}}_{2,nd}(t)$$
$$\mathcal{L}^{\boldsymbol{X}}_{1,ad}(t) \leq \Pr[T_1(1) \leq t|ad] - \Pr[T_1(0) \leq t|ad] \leq \mathcal{U}^{\boldsymbol{X}}_{1,ad}(t)$$

where

$$\mathcal{L}^{\boldsymbol{X}}_{2,ad}(t) = \max\left\{0\ , 1 - \frac{S^{\boldsymbol{X}}_{2|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}}\right\} - F^{\boldsymbol{X}}_{2|A=0,T_1\leq T_2}(t)$$
$$\mathcal{U}^{\boldsymbol{X}}_{2,ad}(t) = \min\left\{1\ , F^{\boldsymbol{X}}_{2|A=1}(t)\frac{\eta^{\boldsymbol{X}}_{A=1,T_2\leq t}}{\eta^{\boldsymbol{X}}_{A=0}}\right\} - F^{\boldsymbol{X}}_{2|A=0,T_1\leq T_2}(t)$$
$$\mathcal{L}^{\boldsymbol{X}}_{2,nd}(t) = F^{\boldsymbol{X}}_{2|A=1,T_1>T_2}(t) - \min\left\{1\ , F^{\boldsymbol{X}}_{2|A=0}(t)\frac{1-\eta^{\boldsymbol{X}}_{A=0,T_2\leq t}}{1-\eta^{\boldsymbol{X}}_{A=1}}\right\}$$
$$\mathcal{U}^{\boldsymbol{X}}_{2,nd}(t) = F^{\boldsymbol{X}}_{2|A=1,T_1>T_2}(t) - \max\left\{0\ , 1 - \frac{F^{\boldsymbol{X}}_{2|A=0}(t)}{1-\eta^{\boldsymbol{X}}_{A=1}}\right\}$$
$$\mathcal{L}^{\boldsymbol{X}}_{1,ad}(t) = \max\left\{0\ , 1 - \frac{S^{\boldsymbol{X}}_{1|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}}\right\} - F^{\boldsymbol{X}}_{1|A=0,T_1\leq T_2}(t)$$
$$\mathcal{U}^{\boldsymbol{X}}_{1,ad}(t) = \min\left\{1\ , \frac{F^{\boldsymbol{X}}_{1|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}}\right\} - F^{\boldsymbol{X}}_{1|A=0,T_1\leq T_2}(t).$$

Each of the causal contrasts in (2.1)–(2.3) of the main text have one component identified by our assumptions and one that is not. We start with the former:

$$
\begin{aligned}
\Pr[T_2(0) \leq t|ad] &= \int_{\boldsymbol{x}\in\mathcal{X}} \Pr[T_2(0) \leq t|T_1(0) \leq T_2(0), \boldsymbol{X} = \boldsymbol{x}] f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}\in\mathcal{X}} \Pr[T_2(0) \leq t|A = 0, T_1 \leq T_2, \boldsymbol{X} = \boldsymbol{x}] f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}\in\mathcal{X}} F_{2|A=0,T_1(0)\leq T_2(0),\boldsymbol{X}=\boldsymbol{x}} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= F^{\boldsymbol{X}}_{2|A=0,T_1\leq T_2}(t), \\
\Pr[T_2(1) \leq t|nd] &= \Pr[T_2(1) \leq t|T_1(1) > T_2(1), \boldsymbol{X} = \boldsymbol{x}] f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}\in\mathcal{X}} \Pr[T_2(1) \leq t|A = 1, T_1(1) > T_2(1), \boldsymbol{X} = \boldsymbol{x}] f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}\in\mathcal{X}} F_{2|A=1,T_1>T_2,\boldsymbol{X}=\boldsymbol{x}} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= F^{\boldsymbol{X}}_{2|A=1,T_1>T_2}(t), \\
\Pr[T_1(0) \leq t|ad] &= \Pr[T_1(0) \leq t|T_1(0) \leq T_2(0), \boldsymbol{X} = \boldsymbol{x}] f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}\in\mathcal{X}} \Pr[T_1(0) \leq t|A = 0, T_1(0) \leq T_2(0), \boldsymbol{X} = \boldsymbol{x}] f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}\in\mathcal{X}} F_{1|A=0,T_1\leq T_2,\boldsymbol{X}=\boldsymbol{x}} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \\
&= F^{\boldsymbol{X}}_{1|A=0,T_1\leq T_2}(t).
\end{aligned}
$$

In each of the equations, the first equality is by Assumption 3 and law of total probability, the second by Assumption A.1 and the third by Assumption 1. Turning to the partially identified components, lower bounds are obtained by:

$$\begin{aligned}
\Pr[T_2(1) \leq t|ad] &= \Pr[T_2(1) \leq t|T_1(0) \leq T_2(0)] \\
&\geq \frac{\Pr[T_2(1) \leq t] + \Pr[T_1(0) \leq T_2(0)] - 1}{\Pr[T_1(0) \leq T_2(0)]} \\
&= 1 - \frac{\int_{\boldsymbol{x}\in\mathcal{X}} S_{2|A=1,\boldsymbol{X}=\boldsymbol{x}}(t) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}{\int_{\boldsymbol{x}\in\mathcal{X}} \eta_{A=0,\boldsymbol{X}=\boldsymbol{x}} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}} \\
&= 1 - \frac{S^{\boldsymbol{X}}_{2|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}}, \\
\Pr[T_2(0) \leq t|nd] &= \Pr[T_2(0) \leq t|T_1(1) > T_2(1)] \\
&\geq \frac{\Pr[T_2(0) \leq t] + \Pr[T_1(1) > T_2(1)] - 1}{\Pr[T_1(1) > T_2(1)]} \\
&= 1 - \frac{\int_{\boldsymbol{x}\in\mathcal{X}} S_{2|A=0,\boldsymbol{X}=\boldsymbol{x}}(t) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}{1 - \int_{\boldsymbol{x}\in\mathcal{X}} \eta_{A=1,\boldsymbol{X}=\boldsymbol{x}} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}} \\
&= 1 - \frac{S^{\boldsymbol{X}}_{2|A=0(t)}}{1 - \eta^{\boldsymbol{X}}_{A=1}}, \\
\Pr[T_1(1) \leq t|ad] &= \Pr[T_1(1) \leq t|T_1(0) \leq T_2(0)] \\
&\geq \frac{\Pr[T_1(1) \leq t] + \Pr[T_1(0) \leq T_2(0)] - 1}{\Pr[T_1(0) \leq T_2(0)]} \\
&= 1 - \frac{\int_{\boldsymbol{x}\in\mathcal{X}} S_{1|A=1,\boldsymbol{X}=\boldsymbol{x}}(t) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}}{\int_{\boldsymbol{x}\in\mathcal{X}} \eta_{A=0,\boldsymbol{X}=\boldsymbol{x}} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}} \\
&= 1 - \frac{S^{\boldsymbol{X}}_{1|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}}.
\end{aligned}$$

In each of the equations, the first equality is by Assumption 3, the second line is by the inequality $\Pr(B_1 \cap B_2) \geq \Pr(B_1) + \Pr(B_2) - 1$ for any two events B_1, B_2, and the third line is by Assumptions 1 and A.1.

The upper bounds for the components are obtained by:

$$\begin{aligned}
\Pr[T_2(1) \le t|ad] &= \Pr[T_2(1) \le t|T_1(0) \le T_2(0), T_1(1) \le T_2(1)] \\
&= \Pr[T_1(0) \le T_2(0), T_1(1) \le T_2(1)|T_2(1) \le t]\frac{F^{\boldsymbol{X}}_{2|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}} \\
&\le \Pr[T_1(1) \le T_2(1)|T_2(1) \le t]\frac{F^{\boldsymbol{X}}_{2|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}} \\
&= \eta^{\boldsymbol{X}}_{A=1,T_2\le t}\frac{F^{\boldsymbol{X}}_{2|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}}, \\
\Pr[T_2(0) \le t|nd] &= \Pr[T_2(0) \le t|T_1(0) > T_2(0), T_1(1) > T_2(1)] \\
&= \Pr[T_1(0) > T_2(0), T_1(1) > T_2(1)|T_2(0) \le t]\frac{F^{\boldsymbol{X}}_{2|A=0}(t)}{1-\eta^{\boldsymbol{X}}_{A=1}} \\
&\le \Pr[T_1(0) > T_2(0)|T_2(0) \le t]\frac{F^{\boldsymbol{X}}_{2|A=0}(t)}{1-\eta^{\boldsymbol{X}}_{A=1}} \\
&= (1-\eta^{\boldsymbol{X}}_{A=0,T_2\le t})\frac{F^{\boldsymbol{X}}_{2|A=0}(t)}{1-\eta^{\boldsymbol{X}}_{A=1}}, \\
\Pr[T_1(1) \le t|ad] &= \Pr[T_1(1) \le t|T_1(0) \le T_2(0), T_1(1) \le T_2(1)] \\
&= \Pr[T_1(0) \le T_2(0), T_1(1) \le T_2(1)|T_1(1) \le t]\frac{F^{\boldsymbol{X}}_{1|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}} \\
&\le \Pr[T_1(1) \le T_2(1)|T_2(1) \le t]\frac{F^{\boldsymbol{X}}_{1|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}} \\
&= \eta^{\boldsymbol{X}}_{A=1,T_1\le t}\frac{F^{\boldsymbol{X}}_{1|A=1}(t)}{\eta^{\boldsymbol{X}}_{A=0}}.
\end{aligned}$$

In each of the equations, the second line is by Bayes' Theorem and Assumptions 1,3 and A.1, the inequality is by increasing the event probability by omission, and the final equality is again by Assumptions 1 and A.1.

A.6.2 Analogue of Proposition 4

We first adapt Assumption 6 for the case of covariates $\boldsymbol{X}$ (this is a weaker version of the assumption).

Assumption A.2. *There exists a bivariate random variable* $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)$ *such that*

(i) $A \perp\!\!\!\perp \{T_1(a), T_2(a), \gamma_a\} | \boldsymbol{X}$ *for* $a = 0, 1$.

(ii) Given $\boldsymbol{\gamma}$*, the joint distribution of the potential event times can be factored as follows*

$$f[T_1(0), T_2(0), T_1(1), T_2(1)|\boldsymbol{X}, \boldsymbol{\gamma}] = f[T_1(0), T_2(0)|\boldsymbol{X}, \gamma_0]P[T_1(1), T_2(1)|\boldsymbol{X}, \gamma_1],$$

where $f(\cdot)$ *denotes a density function of a possibly-multivariate random variable.*

(iii) The frailty variables and the covariates are independent $\boldsymbol{\gamma} \perp\!\!\!\perp \boldsymbol{X}$.

(iv) The frailty variable operates multiplicatively on the hazard functions. That is, $\lambda_{jk}(t|a, \boldsymbol{X}, \boldsymbol{\gamma}) = \gamma_a \widetilde{\lambda}_{jk}(t|a, \boldsymbol{X})$ *for* $jk = 01, 02, a = 0, 1$ *and* $\lambda_{12}(t|t_1, a, \boldsymbol{X}, \boldsymbol{\gamma}) = \gamma_a \widetilde{\lambda}_{12}(t|t_1, a, \boldsymbol{X})$ *for* $a = 0, 1$*, for some* $\widetilde{\lambda}_{jk}$ *functions.*

(v) The probability density function of $\boldsymbol{\gamma}$*,* $f_{\boldsymbol{\theta}}(\boldsymbol{\gamma})$*, is known up to a finite dimensional parameter* $\boldsymbol{\theta}$ *that is identifiable from the observed data distribution.*

Then, the analogue of Proposition 4 is

Proposition A.2. *Under Assumptions 1 and A.2, the causal effects (2.1)–(2.3) are identified by*

$$\Pr[T_2(1) \le t|ad] - \Pr[T_2(0) \le t|ad]$$
$$= \int_{\mathcal{X}} \int_0^\infty \int_0^\infty [F_{2|T_1 \le T_2, A=1, \boldsymbol{x}, \gamma_1}(t) - F_{2|T_1 \le T_2, A=0, \boldsymbol{x}, \gamma_0}(t)\}] \frac{\eta_{A=0,\boldsymbol{x},\gamma_0}\eta_{A=1,\boldsymbol{x},\gamma_1} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{\gamma} d\boldsymbol{x}}{\int_{\mathcal{X}} \int_0^\infty \int_0^\infty \eta_{A=0,\boldsymbol{x}',\gamma_0'}\eta_{A=1,\boldsymbol{x}',\gamma_1'} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') f_{\boldsymbol{X}}(\boldsymbol{x}') d\boldsymbol{\gamma}' d\boldsymbol{x}'},$$

$$\Pr[T_2(1) \le t|nd] - \Pr[T_2(0) \le t|nd]$$
$$= \int_{\mathcal{X}} \int_0^\infty \int_0^\infty [F_{2|T_1 > T_2, A=1, \boldsymbol{x}, \gamma_1}(t) - F_{2|T_1 > T_2, A=0, \boldsymbol{x}, \gamma_0}(t)] \frac{(1-\eta_{A=0,\boldsymbol{x},\gamma_0})(1-\eta_{A=1,\boldsymbol{x},\gamma_1}) f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{\gamma} d\boldsymbol{x}}{\int_{\mathcal{X}} \int_0^\infty \int_0^\infty (1-\eta_{A=0,\boldsymbol{x}',\gamma_0'})(1-\eta_{A=1,\boldsymbol{x}',\gamma_1'}) f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') f_{\boldsymbol{X}}(\boldsymbol{x}') d\boldsymbol{\gamma}' d\boldsymbol{x}'},$$

$$\Pr[T_1(1) \le t|ad] - \Pr[T_1(0) \le t|ad]$$
$$= \int_{\mathcal{X}} \int_0^\infty \int_0^\infty [F_{1|T_1 \le T_2, A=1, \boldsymbol{x}, \gamma_1}(t) - F_{1|T_1 \le T_2, A=0, \boldsymbol{x}, \gamma_0}(t)\}] \frac{\eta_{A=0,\boldsymbol{x},\gamma_0}\eta_{A=1,\boldsymbol{x},\gamma_1} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{\gamma} d\boldsymbol{x}}{\int_{\mathcal{X}} \int_0^\infty \int_0^\infty \eta_{A=0,\boldsymbol{x}',\gamma_0'}\eta_{A=1,\boldsymbol{x}',\gamma_1'} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') f_{\boldsymbol{X}}(\boldsymbol{x}') d\boldsymbol{\gamma}' d\boldsymbol{x}'},$$

where for any event $\mathcal{Q}$, $F_{j|\mathcal{Q},\boldsymbol{x},\boldsymbol{\gamma}}(t) = \Pr(T_j \le t|\mathcal{Q}, \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\gamma})$. Furthermore, all the integrals can be consistently estimated from the data.

Proof. For $j = 1, 2$ and $a = 0, 1$ we may write

$$\begin{aligned}
\Pr[T_j(a) \le t|ad] &= E_{\boldsymbol{X},\boldsymbol{\gamma}|ad}\{\Pr[T_j(a) \le t \mid T_1(0) \le T_2(0), T_1(1) \le T_2(1), \boldsymbol{X}, \boldsymbol{\gamma}]\} \\
&= E_{\boldsymbol{X},\boldsymbol{\gamma}|ad}\{\Pr[T_j(a) \le t|T_1(a) \le T_2(a), \boldsymbol{X}, \gamma_a]\} \\
&= E_{\boldsymbol{X},\boldsymbol{\gamma}|ad}[\Pr(T_j \le t|T_1 \le T_2, A = a, \boldsymbol{X}, \gamma_a)] \\
&= \int_{\mathcal{X}} \int_0^\infty \int_0^\infty F_{j|T_1 \le T_2, A=a, \boldsymbol{x}, \gamma_a}(t) f(\boldsymbol{\gamma}, \boldsymbol{x}|ad) d\boldsymbol{\gamma} d\boldsymbol{x},
\end{aligned}$$

where the second line is by Part (i) of Assumption A.2, the third by Part (ii) of Assumption A.2, and the fourth by Part (iii) of Assumption A.2 and where $f(\boldsymbol{\gamma}, \boldsymbol{X}|ad)$ is the joint density function of $\boldsymbol{\gamma}$ and $\boldsymbol{X}$ within the ad stratum. Note that $\Pr[T_j \le t|T_1 \le T_2, A = a, \boldsymbol{X} = \boldsymbol{x}, \gamma_a]$ is identifiable from the data. Now, by

Bayes' Theorem,

$$f(\boldsymbol{\gamma}, \boldsymbol{x}|ad) = \frac{\Pr[T_1(0) \le T_2(0), T_1(1) \le T_2(1)|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\gamma}] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) f_{\boldsymbol{X}}(\boldsymbol{x})}{\int_{\mathcal{X}} \int_0^\infty \int_0^\infty \Pr[T_1(0) \le T_2(0), T_1(1) \le T_2(1)|\boldsymbol{X} = \boldsymbol{x}', \boldsymbol{\gamma}'] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') f_{\boldsymbol{X}}(\boldsymbol{x}') d\boldsymbol{\gamma}' d\boldsymbol{x}'}$$

$$= \frac{\Pr[T_1(0) \le T_2(0)|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\gamma}] \Pr[T_1(1) \le T_2(1)|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\gamma}] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) f_{\boldsymbol{X}}(\boldsymbol{x})}{\int_{\mathcal{X}} \int_0^\infty \int_0^\infty \Pr[T_1(0) \le T_2(0)|\boldsymbol{X} = \boldsymbol{x}', \boldsymbol{\gamma}'] \Pr[T_1(1) \le T_2(1)|\boldsymbol{X} = \boldsymbol{x}', \boldsymbol{\gamma}'] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') f_{\boldsymbol{X}}(\boldsymbol{x}') d\boldsymbol{\gamma}' d\boldsymbol{x}'}$$

$$= \frac{\Pr[T_1 \le T_2|A = 0, \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\gamma}] \Pr[T_1 \le T_2|A = 1, \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\gamma}] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) f_{\boldsymbol{X}}(\boldsymbol{x})}{\int_{\mathcal{X}} \int_0^\infty \int_0^\infty \Pr[T_1 \le T_2|A = 0, \boldsymbol{X} = \boldsymbol{x}', \gamma_0'] \Pr[T_1 \le T_2|A = 1, \boldsymbol{X} = \boldsymbol{x}', \gamma_1'] f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') f_{\boldsymbol{X}}(\boldsymbol{x}') d\boldsymbol{\gamma}' d\boldsymbol{x}'}$$

$$= \frac{\eta_{A=0,\boldsymbol{x},\boldsymbol{\gamma}} \eta_{A=1,\boldsymbol{x},\boldsymbol{\gamma}} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) f_{\boldsymbol{X}}(\boldsymbol{x})}{\int_{\mathcal{X}} \int_0^\infty \int_0^\infty \eta_{A=0,\boldsymbol{x}',\boldsymbol{\gamma}'} \eta_{A=1,\boldsymbol{x}',\boldsymbol{\gamma}'} f_{\boldsymbol{\theta}}(\boldsymbol{\gamma}') f_{\boldsymbol{X}}(\boldsymbol{x}') d\boldsymbol{\gamma}' d\boldsymbol{x}'},$$

where the first line is by Part (*iii*) of Assumption A.2, the second by Part (*i*) of Assumption A.2, and the third by Part (*ii*) of Assumption A.2. Note that $f_{\boldsymbol{X}}(\boldsymbol{x})$, the probability density function of $\boldsymbol{X}$ and $f_{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ are identifiable from the data. Therefore, estimation of $F_{j|T_1 \le T_2, A=a, \boldsymbol{x}, \gamma_a}(t)$ for $j = 1, 2$ and $a = 0, 1$ as well as estimation of $\eta_{A=0,\boldsymbol{X},\gamma_0}, \eta_{A=1,\boldsymbol{X},\gamma_1}$ and $\boldsymbol{\theta}$ is possible using, for example, the EM algorithm we proposed in Section 4.2 of the main text (see details in Section C.3). The proof for identification of $\Pr[T_2(a) \le t|nd]$ under Assumptions 1 and A.2 is similar.

□

B Sensitivity analysis without frailty assumptions

As can be seen from the bounds given in Equation (3.11) of the main text, out of the six probability functions appearing in (3.8)–(3.10), three functions are non-parametrically identified from the data, $\Pr[T_2(0) \le t \mid ad], \Pr[T_2(1) \le t \mid nd]$ and $\Pr[T_1(0) \le t \mid ad]$, while the three other functions, $\Pr[T_1(1) \le t \mid ad], \Pr[T_2(1) \le t \mid ad]$, and $\Pr[T_2(0) \le t \mid nd]$, are only partially non-parametrically identified. Using

Bayes' theorem, under Assumptions 1–3

$$\Pr[T_2(1) \le t \mid ad] = \Pr[T_1(0) < T_2(0) | T_2(1) \le t] \frac{F_{2|A=1}(t)}{\eta_{A=0}}, \tag{B.2}$$

$$\Pr[T_1(1) \le t \mid ad] = \Pr[T_1(0) < T_2(0) | T_1(1) \le t] \frac{F_{1|A=1}(t)}{\eta_{A=0}}, \tag{B.3}$$

$$\Pr[T_2(0) \le t \mid nd] = \Pr[T_1(1) > T_2(1) | T_2(0) \le t] \frac{F_{2|A=0}(t)}{1 - \eta_{A=1}}. \tag{B.4}$$

Focusing temporarily on (B.2), the unidentified component $\Pr[T_1(0) < T_2(0)|T_2(1) \le t]$ is the proportion of ad out of those who would have died by time t, under $A = 1$. We propose the following sensitivity analysis, partially inspired by Shepherd et al. (2007): let $g(t; \xi^{ad,T_2}) = \Pr[T_1(0) < T_2(0)|T_2(1) \le t]$ be a working parametric function. One natural choice is $g(t; \xi^{ad,T_2}) = \exp(\xi_0^{ad,T_2} + \xi_1^{ad,T_2} t)/[1 + \exp(\xi_0^{ad,T_2} + \xi_1^{ad,T_2} t)]$. For a fixed value of ξ_1^{ad,T_2}, ξ_0^{ad,T_2} is identified because $\lim_{t\to\infty} g(t; \xi^{ad,T_2}) = \Pr[T_1(0) < T_2(0)] = \eta_{A=0}$. Practically, we can choose a value $\widetilde{t^\star}$ representing a time by which nearly all the population is expected to have the terminal event. That is, $\widetilde{t^\star}_2$ such that $Pr[T_2(1) > \widetilde{t^\star}_2] < \epsilon$ for some small ϵ. Then, we set $\xi_0^{ad,T_2} = \text{logit}(\eta_{A=0}) - \xi_1^{ad,T_2} \widetilde{t^\star}_2$. The analyst can examine plausible values of ξ_1^{ad,T_2}, each gives rise to a different $\Pr[T_1(0) < T_2(0)|T_2(1) \le t]$, and hence to different $\Pr[T_1(1) \le t \mid ad] - \Pr[T_1(0) \le t \mid ad]$. By specifying $\xi_1^{ad,T_2} > 0$, the proportion of ad out of those who would have died under $A = 1$ increases with t. This makes sense if one expects that many of those who would have died early, would have died before experiencing the disease, or if generally speaking, death following disease diagnosis at a young age does not tend to be immediate. This is probably the case for AD, which is less likely to be the reason of death for those who died early (as oppose to, e.g., aggressive cancer tumors), even if they had APOE. $\xi_1^{ad,T_2} < 0$ is be interpreted in an analogue fashion.

We employ the same strategy in constructing sensitivity analyses for (B.3) and (B.4). $\Pr[T_1(0) < T_2(0)|T_1(1) \le t]$ can be understood as the proportion of ad out of those who would have had AD by time t if they had APOE. Therefore, if we adopt the working parametric function $g(t; \xi^{ad,T_1}) = \exp(\xi_0^{ad,T_1} + \xi_1^{ad,T_1} t)/[1 + \exp(\xi_0^{ad,T_1} + \xi_1^{ad,T_1} t)]$, then ξ_0^{ad,T_1} is identified for a fixed value of ξ_1^{ad,T_1}. Here, because not everyone are expected to be diagnosed with AD, we can think of a time $\widetilde{t^\star}_1$ by which nearly all of those who would have diagnosed with AD under APOE, would already be diagnosed. Then, $\Pr[T_1(0) < T_2(0)|T_1(1) < \widetilde{t^\star}_1]$ is approximately the proportion of ad our of those who would have been diagnosed with AD at some point of their life under APOE, i.e. $\pi_{ad}/(\pi_{ad} + \pi_{dh})$. Therefore, $\xi_0^{ad,T_1} = \text{logit}(\eta_{A=0}/\eta_{A=1}) - \xi_1^{ad,T_2}\widetilde{t^\star}_2$.

Turning to (B.4), $\Pr[T_1(1) > T_2(1)|T_2(0) \le t]$ is the proportion of nd out of those who would have died without APOE. If we adopt the working parametric function $g(t;\xi^{nd,T_2}) = \text{expit}(\xi_0^{nd,T_2} + \xi_1^{nd,T_2}t)$, then ξ_0^{nd,T_2} is identified for a fixed value of ξ_1^{nd,T_2} by $\xi_0^{nd,T_2} = \text{logit}(1-\eta_{A=1}) - \xi_1^{nd,T_2}\widetilde{t^\star}_1$.

C Theory and details on the estimation methods

C.1 Detailed calculations for the nonparametric estimators

We present here the calculations leading to the expressions given in Equations (4.17)–(4.20) of the main text, that underpin the proposed non-parametric estimation. Let $f_2(s|\mathcal{Q})$ be the probability density function of T_2 conditionally on $\mathcal{Q}$, and write

$$\begin{aligned}\eta_{A=a,T_2\le t} &= \int_0^\infty \Pr(T_1 \le s|A=a, T_2=s, T_2 \le t) f_2(s|A=a, T_1 \le T_2)\\ &= -\frac{1}{F_{2|A=a}(t)}\int_0^t F_{1|A=a,T_2=s}(s)dS_{2|A=a}(s)\end{aligned} \tag{C.5}$$

and using similar arguments,

$$\begin{aligned}S_{1|A=a,T_1\le T_2}(t) &= \int_0^\infty \Pr(T_1 > t|A=a, T_1\le T_2, T_2=s) f_2(s|A=a, T_1\le T_2)ds\\ &= -\frac{1}{\eta_{A=a}}\int_t^\infty \Pr(T_1 > t|A=a, T_1 \le s, T_2=s)\Pr(T_1\le s|A=a,T_2=s)dS_{2|A=a}(s)\\ &= -\frac{1}{\eta_{A=a}}\int_t^\infty [S_{1|A=a,T_2=s}(t) - S_{1|A=a,T_2=s}(s)]dS_{2|A=a}(s),\\ S_{2|A=a,T_1\le T_2}(t) &= 1 - \frac{\eta_{A=a,T_2\le t}F_{2|A=a}}{\eta_{A=a}},\\ S_{2|A=a,T_1> T_2}(t) &= 1 - \frac{(1-\eta_{A=a,T_2\le t})F_{2|A=a}}{1-\eta_{A=a}}.\end{aligned}$$

C.2 Asymptotic properties of the nonparametric estimators

Denote $\widehat{S}_{2|A=a}(t) \equiv \widehat{S}_{2|a}(t)$ for the Kaplan-Meier estimator, and $\widehat{F}_{1|A=a,T_2=s}(t) \equiv \widehat{F}_{1|a,s}(t)$ for the proposed Beran (1981) estimator of $F_{1|A=a,T_2=s}(t)$. We present a sketch of the proof for consistency and asymptotic normality of the estimator for the more complex term $\eta_{A=a,T_2\leq t}$. That is, for consistency and asymptotic normality of

$$\frac{1}{\widehat{F}_{2|a}(t)} \int_0^t \widehat{F}_{1|a,s}(s) d\widehat{S}_{2|a}(s) \tag{C.6}$$

for any $t \in [\underline{\tau}(a), \overline{\tau}(a)]$, where $\underline{\tau}(a) > 0$ ensures that $\widehat{F}_{2|A=a}(t)$ is bounded away from zero and $\overline{\tau}(a) = \inf\{t : \Pr(\min(T_2, C) \leq t|A = a) < 1)\}$ is the standard condition for KM consistency. In principal, the results below for the smoothed KM estimator also demand that $\overline{\tau}(a) < \sup\{t : \Pr(\min\{T_1, C\} > t|A = a, T_2 = s) > 0\}$ but this condition is met automatically in our case, as for any t, among those with $T_2 = t$, there are always people who died without experiencing the non-terminal event.

Interestingly, the integral $\int_0^t \widehat{F}_{1|a,s}(s) d\widehat{S}_{2|a}(s)$ can be seen as a special case of the estimator presented in Equation (2.4) in Akritas and Van Keilegom (2003) for a bivariate distribution function (in the non semi-competing case); this can be seen by plugging in $w(y) = 1$ in Equation (2.4) in Akritas and Van Keilegom (2003). Therefore, consistency and asymptotic normality of $\int_0^t \widehat{F}_{1|a,s}(s) d\widehat{S}_{2|a}(s)$ follow from the results in Akritas and Van Keilegom (2003).

Consistency and asymptotic normality of the KM estimator $\widehat{F}_{2|a}(t)$ is standard (Andersen et al., 1993). Therefore, consistency of (C.6) follows from Slutzki's Theorem. Regarding asymptotic normality of (C.6), by Cramér-Wold device we have that for any $t \in (\underline{\tau}(a), \overline{\tau}(a))$ the joint distribution of

$$\sqrt{n}\left(\widehat{F}_{2|A=a}(t), \int_0^t \widehat{F}_{1|A=a,T_2=s}(s) d\widehat{S}_{2|A=a}(s)\right)$$

is multivariate normal. Therefore, asymptotic normality of (C.6) is obtained by applying the extended continuous mapping theorem (Theorem 7.24 of Kosorok (2007)).

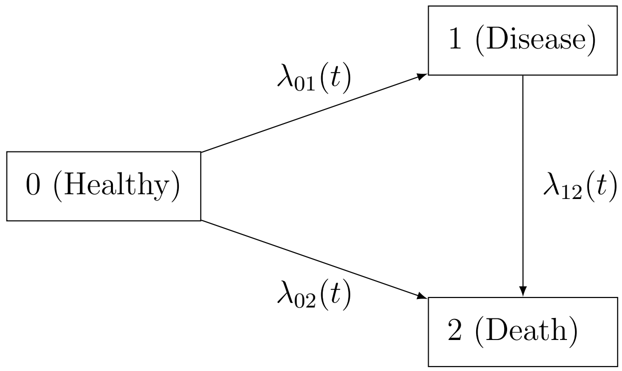

Figure C.1: Illustration of the semi-competing risks data

C.3 Details on the semi-parametric model and estimation by the EM algorithm

C.3.1 EM algorithm

The EM algorithm hinges on the fact that, given A_i, γ_{A_i}, $i = 1, \ldots, n$, the six partial likelihoods for estimating $\boldsymbol{\beta}$ are given by

$$\prod_{i=1}^{n} \left\{ \frac{\gamma_a \exp\left(\boldsymbol{X}_i^T \beta_{01}^a\right)}{\sum_{j=1}^{n} I(\widetilde{T}_{i1} \leq \widetilde{T}_{j2}) I(A_j = a) \gamma_a \exp\left(X_j^T \beta_{01}^a\right)} \right\}^{\delta_{i1} I(A_i = a)}, \quad a = 0, 1\,,$$

$$\prod_{i=1}^{n} \left\{ \frac{\gamma_a \exp\left(\boldsymbol{X}_i^T \beta_{02}^a\right)}{\sum_{j=1}^{n} I(\widetilde{T}_{i1} \leq \widetilde{T}_{j2}) I(A_j = a) \gamma_a \exp\left(X_j^T \beta_{02}^a\right)} \right\}^{(1-\delta_{i1})\delta_{i2} I(A_i = a)}, \quad a = 0, 1\,,$$

$$\prod_{i=1}^{n} \left\{ \frac{\gamma_a \exp\left(\boldsymbol{X}_i^T \beta_{12}^a\right)}{\sum_{j=1}^{n} I(\widetilde{T}_{j1} \leq \widetilde{T}_{i2} \leq \widetilde{T}_{j2}) I(A_j = a) \gamma_a \exp\left(X_j^T \beta_{12}^a\right)} \right\}^{\delta_{i1}\delta_{i2} I(A_i = a)}, \quad a = 0, 1\,.$$

The six respective Breslow-type estimators, given γ_{A_i}, $i = 1, \ldots, n$, and β_{jk}^a are

$$\widetilde{\Delta}\Lambda_{01}^{0a}(t) = \frac{\sum_{i=1}^{n} \delta_{i1} I(A_i = a) I(\widetilde{T}_{i1} = t)}{\sum_{i=1}^{n} I(t \leq \widetilde{T}_{i1}) I(A_i = a) \gamma_a \exp\left(\boldsymbol{X}_i^T \beta_{01}^a\right)}, \quad a = 0, 1\,,$$

$$\widetilde{\Delta}\Lambda_{02}^{0a}(t) = \frac{\sum_{i=1}^{n} (1 - \delta_{i1}) \delta_{i2} I(A_i = a) I(\widetilde{T}_{i1} = t)}{\sum_{i=1}^{n} I(t \leq \widetilde{T}_{i1}) I(A_i = a) \gamma_a \exp\left(\boldsymbol{X}_i^T \beta_{02}^a\right)}, \quad a = 0, 1\,,$$

$$\widetilde{\Delta}\Lambda_{12}^{0a}(t) = \frac{\sum_{i=1}^{n} \delta_{i1}\delta_{i2} I(A_i = a) I(\widetilde{T}_{i2} = t)}{\sum_{i=1}^{n} I(\widetilde{T}_{i1} \leq t \leq \widetilde{T}_{i2}) I(A_i = a) \gamma_a \exp\left(\boldsymbol{X}_i^T \beta_{12}^a\right)} \quad a = 0, 1\,.$$

For the E-step, the posterior expectation of the frailty variate γ_{A_i} given the entire observed data equals to the posterior expectation given participant i observed data, $E\left(\gamma_{A_i}|\mathcal{D}_i\right)$, is given by

$$\begin{aligned} E\left(\gamma_{A_i}|\mathcal{D}_n\right) &= \int_0^\infty \gamma_{A_i} f(\gamma_{A_i}|\mathcal{D}_n) d\gamma_{A_i} \\ &= \frac{\int_0^\infty \gamma_{A_i} f(\mathcal{D}_i|\gamma_{A_i}) f_{\theta_{A_i}}(\gamma_{A_i}) d\gamma_{A_i}}{\int_0^\infty f(\mathcal{D}_i|\gamma_{A_i}) f_{\theta_{A_i}}(\gamma_{A_i}) d\gamma_{A_i}} = \frac{(-1)^{\delta_{i\cdot}+1}\phi_{A_i}^{(\delta_{i\cdot}+1)}(s_i)}{(-1)^{\delta_{i\cdot}}\phi_{A_i}^{(\delta_{i\cdot})}(s_i)}, \end{aligned}$$

for $i = 1, \ldots, n$, and where $f_{\theta_a}(\gamma_a)$ is the marginal probability density function of γ_a.

Finally estimation of θ_a is carried out in each M-step by maximizing the conditional expectation of the log-likelihood of γ_a, given the observed data and the current value of the parameters. Denote this function by $g(\theta_a)$, which can be written as

$$g(\theta_a) = \frac{1}{n_a} \sum_{i=1}^{n} I(A_i = a) \widehat{E}\left\{\log f_{\theta_a}(\gamma_{A_i})|\mathcal{D}_i, \boldsymbol{\beta}^a, \boldsymbol{\Lambda}_0^a\right\}$$

where $n_a = \sum_{i=1}^n I(A_i = a)$, $\boldsymbol{\beta}^a = (\boldsymbol{\beta}_{01}^{aT}, \boldsymbol{\beta}_{02}^{aT}, \boldsymbol{\beta}_{12}^{aT})^T$, and $\boldsymbol{\Lambda}_0^a = (\Lambda_{01}^0(\cdot|a), \Lambda_{02}^0(\cdot|a), \Lambda_{12}^0(\cdot|a))$ and $\widehat{E}$ is the expectation under the current parameter values. Note that in this case, $\boldsymbol{\theta} = (\theta_0, \theta_1)$.

In summary, the following is our proposed estimation procedure. Initial values are obtained by six standard Cox regression models for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}_0$ assuming $\gamma_{A_i} = 1$, $i = 1, \ldots, n$. Then the following three steps are iterated until convergence is met.

Step 1. Use current values of $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}_0$, and $\boldsymbol{\theta}$, to get $\widehat{E}\left(\log \gamma_{A_i}|\mathcal{D}_i\right)$ and $\widehat{E}\left(\gamma_{A_i}|\mathcal{D}_i\right)$, $i = 1, \ldots, n$.

Step 2. Estimate $\boldsymbol{\beta}$ with six separate Cox model analysis, each with covariates $\boldsymbol{X}_i$ and offset terms $\log \widehat{E}\left(\gamma_{A_i}|\mathcal{D}_i\right)$ (which replaces the unknown γ_{A_i}). Estimate $\boldsymbol{\Lambda}_0$ using Breslow estimators with $\log \widehat{E}\left(\gamma_{A_i}|\mathcal{D}_i\right)$. This can be done by coxph package of R.

Step 3. Estimate θ_a by maximizing $g(\theta_a)$, $a = 0, 1$.

C.3.2 The EM algorithm under Gamma frailty

Arguably the most common choice for the frailty variate distribution is the Gamma distribution with mean 1 and variance θ. If we assume that γ_{ai}, $i = 1, \dots, n$, are Gamma distributed with mean 1 and variance θ_a, then Step 1 above is based on

$$\widehat{E}\left(\log \gamma_{A_i} | \mathcal{D}_i\right) = \Psi\left(\frac{1}{\theta_{A_i}} + \delta_{i\cdot}\right) - \log\left(\frac{1}{\theta_{A_i}} + s_i\right), \quad i = 1, \dots, n\,,$$

where $\Psi(x) = \Psi'(x)/\Psi(x)$ the digamma function, and

$$\widehat{E}\left(\gamma_{A_i} | \mathcal{D}_i\right) = \frac{\theta_{A_i}^{-1} + \delta_{i\cdot}}{\theta_{A_i}^{-1} + s_i}.$$

Step 3 is carried out by maximizing

$$\begin{aligned} g(\theta_a) &= \frac{1}{n_a}\sum_{i=1}^{n} I(A_i = a)\widehat{E}\left\{\log f_{\theta_a}(\gamma_{A_i}) | \mathcal{D}_i\right\} \\ &= -\frac{1}{\theta_a}\log\left(\frac{1}{\theta_a}\right) + \left(\frac{1}{\theta_a} - 1\right)\frac{1}{n_a}\sum_{i=1}^{n} I(A_i = a)\widehat{E}\left(\log \gamma_{A_i} | \mathcal{D}_i\right) \\ &\quad \frac{1}{\theta_a}\frac{1}{n_a}\sum_{i=1}^{n} I(A_i = a)\widehat{E}\left(\gamma_{A_i} | \mathcal{D}_i\right) - \log\Gamma\left(\frac{1}{\theta_a}\right) \end{aligned}$$

as a function of θ_a for $a = 0, 1$.

D Additional details on the simulation study

We first note that code, simulation scripts, and results are available through our **R** package `CausalSemiComp` and the Github of the first author (Repository name: `CausalSemiCompReproduce`)

D.1 Details on the data-generating mechanism

The function `SimDataWeibFrail` from the **R** package `CausalSemiComp` accompanying this paper was use to simulate data in all simulation studies. In all simulation studies, potential outcomes $\{T_1(0), T_2(0)\}$ and $\{T_1(1), T_2(1)\}$ were initially simulated from models (4.21) with two covariates: $X_1 \sim Ber(0.5)$ and $X_2 \sim N(0,1)$, and with Gamma frailty variables γ_0, γ_1 and with $\theta_0 = \theta_1 = \theta$. Observed data was created by simulating A, with $\Pr(A = 1) = 0.5$ and determining which pair is potentially observed, $\{T_1(0), T_2(0)\}$ or $\{T_1(1), T_2(1)\}$ according to A value. We took Weibull baseline hazards (see details below), under a parameterization such that for a random variable $V \sim Weibull(\widetilde{\alpha}, \widetilde{\mu})$, $Pr(V > v) = \exp[-(v/\widetilde{\mu})^{\widetilde{\alpha}}]$. For calculating the large-sample bounds (Figure 1) and the RMST causal effects (Figure 2), no censoring was applied. True causal effects were calculated by a Monte Carlo simulation with sample size 10^7. For the simulation studies, we took administrative censoring at time 100, and additional Exponential censoring time, shared between $a = 0, 1$, with mean determined to obtain reported censoring rates.

For the simulations calculating the large-sample bounds and assessing the non-parametric estimation methods, we made sure that Assumption 3 holds by throwing out observations and re-simulate whenever a *dp* observation was simulated. The estimation procedure for these scenarios was non-parametric, so the fact Model (4.20) no longer holds with Weibull baseline hazards, due to throwing out observations just changed the DGM and the estimators were still expected to perform well. The DGM of Scenario (I) also implies that Assumption 4 holds as we verified from the simulated data. For the nonparametric simulations, Scenarios (I)–(III), we also re-simulated the data whenever $T_2(a)$ was too large, to avoid heavy tails of the distribution that are unlikely for realistic data. Table D.1 gives the parameter values for scenarios (I)–(III), and Figure D.2 presents the resulting cumulative distribution functions within the different strata. At each figure the difference between the two curves is the true causal effects portrayed by the solid black line in Figure 1. For Scenarios (II) and (III), within each stratum the CDF is the same for $a = 0, 1$ and causal effects are zero.

For the simulations assessing the semi-parametric estimation methods, data at each simulation iteration were simulated under model (4.20), with Gamma frailty distribution with shared variance parameter θ, and without further restrictions; parameter values are given in Table D.2. Censoring time was simulated as described above. For RMST mean and median causal effects, we considered a target population with equal percentages (for simplicity) of 18 covariate profiles generated from the cross-product values of $X_1 \in \{0, 1\}$ and $X_2 \in \{-2, -1.5, -1, 0, , ..., 2\}$. Then true effects were calculated by simulating potential outcomes $\{T_1(0), T_2(0)\}$ and $\{T_1(1), T_2(1)\}$ for a population of 10^8 individuals from this population structure, and RMST casual effects were calculated simply as the appropriate differences between means and medians within relevant strata.

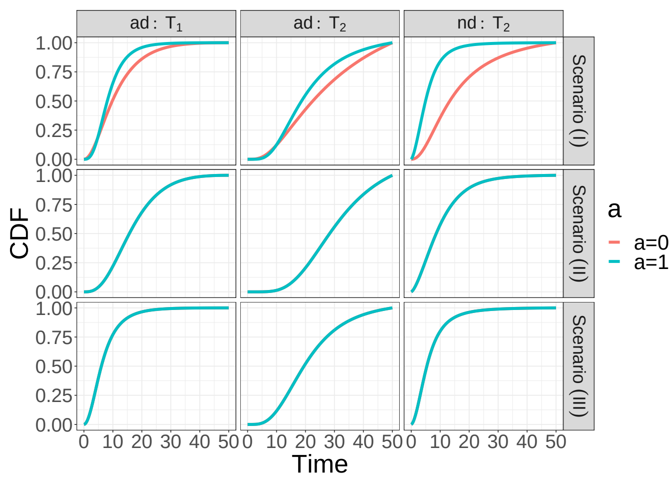

Figure D.2: True stratum-specific CDF of $T_1(a)$ and $T_2(a)$ for Scenarios (I)–(III). In Scenarios (II) and (III), CDFs are the same for $a = 0$ and $a = 1$.

Table D.1: Parameter specifications for Scenarios (I)–(III)

Scenario	Baseline Shapes	Baseline Scales	Covariate coefficients	Frailty variance and correlation
(I)	$\widetilde{\alpha}_{01}^{0} = 2$	$\widetilde{\mu}_{01}^{0} = 12.5$	$\boldsymbol{\beta}_{01}^{0} = (\log(0.25), 0)$	
	$\widetilde{\alpha}_{02}^{0} = 2.25$	$\widetilde{\mu}_{02}^{0} = 25$	$\boldsymbol{\beta}_{02}^{0} = (0, 0)$	$\theta_0 = \theta_1 = 2$
	$\widetilde{\alpha}_{12}^{0} = 1.5$	$\widetilde{\mu}_{12}^{0} = 25$	$\boldsymbol{\beta}_{12}^{0} = (0, 0)$	(Kendall's τ=0.5)
	$\widetilde{\alpha}_{01}^{1} = 3$	$\widetilde{\mu}_{01}^{1} = 10$	$\boldsymbol{\beta}_{01}^{1} = (0, 0)$	$\rho = 0.75$
	$\widetilde{\alpha}_{02}^{1} = 1.5$	$\widetilde{\mu}_{02}^{1} = 17.5$	$\boldsymbol{\beta}_{02}^{1} = (0, 0)$	
	$\widetilde{\alpha}_{12}^{1} = 2.5$	$\widetilde{\mu}_{12}^{1} = 20$	$\boldsymbol{\beta}_{12}^{1} = (0, 0)$	
(II)	$\widetilde{\alpha}_{01}^{0} = 3$	$\widetilde{\mu}_{01}^{0} = 20$	$\boldsymbol{\beta}_{01}^{0} = (0, 0)$	
	$\widetilde{\alpha}_{02}^{0} = 1.5$	$\widetilde{\mu}_{02}^{0} = 15$	$\boldsymbol{\beta}_{02}^{0} = (0, 0)$	$\theta_0 = \theta_1 = 2/3$
	$\widetilde{\alpha}_{12}^{0} = 2.75$	$\widetilde{\mu}_{12}^{0} = 25$	$\boldsymbol{\beta}_{12}^{0} = (0, 0)$	(Kendall's $\tau = 0.25$)
	$\widetilde{\alpha}_{01}^{1} = 3$	$\widetilde{\mu}_{01}^{1} = 20$	$\boldsymbol{\beta}_{01}^{1} = (0, 0)$	$\rho = 0$
	$\widetilde{\alpha}_{02}^{1} = 1.5$	$\widetilde{\mu}_{02}^{1} = 15$	$\boldsymbol{\beta}_{02}^{1} = (0, 0)$	
	$\widetilde{\alpha}_{12}^{1} = 2.75$	$\widetilde{\mu}_{12}^{1} = 25$	$\boldsymbol{\beta}_{12}^{1} = (0, 0)$	
(III)	$\widetilde{\alpha}_{01}^{0} = 2$	$\widetilde{\mu}_{01}^{0} = 7.5$	$\boldsymbol{\beta}_{01}^{0} = (0, 0)$	
	$\widetilde{\alpha}_{02}^{0} = 1.75$	$\widetilde{\mu}_{02}^{0} = 15$	$\boldsymbol{\beta}_{02}^{0} = (0, 0)$	$\theta_0 = \theta_1 = 2/3$
	$\widetilde{\alpha}_{12}^{0} = 2.5$	$\widetilde{\mu}_{12}^{0} = 20$	$\boldsymbol{\beta}_{12}^{0} = (0, 0)$	(Kendall's $\tau = 0.25$)
	$\widetilde{\alpha}_{01}^{1} = 2$	$\widetilde{\mu}_{01}^{1} = 7.5$	$\boldsymbol{\beta}_{01}^{1} = (0, 0)$	$\rho = 0$
	$\widetilde{\alpha}_{02}^{1} = 1.75$	$\widetilde{\mu}_{02}^{1} = 15$	$\boldsymbol{\beta}_{02}^{1} = (0, 0)$	
	$\widetilde{\alpha}_{12}^{1} = 2.5$	$\widetilde{\mu}_{12}^{1} = 20$	$\boldsymbol{\beta}_{12}^{1} = (0, 0)$	

Table D.2: Parameter specifications for Scenarios (I)–(III)

Scenario	Baseline Shapes	Baseline Scales	Covariate coefficients	Frailty variance and correlation
(IV)	$\widetilde{\alpha}_{01}^{0} = 2/5$	$\widetilde{\mu}_{01}^{0} = 4$	$\boldsymbol{\beta}_{01}^{0} = (\log(0.25), \log(3))$	
	$\widetilde{\alpha}_{02}^{0} = 2$	$\widetilde{\mu}_{02}^{0} = 5$	$\boldsymbol{\beta}_{02}^{0} = (\log(0.75), \log(1.5))$	Sub scenarios: $\theta_0 = \theta_1 = 2/3, 1, 2$
	$\widetilde{\alpha}_{12}^{0} = 2.5$	$\widetilde{\mu}_{12}^{0} = 15$	$\boldsymbol{\beta}_{12}^{0} = (0, \log(2))$	(Kendall's $\tau = 1/4, 1/3, 1/2$)
	$\widetilde{\alpha}_{01}^{1} = 2.5$	$\widetilde{\mu}_{01}^{1} = 2$	$\boldsymbol{\beta}_{01}^{1} = (0, 0)$	Variety of $\rho \in [0, 1]$
	$\widetilde{\alpha}_{02}^{1} = 2$	$\widetilde{\mu}_{02}^{1} = 3$	$\boldsymbol{\beta}_{02}^{1} = (\log(0.75), \log(1.5))$	
	$\widetilde{\alpha}_{12}^{1} = 2.5$	$\widetilde{\mu}_{12}^{1} = 10$	$\boldsymbol{\beta}_{12}^{1} = (\log(0.5), \log(2))$	

D.2 Details on the different analyses

D.2.1 Non-parametric simulations

The non-parametric estimation was carried out for each simulation iteration as described in Section 4.1 of the main text. That is, first the standard KM estimator was calculated for $S_{2|A=a}(t)$ at each intervention group $a = 0, 1$. Then, the smoothed KM estimator (Beran, 1981) was calculated for $S_{1|A=a,T_2=s}(t)$, within each intervention group $a = 0, 1$, and for all s values for which terminal events were observed at each group. This was done using our new **R** package, `CausalSemiComp` that utilize the `prodlim` package (Gerds, 2019) for calculating the smoothed KM estimator. Then, estimators $\widehat{S}_{1|A=a}(t)$, $\widehat{\eta}_{A=a}$, $\widehat{\eta}_{A=a,T_2\leq t}$, $\widehat{S}_{2|A=a,T_1\leq T_2}(t)$ and $\widehat{S}_{2|A=a,T_1>T_2}(t)$, were obtained by plugging-in the KM and the smoothed KM estimators into Equations (4.14)–(4.19). Standard errors were estimated using bootstrap with 100 repetitions, and Wald-type 95% confidence intervals were then calculated.

D.2.2 Semi-parametric simulations

For each dataset, we first estimated the statistical parameters $\boldsymbol{\psi}$ using the EM algorithm we proposed to use. A single $\hat{\theta}$ was taken as the mean of the $\hat{\theta}_0$ and $\hat{\theta}_1$ obtained through the EM estimation. The RMST causal effects were then estimated using a Monte Carlo estimation that, using the obtained $\hat{\boldsymbol{\psi}}$, simulated data of sample size 180,000 as follows:

1. For each value of $\boldsymbol{X}$ in the target population, simulate 100 $\boldsymbol{\gamma}$ values using $\hat{\theta}$, assuming ρ is known.

2. For each of the obtained $\boldsymbol{\gamma}$, calculate the six hazard functions (three under each a value) using the $\boldsymbol{X}$ profile, the simulated $\boldsymbol{\gamma}$, and the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}_0$.

3. For each of the 100 values (per $\boldsymbol{X}$), simulate 100 quadruplets $\{T_1(0), T_2(0), T_1(1), T_2(1)\}$.

These three steps result in population of size $18 \times 100 \times 100$. Then RMST effects are estimated by calculating difference between means and medians as appropriate within relevant strata. The choice to have 100 repetitions of each form ($\boldsymbol{\gamma}$ per $\boldsymbol{X}$ and Potential outcomes given $\boldsymbol{\gamma}, \boldsymbol{X}$) was made arbitrarily, and only minor reduction in the standard error was observed when we increased these numbers to 250.

Standard errors for both the statistical parameters and the causal effects were estimated using bootstrap with 100 repetitions, and Wald-type 95% confidence intervals were then calculated.

D.3 Simulation results

We present in this section results for the non-parametric and semi-parametric estimation methods. For the non-parameteric estimators, the following results are presented.

- For the non-parametric estimators, Table D.3 present results for the estimation of $\eta_{A=a}, S_{1|A=a}(10), S_{1|A=a}(30)$ under Scenarios (I)–(III), negligible bias was observed, standard errors were well-estimated and the confidence intervals exhibited good empirical coverage rate.

- Tables D.4 and D.5 presents the performance of estimators for $\eta_{A=a,T_2 \leq t}$ and $S_{A=a,T_1 \leq T_2}(t)$ for $A = 0, 1$ and $t = 10, 20$ for Scenarios (I), (II) (Table D.4), and (III) D.5. Bias was small, standard errors were well-estimated and were increased when censoring was more substantial. Empirical coverage rate by the confidence intervals was generally good. When the true probabilities were very small, finite sample bias was observed and decreased coverage rate; these phenomenons are expected to disappear for larger sample size.

For the semi-parametric estimation method, the following results are presented.

- Tables D.6, D.7 and D.8 present the results for $\rho = 0, 0.5, 1$. Note, in all these analyses ρ was assumed to be known, as the goal was to focus on the statistical properties of the estimation method. In all three tables it can be seen that minimal bias was observed, the standard errors were well-estimated, and empirical coverage rate was satisfactory.

- Tables D.9 and D.10 present the performance of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ estimators, respectively, for different θ values and censoring rates, for $\rho = 0.5$. For all individual entries of $\boldsymbol{\beta}$ and for all scenarios, bias was negligible, standard errors were well-estimated and the empirical coverage rate was close to the desirable 95%.

- Tables D.11, D.12 and D.13 present the performance of $\widehat{\Lambda}_{01}^{0,a}(t)$, $\widehat{\Lambda}_{02}^{0,a}(t)$, and $\widehat{\Lambda}_{12}^{0,a}(t)$, respectively, each for $a = 0, 1$. Results are presented for different θ values and censoring rates, under for $\rho = 0.5$. The values for t in each table were chosen to obtain a range of values for the different baseline hazards - without having extremely small or extremely large values. Generally, the baseline hazards were well-estimated with minimal finite-sample bias. Standard errors were well estimated and empirical coverage of the confidence intervals was as desired.

- Table D.14 reports the performance of the frailty variance estimator $\hat{\theta}$ for different θ and ρ values, and for different censoring rates. Minimal bias was observed, which is expected to disappear for larger sample size. Standard errors were well estimated, and empirical coverage rate was only slightly below the desired 95% level.

Table D.3: Performance of the proposed non-parametric estimators. Censoring rates considered were low (10%, C-L) and moderate (30%, C-M). True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\eta_{A=0}$		$\eta_{A=1}$		$S_{1\|A=0}(10)$		$S_{1\|A=0}(30)$		$S_{1\|A=1}(10)$		$S_{1\|A=1}(30)$	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
Scenario (I)												
True	0.491		0.848		0.748		0.523		0.481		0.158	
Mean.EST	0.491	0.492	0.845	0.844	0.749	0.748	0.524	0.523	0.479	0.476	0.159	0.159
EMP.SD	0.016	0.020	0.011	0.012	0.014	0.016	0.016	0.020	0.017	0.019	0.012	0.012
EST.SE	0.017	0.019	0.011	0.012	0.014	0.017	0.017	0.019	0.017	0.019	0.011	0.012
CP95%	0.946	0.944	0.951	0.948	0.950	0.949	0.945	0.940	0.953	0.942	0.944	0.951
Scenario (II)												
True	0.144		0.431		0.968		0.867		0.604		0.158	
Mean.EST	0.146	0.145	0.429	0.426	0.968	0.967	0.865	0.866	0.904	0.903	0.601	0.599
EMP.SD	0.012	0.015	0.016	0.018	0.006	0.007	0.012	0.014	0.010	0.012	0.016	0.019
EST.SE	0.012	0.015	0.016	0.019	0.006	0.007	0.012	0.015	0.010	0.013	0.016	0.019
CP95%	0.945	0.946	0.938	0.937	0.939	0.942	0.943	0.949	0.953	0.956	0.945	0.941
Scenario (III)												
True	0.726		0.938		0.440		0.278		0.277		0.067	
Mean.EST	0.729	0.728	0.935	0.934	0.435	0.433	0.274	0.274	0.274	0.271	0.068	0.068
EMP.SD	0.014	0.015	0.007	0.008	0.017	0.019	0.014	0.015	0.015	0.018	0.007	0.008
EST.SE	0.015	0.015	0.007	0.008	0.017	0.019	0.015	0.015	0.015	0.018	0.007	0.008
CP95%	0.936	0.943	0.942	0.930	0.934	0.929	0.931	0.937	0.931	0.928	0.943	0.938

E Additional information on the illustrative data analysis

E.1 Details on the semi-parametric analysis

For each dataset, we first estimated the statistical parameters $\boldsymbol{\psi}$ using the EM algorithm we proposed to use. A single $\hat{\theta}$ was taken as the weighted mean of the $\hat{\theta}_0$ and $\hat{\theta}_1$ obtained through the EM estimation. The RMST causal effects were then estimated using a Monte Carlo estimation that, using the obtained $\hat{\boldsymbol{\psi}}$, simulated 180,000 people as follows:

1. For each value of $\boldsymbol{X}$ in the target population, simulate 100 $\boldsymbol{\gamma}$ values using $\hat{\theta}$, assuming ρ is known.

2. For each of the obtained $\boldsymbol{\gamma}$, calculate the six hazard functions (three under each a value) using the $\boldsymbol{X}$ profile, the simulated $\boldsymbol{\gamma}$, and the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}_0$.

3. For each of the 100 values (per $\boldsymbol{X}$), simulate 100 the quadruple $\{T_1(0), T_2(0), T_1(1), T_2(1)\}$.

These three steps result in population of size $18 \times 100 \times 100$. Then RMST effects are estimated by calculating difference between means and medians as appropiate within relevant strata. The choice to have 100 repetitions of each form ($\boldsymbol{\gamma}$ per $\boldsymbol{X}$ and Potential outcomes given $\boldsymbol{\gamma}, \boldsymbol{X}$) was made arbitrarily, and only minor reduction in the standard error was observed when we increased these numbers to 250.

E.2 Additional results

Table D.4: Performance of the proposed nonparametric estimator of different components under Scenarios (I) and (II). True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\eta_{A=0,T_2\leq t}$		$\eta_{A=1,T_2\leq t}$		$S_{A=0,T_1\leq T_2}(t)$		$S_{A=1,T_1\leq T_2}(t)$	
	$t=10$	$t=20$	$t=10$	$t=20$	$t=10$	$t=20$	$t=10$	$t=20$
Scenario (I)								
True	0.286	0.393	0.423	0.744	0.489	0.139	0.388	0.056
Low censoring (10%)								
Mean.EST	0.272	0.389	0.426	0.742	0.489	0.139	0.383	0.051
EMP.SD	0.028	0.021	0.030	0.018	0.023	0.016	0.017	0.008
EST.SE	0.027	0.021	0.031	0.018	0.024	0.017	0.018	0.008
CP95%	0.909	0.947	0.955	0.951	0.962	0.955	0.942	0.888
Moderate censoring (30%)								
Mean.EST	0.274	0.389	0.425	0.742	0.488	0.138	0.380	0.048
EMP.SD	0.029	0.022	0.032	0.019	0.027	0.021	0.020	0.010
EST.SE	0.029	0.022	0.033	0.019	0.028	0.021	0.021	0.010
CP95%	0.918	0.945	0.954	0.956	0.956	0.946	0.934	0.835
Scenario (II)								
True	0.004	0.038	0.017	0.149	0.780	0.318	0.780	0.318
Low censoring (10%)								
Mean.EST	0.004	0.041	0.019	0.153	0.780	0.316	0.776	0.306
EMP.SD	0.003	0.007	0.006	0.014	0.038	0.042	0.022	0.024
EST.SE	0.002	0.006	0.006	0.014	0.038	0.043	0.022	0.025
CP95%	0.859	0.948	0.955	0.945	0.936	0.948	0.947	0.916
Moderate censoring (30%)								
Mean.EST	0.004	0.041	0.019	0.153	0.780	0.316	0.776	0.306
EMP.SD	0.003	0.007	0.006	0.014	0.038	0.042	0.022	0.024
EST.SE	0.002	0.006	0.006	0.014	0.038	0.043	0.022	0.025
CP95%	0.859	0.948	0.955	0.945	0.936	0.948	0.947	0.916

Table D.5: Performance of the proposed nonparametric estimator of different components under Scenarios (I) and (II). True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\eta_{A=0,T_2\leq t}$		$\eta_{A=1,T_2\leq t}$		$S_{A=0,T_1\leq T_2}(t)$		$S_{A=1,T_1\leq T_2}(t)$	
	$t=10$	$t=20$	$t=10$	$t=20$	$t=10$	$t=20$	$t=10$	$t=20$
Scenario (III)								
True	0.277	0.589	0.686	0.891	0.229	0.034	0.229	0.034
Low censoring (10%)								
Mean.EST	0.291	0.595	0.685	0.889	0.225	0.031	0.224	0.031
EMP.SD	0.025	0.020	0.033	0.012	0.017	0.007	0.015	0.006
EST.SE	0.025	0.020	0.032	0.012	0.017	0.007	0.015	0.006
CP95%	0.926	0.934	0.939	0.953	0.928	0.878	0.909	0.873
Moderate censoring (30%)								
Mean.EST	0.290	0.594	0.685	0.889	0.221	0.028	0.220	0.028
EMP.SD	0.026	0.021	0.035	0.013	0.020	0.008	0.018	0.008
EST.SE	0.026	0.021	0.034	0.013	0.020	0.008	0.018	0.007
CP95%	0.934	0.936	0.942	0.935	0.922	0.838	0.910	0.814

Table D.6: Performance of the semi-parametric estimators under Scenario (IV) with $\rho = 0.5$, under low (5%, C-L) and moderate (25%, C-M) censoring rates for T_2. Censoring rates for T_1 were 35-40%. The presented effects are (2.4), (2.5) and (2.7) (ATE) and their median versions (MTE). True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$T_1\|ad$				$T_2\|ad$				$T_2\|nd$			
	ATE		MTE		ATE		MTE		ATE		MTE	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
$\theta = 2/3$												
True	-1.16		-0.67		-3.95		-7.17		-2.18		-1.44	
Mean.EST	-1.20	-1.21	-0.68	-0.68	-3.94	-3.95	-6.96	-6.93	-2.18	-2.18	-1.45	-1.45
EMP.SD	0.19	0.19	0.18	0.18	0.39	0.42	0.65	0.72	0.27	0.30	0.23	0.25
EST.SE	0.18	0.19	0.18	0.19	0.40	0.43	0.65	0.72	0.27	0.29	0.24	0.25
CP95%	0.94	0.94	0.94	0.95	0.95	0.95	0.96	0.93	0.95	0.93	0.95	0.94
$\theta = 1$												
True	-1.15		-0.71		-3.65		-6.83		-2.42		-1.58	
Mean.EST	-1.20	-1.20	-0.73	-0.73	-3.66	-3.68	-6.72	-6.69	-2.42	-2.44	-1.59	-1.61
EMP.SD	0.20	0.22	0.19	0.20	0.42	0.45	0.58	0.64	0.32	0.35	0.26	0.27
EST.SE	0.20	0.22	0.20	0.21	0.41	0.45	0.61	0.67	0.34	0.36	0.27	0.28
CP95%	0.96	0.93	0.96	0.95	0.93	0.95	0.98	0.97	0.96	0.95	0.96	0.95
$\theta = 2$												
True	-1.01		-0.78		-3.09		-5.60		-2.23		-2.71	
Mean.EST	-1.02	-1.04	-0.78	-0.80	-3.07	-3.06	-5.59	-5.55	-2.31	-2.30	-2.78	-2.74
EMP.SD	0.25	0.26	0.25	0.25	0.42	0.47	0.67	0.71	0.48	0.51	0.51	0.54
EST.SE	0.25	0.27	0.25	0.26	0.44	0.46	0.69	0.74	0.48	0.54	0.54	0.58
CP95%	0.95	0.96	0.94	0.95	0.95	0.94	0.95	0.95	0.95	0.96	0.96	0.96

Table D.7: Performance of the semi-parametric estimator for RMST causal effects under Scenario (IV) with $\rho = 0$, for different frailty variance (θ) values and under low (5%, C-L) and moderate (25%, C-M) censoring rates for T_2. Censoring rates for T_1 were 35-40%. The presented effects are (2.4), (2.5) and (2.7) (ATE) and their median versions (MTE). True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$T_1\|ad$				$T_2\|ad$				$T_2\|nd$			
	ATE		MTE		ATE		MTE		ATE		MTE	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
$\theta = 2/3$												
True	-1.15		-0.65		-3.95		-7.08		-2.28		-1.50	
Mean.EST	-1.19	-1.19	-0.67	-0.67	-3.94	-3.96	-6.92	-6.91	-2.24	-2.24	-1.49	-1.50
EMP.SD	0.18	0.20	0.17	0.18	0.39	0.43	0.68	0.72	0.30	0.31	0.25	0.26
EST.SE	0.18	0.19	0.18	0.19	0.40	0.44	0.68	0.74	0.30	0.32	0.25	0.26
CP95%	0.95	0.92	0.93	0.93	0.96	0.95	0.94	0.96	0.94	0.94	0.95	0.94
$\theta = 1$												
True	-1.11		-0.69		-3.63		-6.89		-2.58		-1.67	
Mean.EST	-1.16	-1.20	-0.70	-0.72	-3.63	-3.66	-6.72	-6.72	-2.56	-2.50	-1.69	-1.66
EMP.SD	0.20	0.21	0.19	0.20	0.41	0.45	0.56	0.66	0.39	0.42	0.29	0.31
EST.SE	0.21	0.21	0.20	0.20	0.42	0.45	0.63	0.69	0.38	0.41	0.29	0.30
CP95%	0.95	0.94	0.93	0.91	0.96	0.95	0.97	0.97	0.94	0.94	0.94	0.93
$\theta = 2$												
True	-0.90		-0.72		-3.01		-5.55		-2.48		-2.68	
Mean.EST	-0.91	-0.94	-0.72	-0.74	-2.98	-2.98	-5.51	-5.51	-2.53	-2.54	-2.73	-2.69
EMP.SD	0.24	0.26	0.25	0.26	0.44	0.48	0.67	0.74	0.62	0.65	0.53	0.53
EST.SE	0.25	0.26	0.25	0.25	0.44	0.47	0.71	0.75	0.59	0.66	0.54	0.58
CP95%	0.95	0.94	0.94	0.93	0.93	0.95	0.96	0.95	0.93	0.94	0.95	0.94

Table D.8: Performance of the semi-parametric estimator for RMST causal effects under Scenario (IV) with $\rho = 1$, for different frailty variance (θ) values and under low (5%, C-L) and moderate (25%, C-M) censoring rates for T_2. Censoring rates for T_1 were 35-40%. The presented effects are (2.4), (2.5) and (2.7) (ATE) and their median versions (MTE). True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$T_1\|ad$				$T_2\|ad$				$T_2\|nd$			
	ATE		MTE		ATE		MTE		ATE		MTE	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
$\theta = 2/3$												
True	-1.19		-0.68		-3.95		-7.27		-2.04		-1.38	
Mean.EST	-1.21	-1.23	-0.69	-0.70	-3.96	-3.93	-6.98	-6.90	-2.13	-2.14	-1.42	-1.41
EMP.SD	0.18	0.19	0.18	0.18	0.39	0.44	0.62	0.71	0.26	0.28	0.23	0.24
EST.SE	0.18	0.19	0.18	0.19	0.39	0.43	0.63	0.71	0.25	0.27	0.23	0.24
CP95%	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.94	0.93	0.94	0.93	0.95
$\theta = 1$												
True	-1.23		-0.75		-3.68		-6.76		-2.14		-1.49	
Mean.EST	-1.27	-1.27	-0.78	-0.77	-3.65	-3.67	-6.67	-6.66	-2.25	-2.30	-1.53	-1.55
EMP.SD	0.21	0.22	0.20	0.21	0.40	0.45	0.54	0.62	0.31	0.33	0.26	0.27
EST.SE	0.21	0.22	0.20	0.21	0.40	0.44	0.60	0.66	0.30	0.33	0.26	0.27
CP95%	0.94	0.94	0.95	0.94	0.95	0.94	0.96	0.97	0.93	0.92	0.95	0.94
$\theta = 2$												
True	-1.21		-0.87		-3.18		-5.53		-1.73		-3.43	
Mean.EST	-1.24	-1.22	-0.89	-0.88	-3.18	-3.16	-5.54	-5.52	-1.78	-1.82	-3.54	-3.55
EMP.SD	0.28	0.29	0.26	0.25	0.42	0.47	0.67	0.74	0.31	0.36	0.95	1.03
EST.SE	0.28	0.30	0.26	0.27	0.43	0.46	0.70	0.75	0.34	0.43	1.06	1.13
CP95%	0.95	0.95	0.96	0.96	0.96	0.96	0.95	0.95	0.96	0.96	0.96	0.96

Table D.9: Performance of the semi-parametric estimator for $\boldsymbol{\beta}^0$ under Scenario (IV) with $\rho = 0.5$, for different frailty variance (θ) values and under low (5%, C-L) and moderate (25%, C-M) censoring rates for T_2. Censoring rates for T_1 were 35-40%. True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\beta^0_{01,1}$		$\beta^0_{01,2}$		$\beta^0_{02,1}$		$\beta^0_{02,2}$		$\beta^0_{12,1}$		$\beta^0_{12,2}$	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
True	-1.39		1.10		-0.29		0.41		0.00		0.00	
$\theta = 2/3$												
Mean.EST	-1.38	-1.38	1.09	1.09	-0.27	-0.28	0.40	0.40	0.00	0.00	0.00	0.00
EMP.SD	0.13	0.13	0.07	0.07	0.12	0.13	0.06	0.07	0.13	0.15	0.07	0.08
EST.SE	0.13	0.13	0.07	0.08	0.12	0.13	0.07	0.07	0.13	0.15	0.07	0.08
CP95%	0.94	0.94	0.96	0.96	0.94	0.94	0.96	0.95	0.94	0.94	0.96	0.93
$\theta = 1$												
Mean.EST	-1.38	-1.38	1.09	1.09	-0.28	-0.27	0.40	0.39	0.00	0.00	0.00	0.00
EMP.SD	0.14	0.14	0.07	0.08	0.13	0.13	0.07	0.07	0.15	0.16	0.07	0.08
EST.SE	0.14	0.14	0.08	0.08	0.13	0.14	0.07	0.07	0.14	0.17	0.07	0.08
CP95%	0.94	0.96	0.96	0.95	0.94	0.95	0.94	0.96	0.94	0.94	0.95	0.96
$\theta = 2$												
Mean.EST	-1.38	-1.38	1.09	1.09	-0.28	-0.28	0.40	0.39	-0.02	-0.01	0.01	0.00
EMP.SD	0.16	0.16	0.09	0.09	0.15	0.15	0.08	0.08	0.17	0.18	0.09	0.10
EST.SE	0.16	0.17	0.09	0.09	0.15	0.16	0.08	0.08	0.17	0.19	0.09	0.09
CP95%	0.94	0.96	0.94	0.94	0.95	0.95	0.94	0.94	0.94	0.96	0.94	0.94

Table D.10: Performance of the semi-parametric estimator for $\boldsymbol{\beta}^1$ under Scenario (IV) with $\rho = 0.5$, for different frailty variance (θ) values and under low (5%, C-L) and moderate (25%, C-M) censoring rates for T_2. Censoring rates for T_1 were 35-40%. True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\beta^1_{01,1}$		$\beta^1_{01,2}$		$\beta^1_{02,1}$		$\beta^1_{02,2}$		$\beta^1_{12,1}$		$\beta^1_{12,2}$	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
True	0.00		0.69		-0.29		0.41		-0.69		0.69	
$\theta = 2/3$												
Mean.EST	0.00	0.00	0.69	0.69	-0.28	-0.29	0.41	0.40	-0.68	-0.69	0.69	0.69
EMP.SD	0.10	0.10	0.06	0.06	0.14	0.15	0.08	0.08	0.11	0.13	0.06	0.07
EST.SE	0.10	0.11	0.06	0.06	0.14	0.15	0.08	0.08	0.11	0.12	0.06	0.07
CP95%	0.95	0.95	0.96	0.94	0.95	0.94	0.95	0.94	0.94	0.95	0.94	0.96
$\theta = 1$												
Mean.EST	0.00	0.00	0.69	0.69	-0.29	-0.29	0.40	0.40	-0.69	-0.69	0.69	0.69
EMP.SD	0.11	0.12	0.06	0.06	0.15	0.15	0.08	0.08	0.12	0.13	0.07	0.07
EST.SE	0.11	0.12	0.06	0.06	0.15	0.15	0.08	0.08	0.12	0.13	0.07	0.07
CP95%	0.94	0.94	0.96	0.94	0.95	0.96	0.94	0.95	0.95	0.96	0.95	0.95
$\theta = 2$												
Mean.EST	0.00	0.00	0.69	0.69	-0.29	-0.29	0.40	0.40	-0.69	-0.68	0.69	0.69
EMP.SD	0.14	0.14	0.07	0.07	0.17	0.17	0.09	0.09	0.14	0.16	0.07	0.08
EST.SE	0.14	0.14	0.07	0.08	0.17	0.18	0.09	0.09	0.15	0.16	0.08	0.08
CP95%	0.96	0.95	0.95	0.95	0.96	0.96	0.95	0.95	0.96	0.95	0.95	0.95

Table D.11: Performance of the semi-parametric estimator for $\Lambda_{01}^{0,a}(t)$ under Scenario (IV) with $\rho = 0.5$, for different frailty variance (θ) values and under low (5%, C-L) and moderate (25%, C-M) censoring rates for T_2. Censoring rates for T_1 were 35-40%. True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\Lambda_{01}^{0,0}(2)$		$\Lambda_{01}^{0,0}(3)$		$\Lambda_{01}^{0,0}(4)$		$\Lambda_{01}^{0,1}(2)$		$\Lambda_{01}^{0,1}(3)$		$\Lambda_{01}^{0,1}(4)$	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
True	0.18		0.49		1.00		1.00		2.76		5.66	
$\theta = 2/3$												
Mean.EST	0.17	0.18	0.48	0.48	0.98	0.98	0.99	0.99	2.72	2.73	5.55	5.59
EMP.SD	0.02	0.02	0.05	0.05	0.11	0.11	0.09	0.09	0.31	0.33	0.77	0.86
EST.SE	0.02	0.02	0.05	0.05	0.11	0.12	0.09	0.10	0.31	0.34	0.81	0.91
CP95%	0.93	0.95	0.93	0.94	0.94	0.93	0.94	0.94	0.93	0.93	0.94	0.94
$\theta = 1$												
Mean.EST	0.18	0.17	0.48	0.48	0.98	0.98	0.99	0.99	2.72	2.71	5.59	5.54
EMP.SD	0.02	0.02	0.05	0.05	0.11	0.12	0.10	0.11	0.33	0.35	0.82	0.88
EST.SE	0.02	0.02	0.05	0.06	0.12	0.13	0.10	0.10	0.33	0.36	0.83	0.92
CP95%	0.93	0.94	0.94	0.94	0.94	0.93	0.94	0.93	0.94	0.92	0.92	0.92
$\theta = 2$												
Mean.EST	0.18	0.18	0.48	0.48	0.98	0.98	0.99	0.99	2.72	2.71	5.58	5.54
EMP.SD	0.02	0.03	0.06	0.07	0.13	0.15	0.12	0.12	0.38	0.39	0.88	0.92
EST.SE	0.02	0.03	0.06	0.07	0.14	0.15	0.12	0.13	0.39	0.41	0.91	0.99
CP95%	0.96	0.93	0.94	0.94	0.93	0.93	0.93	0.94	0.94	0.93	0.94	0.93

Table D.12: Performance of the semi-parametric estimator for $\Lambda_{02}^{0,a}(t)$ under Scenario (IV) with $\rho = 0.5$, for different frailty variance (θ) values and under low (5%, C-L) and moderate (25%, C-M) censoring rates for T_2. Censoring rates for T_1 were 35-40%. True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\Lambda_{02}^{0,0}(2)$		$\Lambda_{02}^{0,0}(3)$		$\Lambda_{02}^{0,0}(4)$		$\Lambda_{02}^{0,1}(2)$		$\Lambda_{02}^{0,1}(3)$		$\Lambda_{02}^{0,1}(4)$	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
True	0.16		0.36		0.64		0.44		1.00		1.78	
$\theta = 2/3$												
Mean.EST	0.16	0.16	0.35	0.36	0.63	0.63	0.44	0.44	0.99	0.99	1.74	1.76
EMP.SD	0.02	0.02	0.04	0.04	0.07	0.07	0.05	0.05	0.13	0.14	0.29	0.33
EST.SE	0.95	0.94	0.94	0.94	0.92	0.95	0.93	0.94	0.94	0.93	0.92	0.92
CP95%	0.95	0.95	0.96	0.94	0.95	0.94	0.95	0.94	0.94	0.95	0.94	0.96
$\theta = 1$												
Mean.EST	0.16	0.16	0.35	0.35	0.63	0.62	0.44	0.44	0.99	0.99	1.78	1.74
EMP.SD	0.02	0.02	0.04	0.04	0.07	0.08	0.05	0.06	0.14	0.15	0.31	0.32
EST.SE	0.02	0.02	0.04	0.04	0.07	0.08	0.06	0.06	0.15	0.15	0.32	0.33
CP95%	0.94	0.94	0.94	0.94	0.93	0.93	0.94	0.94	0.95	0.94	0.94	0.92
$\theta = 2$												
Mean.EST	0.16	0.16	0.35	0.35	0.63	0.63	0.44	0.44	0.99	0.98	1.75	1.74
EMP.SD	0.02	0.02	0.05	0.05	0.09	0.09	0.06	0.06	0.16	0.16	0.33	0.33
EST.SE	0.02	0.02	0.05	0.05	0.09	0.09	0.06	0.06	0.16	0.16	0.33	0.33
CP95%	0.93	0.95	0.94	0.94	0.93	0.93	0.94	0.94	0.93	0.93	0.92	0.93

Table D.13: Performance of the semi-parametric estimator for $\Lambda_{12}^{0;a}(t)$ under Scenario (IV) with $\rho = 0.5$, for different frailty variance (θ) values and under low (5%, C-L) and moderate (25%, C-M) censoring rates for T_2. Censoring rates for T_1 were 35-40%. True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\Lambda_{12}^{0,0}(8)$		$\Lambda_{02}^{0,0}(12)$		$\Lambda_{02}^{0,0}(16)$		$\Lambda_{12}^{0,1}(8)$		$\Lambda_{02}^{0,1}(12)$		$\Lambda_{02}^{0,1}(16)$	
	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M	C-L	C-M
True	0.21		0.57		1.18		0.57		1.58		3.24	
$\theta = 2/3$												
Mean.EST	0.21	0.21	0.57	0.57	1.16	1.16	0.57	0.57	1.56	1.57	3.19	3.21
EMP.SD	0.03	0.03	0.06	0.06	0.12	0.13	0.06	0.06	0.15	0.17	0.34	0.40
EST.SE	0.03	0.03	0.06	0.07	0.12	0.14	0.05	0.06	0.15	0.17	0.35	0.41
CP95%	0.94	0.95	0.95	0.95	0.94	0.93	0.93	0.94	0.94	0.94	0.94	0.94
$\theta = 1$												
Mean.EST	0.21	0.21	0.57	0.56	1.16	1.15	0.57	0.57	1.57	1.56	3.21	3.20
EMP.SD	0.03	0.03	0.06	0.07	0.13	0.14	0.06	0.06	0.16	0.18	0.38	0.41
EST.SE	0.03	0.03	0.07	0.07	0.13	0.15	0.06	0.06	0.17	0.18	0.39	0.43
CP95%	0.94	0.94	0.93	0.93	0.93	0.94	0.94	0.95	0.94	0.95	0.94	0.94
$\theta = 2$												
Mean.EST	0.21	0.21	0.57	0.57	1.16	1.17	0.57	0.56	1.57	1.55	3.20	3.18
EMP.SD	0.03	0.04	0.08	0.08	0.16	0.17	0.07	0.07	0.19	0.21	0.42	0.47
EST.SE	0.03	0.04	0.08	0.08	0.16	0.18	0.07	0.07	0.20	0.22	0.45	0.49
CP95%	0.94	0.94	0.93	0.94	0.93	0.94	0.94	0.95	0.95	0.94	0.95	0.94

Table D.14: Performance of the semi-parametric estimator for the frailty variance θ under Scenario (IV) for different ρ and θ values, and under low (5%, C-L) and moderate (25%, C-M) censoring rates of T_2. The presented effects are (2.4), (2.5) and (2.7) (ATE) and their median versions (MTE). True: True parameter values; Mean.EST: mean estimate; EMP.SD: empirical standard deviation of the estimates; EST.SE: mean estimated standard error; CP95%: empirical coverage rate of 95% confidence interval.

	$\rho = 0$		$\rho = 0.5$		$\rho = 1$	
	C-L	C-M	C-L	C-M	C-L	C-M
$\theta = 2/3$						
Mean.EST	0.64	0.64	0.64	0.64	0.64	0.64
EMP.SD	0.06	0.07	0.06	0.06	0.05	0.06
EST.SE	0.06	0.06	0.06	0.06	0.06	0.06
CP95%	0.91	0.90	0.91	0.93	0.92	0.91
$\theta = 1$						
Mean.EST	0.97	0.96	0.97	0.96	0.97	0.96
EMP.SD	0.07	0.08	0.07	0.08	0.07	0.08
EST.SE	0.07	0.08	0.07	0.08	0.07	0.08
CP95%	0.90	0.92	0.92	0.91	0.92	0.92
$\theta = 2$						
Mean.EST	1.95	1.93	1.95	1.94	1.95	1.94
EMP.SD	0.11	0.12	0.11	0.12	0.11	0.12
EST.SE	0.11	0.12	0.11	0.12	0.11	0.12
CP95%	0.91	0.91	0.91	0.92	0.92	0.91

Table E.15: Estimated $\exp(\boldsymbol{\beta})$ and associated confidence intervals in the ACT dataset.

	$a = 0$	$a = 1$
Healthy $\rightarrow$ AD		
White race	1.01 (0.77, 1.32)	0.93 (0.76, 1.13)
Female	0.93 (0.72, 1.2)	0.99 (0.82, 1.2)
Healthy $\rightarrow$ Death		
White race	1.08 (0.83, 1.39)	0.72 (0.58, 0.88)
Gender	0.63 (0.55, 0.73)	0.67 (0.6, 0.75)
AD $\rightarrow$ Death		
White race	1.32 (1.13, 1.55)	1.36 (1.21, 1.52)
Female	0.81 (0.69, 0.94)	0.75 (0.68, 0.83)

References

Akritas, M. G. and I. Van Keilegom (2003). Estimation of bivariate and marginal distributions with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 457–471.

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical models based on counting processes*. New York: Springer-Verlag.

Beran, R. (1981). Nonparametric regression with randomly censored survival data.

Gerds, T. A. (2019). *prodlim: Product-Limit Estimation for Censored Event History Analysis*. R package version 2019.11.13.

Kosorok, M. R. (2007). *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media.

Long, D. M. and M. G. Hudgens (2013). Sharpening bounds on principal effects with covariates. *Biometrics 69*(4), 812–819.

Shepherd, B., P. Gilbert, and T. Lumley (2007). Sensitivity analyses comparing time-to-event outcomes existing only in a subset selected postrandomization. *Journal of the American Statistical Association 102*(478), 573–582.