Supporting Information for Modeling semi-competing risks data as a longitudinal bivariate process

Daniel Nevo, Deborah Blacker, Eric B. Larson and Sebastien Haneuse

Description of contents in the Supplementary Materials

- Section A.1 provides a brief review of existing approaches to the analysis of semi-competing risks data, with a focus on how dependence between T_1 and T_2 is structured and how the nature of the assumed structure contributes to gaining knowledge.
- Section A.2 provides technical details complementary to the main text. It specifically shows how a bivariate binary distribution is derived from the marginal probabilities and the odds ratio (Section A.2.1), that any choice for $\boldsymbol{\tau}$ gives rise to well-defined $\pi_{1i,k}, \pi_{2i,k}$ and $\pi_{12i,k}$ (Section A.2.2), an expanded version of Figure 1 (Section A.2.3), how likelihood contributions for let-truncated data are calculated (Section A.2.4), and, finally, how the difference operation Δ_m is represented in the matrix form D_m (Section A.2.5).
- Section A.3 gives the details of the asymptotic distribution and variance of our proposed estimator without (Section A.3.1) and with (Section A.3.2) penalty applied to the likelihood approach.
- Section A.4 presents details about and results from our extensive simulations. It describes the data-generating mechanism, details about the different analyses we considered and then it presents the results along with conclusions. Section A.4.4 presents additional simulation study, tailored to study the impact of censoring level on the different approaches to determine k_i^l and k_i^r.
- Section A.5 presents two tables showing the outcomes represented using our partition approach, under 2.5 years and 5 years interval length partitions. It also includes a summary table of key participant characteristics measured at study entry.
- Section A.6 presents further details and results from the ACT data analysis.

A.1 Existing methods

Beyond the usual challenges of time-to-event analyses (i.e. structuring covariate effects, handling functions of time, and accommodating various forms of censoring and truncation), key challenges that arise in the analysis of semi-competing risks data are: (i) respecting the terminal event as a competing risk; and, (ii) structuring dependence between T_1 and T_2. Here we present an overview of methods that explicitly structure the dependence between T_1 and T_2, and how the nature of the assumed structure contributes to gaining knowledge.

A.1.1 The causal inference paradigm

When primary scientific interest lies in comparing the relative impact of two (or more) treatment options on the risk of the non-terminal event, researchers may embed the analysis within the causal counterfactual paradigm (Zhang and Rubin, 2003; Egleston et al., 2006; Tchetgen Tchetgen, 2014). Throughout this work, the overarching strategy has been to first define some causal contrast, then specify the assumptions under which the contrast is identified with the observed data, and finally propose consistent estimators. One such contrast is the so-called *survivor average causal effect* which considers the impact of treatment choice among 'always survivors', that is the sub-population of patients who would survive under either treatment and, as such, for whom the non-terminal event is always well-defined. Note that, in adopting a strategy of restricting to 'always survivors' to handle death as a competing risk, analysts do not need to consider the nature of the dependence between the two events. While this has appeal, it also precludes learning about dependence.

A.1.2 Copula-based methods

In standard bivariate survival analyses (i.e. where neither event is terminal for the other), one can model the joint survivor as $\Pr(T_1 > t, T_2 > s) = C_\theta(S_1(t), S_2(s))$, where $S_1(\cdot)$ and $S_2(\cdot)$ are the marginal survivor functions for T_1 and T_2, respectively, and $C_\theta(\cdot, \cdot)$ is a copula indexed by the (unknown) parameter θ. One well-known example of the latter is the Clayton copula (Clayton, 1978), one representation of which is:

$$C_\theta(S_1(t), S_2(s)) = \{S_1(t)^{(1-\theta)} + S_2(t)^{(1-\theta)} - 1\}^{1/(1-\theta)},$$

with θ interpretable as the variance of a common frailty for which the population distribution is a Gamma(θ^{-1}, θ^{-1}). In the semi-competing risks setting, however, since the terminal event is a competing risk for the non-terminal event, mass can only be attributed to joint events in the upper wedge of the (T_1, T_2) plane. To accommodate this, Fine et al. (2001) proposed to use the Clayton copula to define a model solely on the upper wedge and, thus, leave the joint density on the lower wedge as 'unspecified'. While numerous extensions have been proposed (Peng and Fine, 2007; Hsieh et al., 2008; Li and Peng, 2015), one limitation of copula-based methods is that $S_1(t)$ in $C_\theta(S_1(t), S_2(s))$ is not the same as the (unidentifiable) marginal distribution of T_1, even though it is sometimes interpreted as such. Furthermore, directly relevant to this paper is that, as mathematical devices, copulas can only encompasses a narrow range of dependence structures. The Clayton copula, for example, only permits positive dependence between T_1 and T_2. Consequently, copula-based strategies may be limited if interests lies (in part, at least) on how T_1 and T_2 covary.

A.1.3 Illness-death models

A second general framework views semi-competing risks data as arising from an *illness-death* model (Xu et al., 2010). Briefly, such models posit that at any given point in time the study unit is in one of three 'states': (I) an initial state, prior to experiencing either event; (II) a state of having experienced the non-terminal event, prior to experiencing the terminal event;

and (III) an absorbing state of having experienced the terminal event. Modeling then proceeds by structuring the three intensity or hazard functions that dictate the rates at which study units transition between the states:

$\lambda_1(t_1)$ for the (I) $\Rightarrow$ (II) transition;

$\lambda_2(t_2)$ for the (I) $\Rightarrow$ (III) transition; and,

$\lambda_3(t_2|t_1)$ for the (II) $\Rightarrow$ (III) transition.

To investigate covariate effects, one can fit Cox models (Xu et al., 2010; Lee et al., 2015) or accelerated failure time models (Lee et al., 2017) for the three hazard functions. For a given specification, one can then consider the interplay between $\lambda_2(t_2)$ and $\lambda_3(t_2|t_1)$, sometimes via the *explanatory hazard ratio*:

$$\text{EHR}(t_2;\, t_1)=\lambda_2(t_2)/\lambda_3(t_2|t_1),$$

as a means to characterize dependence between T_1 and T_2. Additionally, a patient-specific frailty or random effect, usually taken to arise from some common distribution with mean 0.0 (or 1.0, as appropriate) and a constant variance, say θ, may be included to accommodate residual dependence not structured through the hazard functions. While a broad class of dependence structures can be captured by these two components of dependence (i.e. the $\text{EHR}(t_2;\, t_1)$ and the frailty variance, θ), the main flexibility of this approach arises from (implicitly) viewing the terminal event as being of primary interest, while a direct representation of how the two events covary is solely through the common frailty.

A.1.4 The cross-quantile residual ratio

Towards directly quantifying dependence between T_1 and T_2, Yang and Peng (2016) proposed the *cross-quantile residual ratio*. Briefly, let $Q_q(Y|\ A) = \inf\{t : \Pr(Y \leq t|\ A) \geq q\}$ denote the q^{th} quantile of Y given condition A holds. The quantile residual time for the terminal event at a given time point t_0 is defined as $Q_q(T_2 - t_0|\ T_2 > t_0)$. The cross-quantile residual ratio is then defined as:

$$\text{CQRR}(q; t_0) \ = \ \frac{Q_q(T_2 - t_0|\ T_2 > t_0, T_1 > t_0)}{Q_q(T_2 - t_0|\ T_2 > t_0, T_1 \leq t_0)}.$$

To interpret this quantity, consider q=0.5 and, in the context of the ACT study, set t_0=75. A value of $\text{CQRR}(q; t_0)$=1.5 can be interpreted as indicating that the median residual lifetime beyond age 75 for patients who have not received diagnosis of AD by age 75 is 50% longer than the corresponding median residual lifetime beyond age 75 for those patients who have had a diagnosis prior to age 75. Intuitively, having not experienced the non-terminal event is associated with slower progression to the terminal event. One key appealing feature of this metric is that it focuses attention directly on the time scale rather than on a comparison of hazards. Like the $\text{EHR}(t_2;\, t_1)$, however, implicit to the use of $\text{CQRR}(q; t_0)$ is that the terminal event is of primary interest.

A.2 Additional technical details

A.2.1 The bivariate binary distribution

In this section, we show the distribution of any bivariate bivariate random variable (Y_1, Y_2) is fully specified by the marginal probabilities of Y_1 and Y_2 and the odds ratio. Our approach utilizes this fact to model time-trends and relationship to covariates for the event times. In our case, the representation is for the bivariate $Y_{1i,k}, Y_{2i,k}$ conditioned upon no event of any type by the beginning of the k-th interval, namely $(Y_{1i,k-1} = 0, Y_{2i,k-1} = 0)$.

Let $q_{rs} = Pr(Y_1 = r, Y_2 = s)$ for $r, s = 0, 1$. Let also $p_1 = Pr(Y_1 = 1), p_2 = Pr(Y_2 = 1)$ and $\theta = OR(Y_1, Y_2) = \frac{q_{01}q_{10}}{q_{00}q_{11}}$. The values of $q_{01}, q_{10}q_{00}, q_{11}$ are obtained by solving

$$\begin{aligned} q_{10} + q_{11} &= p_1 \\ q_{01} + q_{11} &= p_2 \\ \frac{q_{01}q_{10}}{q_{00}q_{11}} &= \theta \\ q_{00} + q_{10} + q_{01} + q_{11} &= 1 \end{aligned} \tag{A.1}$$

By substituting $q_{00} = 1 - q_{10} - q_{01} - q_{11}$, $q_{10} = p_1 - q_{11}$ and $q_{01} = p_2 - q_{11}$ into (A.1), a quadratic function in q_{11} is obtained. It can be verified that the unique solution for q_{11} and the rest of the cell probabilities are thus given by

$$q_{11} = \begin{cases} p_1 p_2 & \theta = 1 \\ \frac{1}{2(\theta-1)} \times \left[1 + (p_1 + p_2)(\theta - 1) - \sqrt{(1 + (p_1 + p_2)(\theta - 1))^2 - 4\theta(\theta - 1)p_1 p_2}\right] & \theta \neq 1 \end{cases}$$

$$q_{10} = p_1 - q_{11}$$

$$q_{01} = p_2 - q_{11}$$

$$q_{00} = 1 - q_{01} - q_{10} + q_{11}$$

A.2.2 A semi-competing risks data-generating mechanism induces a bivariate binary longitudinal data-generating mechanism

We show here that given a distribution for (T_{1i}, T_{2i}) given $\boldsymbol{X}_i$ on the identifiable region $T_{1i} < T_{2i}$, for any choice of the interval partition $\boldsymbol{\tau}$, the quantities $\pi_{1i,k}, \pi_{2i,k}$ and $\theta_{i,k}$, for $k = 1, ..., K$, are well-defined and induced by the distribution of (T_{1i}, T_{2i}) given $\boldsymbol{X}_i$. This implies that there is no such thing as the "right partition". For a given partition $\boldsymbol{\tau} = \{\tau_0, \ldots, \tau_K\}$, we may write

$$\begin{aligned} \pi_{1i,k} &= \mathrm{P}(Y_{1i,k} = 1 |\ Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \boldsymbol{X}^H_{i,k}) \\ &= \mathrm{P}(\tau_{k-1} < T_{1i} \leq \tau_k |\ T_{1i} > \tau_{k-1},\ T_{2i} > \tau_{k-1}, \boldsymbol{X}^H_{i,k}) \\ \pi_{2i,k}(0) &= \mathrm{P}(Y_{2i,k} = 1 |\ Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \boldsymbol{X}^H_{i,k}) \\ &= \mathrm{P}(\tau_{k-1} < T_{2i} \leq \tau_k |\ T_{1i} > \tau_{k-1},\ T_{2i} > \tau_{k-1}, \boldsymbol{X}^H_{i,k}) \\ \pi_{2i,k}(1) &= \mathrm{P}(Y_{2i,k} = 1 |\ Y_{1i,k-1} = 1, Y_{2i,k-1} = 0, \boldsymbol{X}^H_{i,k}) \\ &= \mathrm{P}(\tau_{k-1} < T_{2i} \leq \tau_k |\ T_{1i} \leq \tau_{k-1},\ T_{2i} > \tau_{k-1}, \boldsymbol{X}^H_{i,k}) \end{aligned}$$

For $\theta_{i,k}$ we should show that

$$\mathrm{P}(Y_{1i,k} = y_1, Y_{2i,k} = y_2 |\ Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \boldsymbol{X}^H_{i,k})$$

can be written in terms of the joint distribution of (T_{1i}, T_{2i}) given $\boldsymbol{X}_{i,k}^H$ for $(y_1, y_2) = \{0,1\} \times \{0,1\}$. This is indeed the case:

$$\begin{aligned}
&\mathrm{P}(Y_{1i,k} = 0, Y_{2i,k} = 0 |\ Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \boldsymbol{X}_{i,k}^H) \\
&\quad = \mathrm{P}(T_{1i} > \tau_k,\ T_{2i} > \tau_k |\ T_{1i} > \tau_{k-1},\ T_{2i} > \tau_{k-1}, \boldsymbol{X}_{i,k}^H) \\
&\mathrm{P}(Y_{1i,k} = 1, Y_{2i,k} = 0 |\ Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \boldsymbol{X}_{i,k}^H) \\
&\quad = \mathrm{P}(\tau_{k-1} < T_{1i} \leq \tau_k,\ T_{2i} > \tau_k |\ T_{1i} > \tau_{k-1},\ T_{2i} > \tau_{k-1}, \boldsymbol{X}_{i,k}^H) \\
&\mathrm{P}(Y_{1i,k} = 0, Y_{2i,k} = 1 |\ Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \boldsymbol{X}_{i,k}^H) \\
&\quad = \mathrm{P}(T_{1i} > \tau_k,\ \tau_{k-1} < T_{2i} \leq \tau_k |\ T_{1i} > \tau_{k-1},\ T_{2i} > \tau_{k-1}, \boldsymbol{X}_{i,k}^H) \\
&\mathrm{P}(Y_{1i,k} = 1, Y_{2i,k} = 1 |\ Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \boldsymbol{X}_{i,k}^H) \\
&\quad = \mathrm{P}(\tau_{k-1} < T_{1i} \leq \tau_k,\ \tau_{k-1} < T_{2i} \leq \tau_k |\ T_{1i} > \tau_{k-1},\ T_{2i} > \tau_{k-1}, \boldsymbol{X}_{i,k}^H)
\end{aligned}$$

and each of the joint probabilities can be calculated.

A.2.3 Expanded version of Figure 1

Figure A.1 is an extended version of Figure 1 of the main text, now with five hypothetical participants. It demonstrates both how standard semi-competing risk data is translated to the longitudinal bivariate process. At the same time, it also illustrates the obtained outcome data under the three different approaches towards handling right-censoring and left-truncation.

A.2.4 Likelihood contribution for left-truncated data

Given that k_i^l was chosen, the derivation of the likelihood for left truncated data is as follows. In fact, we use the usual form of conditional probability, conditional on no events observed by the interval $(k_i^l - 1, k_i^l)$. By the way k_i^l was constructed (up to the approximations detailed in Section 3.2 of the main text), we have that for all $k = 1, ...k_i^l$, we have $Y_{1i,k} = 0, Y_{1i,k} = 0$, under the assumption that, like in the ACT study, there are no prevalent cases. Let

$$\mathcal{Q}_{i,k} = \{Y_{1i,1} = 0, ..., Y_{1i,k} = 0, Y_{2i,1} = 0, ..., Y_{2i,k} = 0\}$$

be the event that none of the events were observed by τ_k. The likelihood contribution for an observation with left truncation so no events by the interval $(k_i^l, k_i^l + 1)$ is the conditional probability of the observed data, conditional on no events by k_i^l. That is,

$$\begin{aligned}
&\mathrm{P}(\boldsymbol{Y}_i = \boldsymbol{y}_i | \mathcal{Q}_{i,k_i^l},\ \boldsymbol{X}_i^H) \\
&= \prod_{k=k_i^l}^{k_i^r} \mathrm{P}(Y_{1i,k} = y_{1i,k}, Y_{2i,k} = y_{2i,k} |\ \mathcal{Q}_{i,k_i^l}, Y_{1i,k_i^l+1} = y_{1i,k_i^l+1}, ..., Y_{1i,k} = y_{1i,k-1}, \\
&\qquad\qquad Y_{2i,k_i^l+1} = y_{2i,k_i^l+1}, ..., Y_{2i,k} = y_{2i,k}, \boldsymbol{X}_{i,k}^H) \\
&= \prod_{k=1}^{k_i^r} \mathrm{P}(Y_{1i,k} = y_{1i,k}, Y_{2i,k} = y_{2i,k-1} |\ Y_{1i,k-1} = y_{1i,k-1}, Y_{2i,k-1} = y_{2i,k-1}, \boldsymbol{X}_{i,k}^H).
\end{aligned}$$

In the second line, we decompose the joint distribution to conditional distributions and exclude from the event $\boldsymbol{Y}_i = \boldsymbol{y}_i$ expressions that are constant conditional on $\mathcal{Q}_{i,k}$. In the third line we use the Markov assumption.

A.2.5 The difference operator $\boldsymbol{D}_m$

We show here that the penalty term of the B-spline coefficient differences, $\sum_{j=m+1}^{\widetilde{J}_1} (\Delta^m \eta_{1,j})^2$ can be written as $\boldsymbol{\eta}_1^T \boldsymbol{D}_m^T \boldsymbol{D}_m \boldsymbol{\eta}_1$ (and similarly for $\boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_\theta$), where $\boldsymbol{D}_m$ is the following $(\widetilde{J}_1 - m) \times \widetilde{J}_1$

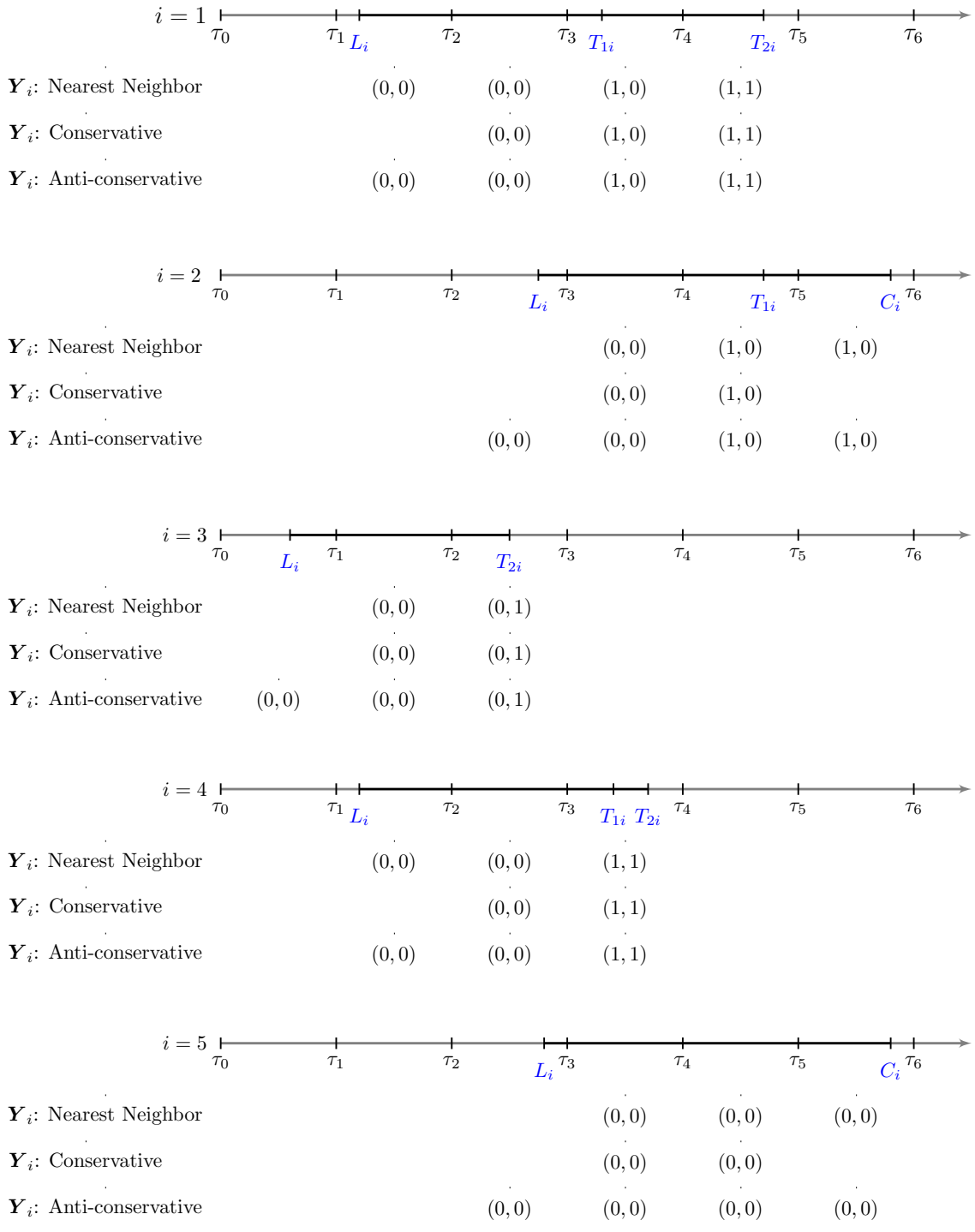


Figure A.1: An extended version of Figure 1. Graphical representation of the interplay between standard notation for semi-competing risks outcome data and the proposed bivariate longitudinal framework.

matrix representation of Δ_m. For $m = 2$, it is

$$\boldsymbol{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

and it can verified that $\boldsymbol{\eta}_1^T \boldsymbol{D}_m^T \boldsymbol{D}_m \boldsymbol{\eta}_1 = \sum_{j=m+1}^{\widetilde{J}_1} (\Delta^m \eta_{1,j})^2$. For example, for $\widetilde{J}_1 = 4$ we have

$$\boldsymbol{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix}, \qquad \boldsymbol{D}_2^T \boldsymbol{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 \\ -2 & 5 & -4 & 0 \\ 0 & -4 & 5 & -2 \\ 0 & 1 & -2 & 1 \end{pmatrix},$$

and therefore

$$\begin{aligned} \boldsymbol{\eta}_1^T \boldsymbol{D}_2^T \boldsymbol{D}_2 \boldsymbol{\eta}_1 &= (\eta_{1,1})^2 + 5(\eta_{1,2})^2 + 5(\eta_{1,3})^2 + (\eta_{1,4})^2 + \eta_{1,1}\eta_{1,3} + \eta_{1,2}\eta_{1,3} - 4\eta_{1,1}\eta_{1,2} - 4\eta_{1,2}\eta_{1,3} - 4\eta_{1,3}\eta_{1,4} \\ &= (\eta_{1,3} - 2\eta_{1,2} + \eta_{1,1})^2 + (\eta_{1,4} - 2\eta_{1,3} + \eta_{1,2})^2 \\ &= \Delta^2 \eta_{1,3} + \Delta^2 \eta_{1,4} = \sum_{j=m+1}^{\widetilde{J}_1} (\Delta^m \eta_{1,j})^2. \end{aligned}$$

A.3 Asymptotic Theory

A.3.1 Theory for the non-penalized estimator

In this subsection we consider the asymptotic theory for the non-penalized estimator $\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}} = \widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}}_{\boldsymbol{\lambda}=0}$ obtained by maximizing

$$\begin{aligned}\mathcal{L}(\boldsymbol{\phi}^{\boldsymbol{\eta}}) \;=\; \prod_{i=1}^{N}\prod_{k=k_i^l}^{k_i^r} \Big\{ &[\pi_{12i,k}]^{Y_{1i,k}Y_{2i,k}}[\pi_{1i,k}-\pi_{12i,k}]^{Y_{1i,k}(1-Y_{2i,k})}[\pi_{2i,k}(0)-\pi_{12i,k}]^{Y_{1i,k}Y_{2i,k}} \\ &\times\; [1-\pi_{1i,k}-\pi_{2i,k}(0)+\pi_{12i,k}]^{(1-Y_{1i,k})(1-Y_{2i,k})}\Big\}^{(1-Y_{1i,k-1})(1-Y_{2i,k-1})} \\ &\times \Big\{[\pi_{2i,k}(1)]^{Y_{2i,k}}[1-\pi_{2i,k}(1)]^{1-Y_{2i,k}}\Big\}^{Y_{1i,k-1}(1-Y_{2i,k-1})}.\end{aligned}$$

As a result, $\widehat{\boldsymbol{\phi}^{\boldsymbol{\eta}}}$ solves the equation $\boldsymbol{U}(\boldsymbol{\phi}^{\boldsymbol{\eta}};0)=\mathbf{0}$, where $\boldsymbol{U}(\boldsymbol{\phi}^{\boldsymbol{\eta}};0)$ is the score function

$$\begin{aligned}\boldsymbol{U}(\boldsymbol{\phi}^{\boldsymbol{\eta}};0) = \sum_{i=1}^{N}\sum_{k=k_i^l}^{k_i^r} &\Big[(1-Y_{1i,k-1})(1-Y_{2i,k-1})\Big]\Big\{\frac{Y_{1i,k}Y_{2i,k}}{\pi_{12i,k}}\frac{\partial\pi_{12i,k}}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}} \\ &+\frac{Y_{1i,k}(1-Y_{2i,k})}{\pi_{1i,k}-\pi_{12i,k}}\Big(\frac{\partial\pi_{1i,k}}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}-\frac{\partial\pi_{12i,k}}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}\Big)+\frac{(1-Y_{1i,k})Y_{2i,k}}{\pi_{2i,k}(0)-\pi_{12i,k}}\Big(\frac{\partial\pi_{2i,k}(0)}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}-\frac{\partial\pi_{12i,k}}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}\Big) \\ &+\frac{(1-Y_{1i,k})(1-Y_{2i,k})}{1-\pi_{1i,k}-\pi_{2i,k}(0)+\pi_{12i,k}}\Big(-\frac{\partial\pi_{1i,k}}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}-\frac{\partial\pi_{2i,k}(0)}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}+\frac{\partial\pi_{12i,k}}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}\Big)\Big\} \\ &+\Big[Y_{1i,k-1}(1-Y_{2i,k-1})\Big]\Big\{\frac{Y_{2i,k}}{\pi_{2i,k}(1)}\frac{\partial\pi_{2i,k}(1)}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}-\frac{1-Y_{2i,k}}{1-\pi_{2i,k}(1)}\frac{\partial\pi_{2i,k}(1)}{\partial\boldsymbol{\phi}^{\boldsymbol{\eta}}}\Big\}.\end{aligned}$$

If we assume that the true functions $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_\theta$ follow the B-spline representation, and let $\boldsymbol{\phi}_0^{\boldsymbol{\eta}}$ denote the true value of the parameter, then by standard maximum likelihood theory (e.g. Van der Vaart, 2000), $E[\boldsymbol{U}(\boldsymbol{\phi}_0^{\boldsymbol{\eta}};0)]=0$, $\widehat{\boldsymbol{\phi}}$ is consistent for $\boldsymbol{\phi}_0^{\boldsymbol{\eta}}$ and $\sqrt{N}(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}}-\boldsymbol{\phi}_0^{\boldsymbol{\eta}})$ converges in distribution to a $N(0,V)$ multivariate normal random variable, and V is estimated by $\widehat{V} = -[\nabla\boldsymbol{\phi}^{\boldsymbol{\eta}}\boldsymbol{U}(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}};0)]^{-1}$.

If we do not assume that the true functions $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_\theta$ follow the B-spline representation, then some bias is expected. In this case, by theory of misspecified models, $\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}}$ converges in probability to $\boldsymbol{\phi}_*^{\boldsymbol{\eta}}$, the solution of $E[\boldsymbol{U}(\boldsymbol{\phi}_*^{\boldsymbol{\eta}};0)]=0$, and $\sqrt{N}(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}}-\boldsymbol{\phi}_*^{\boldsymbol{\eta}})$ converges in distribution to a $N(0,\widetilde{V})$ multivariate normal random variable, and $\widetilde{V}$ is estimated by the sandwich estimator

$$\widehat{\widetilde{V}} = \boldsymbol{U}^T(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}};0)[\nabla\boldsymbol{\phi}^{\boldsymbol{\eta}}\boldsymbol{U}(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}};0)]^{-1}\boldsymbol{U}(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}};0).$$

A.3.2 Theory for the penalized estimator

Turning to the penalized estimator, $\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}} = \widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}}_{\boldsymbol{\lambda}}$ for a fixed value $\boldsymbol{\lambda}$.

$$\boldsymbol{U}(\boldsymbol{\phi}^{\boldsymbol{\eta}};\boldsymbol{\lambda}) = \boldsymbol{U}(\boldsymbol{\phi}^{\boldsymbol{\eta}};0) + \widetilde{\boldsymbol{U}}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_\theta;\boldsymbol{\lambda}) \tag{A.2}$$

where $\widetilde{\boldsymbol{U}}(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3;\boldsymbol{\lambda})$ is zero for the entries associated with $\boldsymbol{\beta}$ and $2\lambda_1\boldsymbol{D}_m^T\boldsymbol{D}_m\boldsymbol{\eta}_1$, $2\lambda_2\boldsymbol{D}_m^T\boldsymbol{D}_m\boldsymbol{\eta}_2$ and $2\lambda_3\boldsymbol{D}_m^T\boldsymbol{D}_m\boldsymbol{\eta}_\theta$ for the entries associated with $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_\theta$, respectively. The estimator $\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}}$ solves $\boldsymbol{U}(\boldsymbol{\phi}^{\boldsymbol{\eta}};\boldsymbol{\lambda})=0$. Let $\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}}_{\boldsymbol{\lambda}}$ be the solution of $E[\boldsymbol{U}(\boldsymbol{\phi}^{\boldsymbol{\eta}};\boldsymbol{\lambda})]=0$. Under standard assumptions for estimating equations (Van der Vaart, 2000), the proof continues with the Taylor expansion (here for $N^{-1/2}\boldsymbol{U}(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda})$)

$$0 = N^{-1/2}\boldsymbol{U}(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda}) + \mathcal{H}(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda})N^{-1/2}(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}} - \widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}}) + o_p(1)$$

which gives

$$N^{-1/2}(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}} - \widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}}) = \mathcal{H}(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda})^{-1}N^{-1/2}\boldsymbol{U}(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda})$$

Now, because $\boldsymbol{U}(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda}) = \sum_{i=1}^{N}\boldsymbol{U}_i(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda})$, by the central limit theorem we have that

$$N^{-1/2}[\boldsymbol{U}(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda})] \xrightarrow{D} N\left[0, \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{U}_i(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda})\boldsymbol{U}_i(\widetilde{\boldsymbol{\phi}}^{\boldsymbol{\eta}};\boldsymbol{\lambda})^T\right].$$

Putting everything together, we get the estimator for $Var(\widehat{\boldsymbol{\phi}}^{\boldsymbol{\eta}})$ given by Equation (9) of the main text.

A.4 Details of the simulation studies

This Section presents details on the simulations we carried out. Our first simulation study took the longitudinal bivariate process as the data generating mechanism. For this study, describe the data generating mechanism (Section A.4.1), before moving to review the different analyses we considered (Section A.4.2). We present and discuss the results (Section A.4.3).

Our second simulation study focused on the impact of censoring on the original time-to-events T_{1i} and T_{i2} on the potential degree of disagreement between the three approaches proposed in Section 3.2 of the main text. The data generating mechanism, analyses we carried and the results are all summarized in Section A.4.4.

A.4.1 Data generation

With the ACT dataset in mind, we considered the following data-generating mechanism. The time-scale was $(65, 100]$ with interval length of 2.5. Therefore, $t_1 = 67.5, t_2 = 70, ...t_K = 100$ and $K = 14$. Data were simulated from Model given by Equations (5)-(7) in the main text taking the intercepts $\alpha_{1,k}, \alpha_{2,k}, \alpha_{\theta,k}$ from the following underline functions

$$\begin{aligned}
\alpha_1(t_k) &= \text{logit}[0.005 + 0.002(t_k - 65) + 0.0008(t_k - 70)^2 - 0.0000128(t_k - 62.5)^3] \\
\alpha_2(t_k) &= \text{logit}[0.03 + 0.003(t_k - 65) + 0.00016(t_k - 65)^2] \\
\alpha_\theta(t_k) &= [0.9 + 0.07(t_k - 65) - 0.0032(t_k - 65)^2]I\{t_k \leq 95\}.
\end{aligned}$$

The black circles in Figure A.2 depict these three time-varying functions. The functions approximate a scenario with an increasing non-terminal event probability until around $t = 92.5$, which is then approximately constant; a terminal event probability that increases with time; and an odds ratio that starts around 3, slowly increases and then decreases with time, such that the odds ratio for $t \geq 95$ is 1. We also included two covariates denoted by X_1 and $X_2(t)$. The first was a baseline continuous covariate with $N(0, 1)$ distribution, and the second was a binary time-dependent covariate, simulated as follows. At baseline, $Pr[X_2(65) = 1] = 0.6$, and then $P[X_2(t + 2.5) = 1|X_2(t) = 1] = 0.9$, while zero values were retained for the rest of the follow-up. We took f_1, f_2 and f_θ to be linear with the following specification

$$\begin{aligned}
f_1(\boldsymbol{X}^H_{i2,k}, Y_{1i,k-1}; \boldsymbol{\beta}_1) &= \boldsymbol{X}^T_{i,k}\boldsymbol{\beta}_1 \\
f_2(\boldsymbol{X}^H_{i2,k}, Y_{1i,k-1}; \boldsymbol{\beta}_2) &= \boldsymbol{X}^T_{i,k}\boldsymbol{\beta}_{2,X} + Y_{1i,k-1}\beta_{2,y} \\
f_\theta(\boldsymbol{X}^H_{i2,k}, Y_{1i,k-1}; \boldsymbol{\theta}) &= \boldsymbol{X}^T_{i,k}\boldsymbol{\beta}_\theta.
\end{aligned}$$

We considered three simulation scenarios for the relationship between X_1 and $X_2(t)$ and terminal and nonterminal events. The scenarios also differ with respect to the value of the long-term dependence parameter $\beta_{2,y}$.

(I) **The null scenario**: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{2,X} = \boldsymbol{\beta}_\theta = (0, 0)$; $\beta_{2,y} = 0$.

(II) **The simple dependence scenario**: $\boldsymbol{\beta}_1 = (\log(0.7), \log(3)); \boldsymbol{\beta}_{2,X} = (\log(0.5), 0); \boldsymbol{\beta}_\theta = (0, 0)$; $\beta_{2,y} = \log(1.4)$.

(III) **The complex dependence scenario**: $\boldsymbol{\beta}_1 = (\log(0.7), \log(3)); \boldsymbol{\beta}_{2,X} = (\log(0.5), 0); \boldsymbol{\beta}_\theta = (\log(1.5), \log(0.5))$; $\beta_{2,y} = \log(1.4)$.

We considered sample sizes $N = 500, 1000, 5000$, and we also compared between right-censoring rates of approximately 0%, 20% and 30%.

Under the simple dependence scenario with 30% censoring rate, on average, 7.5% of the observations were censored after Alzheimer's diagnosis, 37% were censored before diagnosed with Alzheimer's or death, 16.5% died after being diagnosed with Alzheimer's, and 39% died during follow-up without Alzheimer's diagnosis. These rates were comparable with those of the ACT dataset.

A.4.2 Analyses

For each combination of parameter scenario, sample size and censoring rate we simulated 1000 datasets. We compared three variations of the proposed methodology, all fit model (2.5)-(2.7) but differ in the way the $\boldsymbol{\alpha}$ functions are estimated: (1) An unstructured model that concerns each $\alpha_{1,k}, \alpha_{2,k}$ and $\alpha_{\theta,k}$, $k = 1, ...K$ as individual parameters. (2) Penalized B-splines with $\widetilde{J} = 5$ knots for each time-varying function ($\widetilde{J}_1 = \widetilde{J}_2 = \widetilde{J}_\theta$) and (3) Penalized B-splines with $\widetilde{J} = 10$. For the penalized estimators, we compared different $\boldsymbol{\lambda}$ values for the penalty tuning parameter. We took the same $\lambda = \lambda_1 = \lambda_2 = \lambda_\theta$ for all three penalty tuning parameters and considered the values $\lambda = 0, 0.1, 1, 5, 10, 25$. We also considered the estimator that chooses λ to optimize the AIC. The choice $\lambda = 0$ corresponds to maximizing the non-penalized likelihood $\mathcal{L}(\boldsymbol{\phi}^{\boldsymbol{\eta}})$. Standard errors of the coefficients $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_\theta)$ were estimated using the estimator given in Equation (9) of the main text.

The aforementioned methods are all implemented in our **R** package `LongitSemiComp`, currently available from the Github account of the first author. For maximizing the likelihood, both for the unstructured and the B-spline models, a Limited-memory BFGS algorithm was used through the `optim` function in **R**. The algorithm uses the gradient function, which we calculated and coded. Whenever the optimization algorithm failed, a standard BFGS was used instead. Initial values were either all zeros or all -0.1, the latter was used when the algorithm with the former choice of all zeros failed to converge. Table A.1 presents the proportion of times the optimization failed and no estimates were produced. Out of 351 unique data/analysis scenarios in our simulations, 331 had 100% convergence rates; that is, there were no failures. As can be seen from the table, more failures were observed for the combination of low sample size, high censoring rate and for the models richer in parameters. Specifically, when 10 knots were used for low sample size and heavy censoring with no penalty. This demonstrates that adding even a small penalty may mitigate potential convergence issues, and, that when sample size is limited, more parsimonious models should be preferred. To guarantee fast implantation of the optimization algorithm, the likelihood and gradient functions were all written in cpp (and integrated using the **Rcpp** package). Running times for the ACT dataset for $K = 14$ intervals and ~ 10 variables were 3-7 seconds (including SE estimation). For $K = 7$ intervals running times were 1-2.5 seconds.

A.4.3 Results

A.4.3.1 Time-varying components

Figures A.2 and A.3 compare different strategies for estimating the time-varying functions expit($\boldsymbol{\alpha}_1$), expit($\boldsymbol{\alpha}_2$), and exp($\boldsymbol{\alpha}_\theta$), for 30% censoring rate and different sample sizes. Figures A.2 compares the non B-spline estimator, and B-spline estimators with different number of knots and with penalty level ($\lambda = 5$). Figure A.3 summarizes our study of the impact of number of knots selection and compares different penalty levels for the B-spline estimator.

The time-varying terminal reference probability function expit($\boldsymbol{\alpha}_1$) was well-estimated in all methods, regardless of the choice of number of knots and the penalty level. The non-terminal probability function was generally well-estimated until age 85. For later ages, when the number of knots was small ($\widetilde{J} = 5$) and the amount of regularization was substantial ($\lambda \geq 1$) combined, the B-spline estimator was oversmoothed and resulted in some bias in later time points, where less information is available. When the number of knots was $\widetilde{J} = 10$, some bias was observed for ages 90 or older, when sample size was small and the amount of regularization was substantial ($\lambda \geq 5$) combined. This bias disappeared for larger sample sizes.

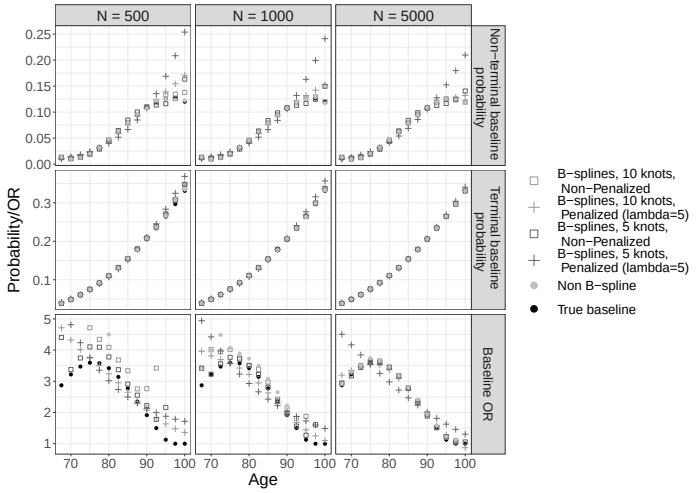
For the time-varying reference odds ratio exp($\boldsymbol{\alpha}_\theta$), values larger than 5 are not presented. Such values were obtained for the non-penalized estimators for the time-varying odds ratio (left bottom square in each panel of Figure A.2), for the later age periods. These findings reflect the instability of non-penalized estimators when sample size was small and censoring was substantial. The time-varying odds ratio function was well estimated by the 10 knots B-spline estimator, as well as by the B-spline non-smoothed 5 knots B-spline estimator. Similar to the non-terminal probability, the 5 knots oversmoothed estimator suffered from bias. Additionally, the model

Table A.1: Proportion of optimization failures. Scenarios with no failures (the vast majority) were omitted.

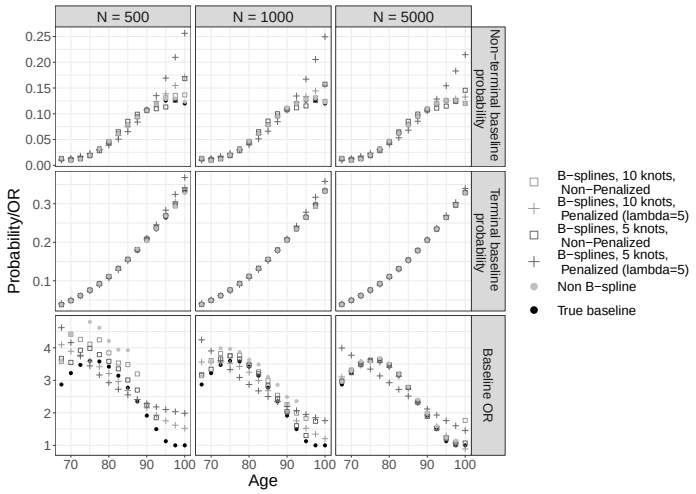
Scenario	Censoring	N	Method	λ	% failed
(I)	0%	500	10 knots	0	2.4
		500	10 knots	5	0.1
		1000	10 knots	0	0.1
	20%	500	10 knots	0	5.7
		1000	10 knots	0	0.3
	30%	500	Unconstrained		0.5
		500	10 knots	0	18.8
		1000	10 knots	0	2.3
(II)	0%	500	10 knots	0	0.8
	20%	500	10 knots	0	5.5
		1000	10 knots	0	0.5
	30%	500	10 knots	0	16.6
		1000	10 knots	0	3.4
(III)	0%	500	10 knots	0	1.6
		5000	5 knots	0	0.2
	20%	500	10 knots	0	7.3
		1000	10 knots	0	0.3
	30%	500	Unconstrained		0.2
		500	10 knots	0	19.5
		1000	10 knots	0	3.5

with completely unrestricted baseline worked well until the last time point, where some bias was observed probably due to smaller sample size. The instability of the undersmoothed estimators for time-varying odds ratio, and biased, yet stable, estimators when using excessive smoothing was strongly demonstrated for lower sample sizes. To summarize, for smaller sample sizes or when using 5 knots only, capturing the non-monotone odds-ratio function was more challenging. Oversmoothing resulted in an undesired monotone function, while no penalization resulted in unstable estimators for the later age periods. Nevertheless, under nearly all scenarios, some penalization ($\lambda = 0.1$ or $\lambda = 0.1$) yielded satisfactory results.

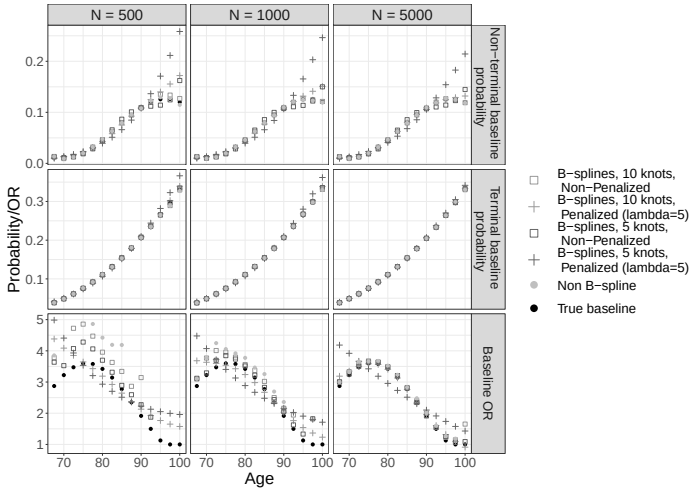
Table A.2 further examines the issue of optimal penalty level. For different sample sizes and censoring rates, and for each choice of number of knots, it shows the (discrete) distribution of the optimal λ values, obtained as the value that minimizes the AIC, as described in Section 4.2 of the main text. Each row describes the proportion of times each λ value was chosen out of 1000 simulation iterations. The results generally agreed with Figure A.3 in that the optimal level of smoothing was usually $\lambda = 0.1$ or $\lambda = 1$. In all simulation scenarios, larger penalty was desirable for $\widetilde{J} = 10$ knots, when compared with $\widetilde{J} = 5$ knots.



Scenario (I): Null

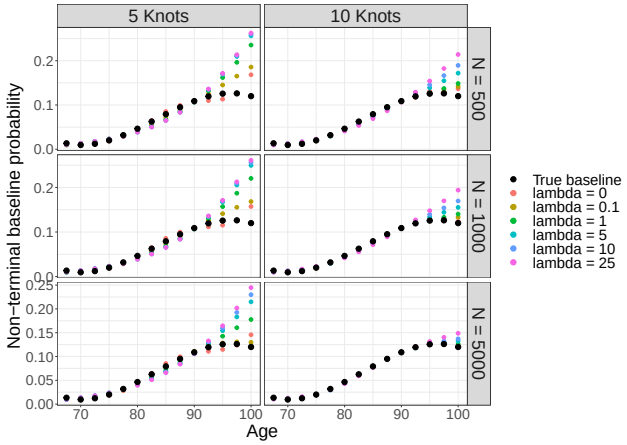


Scenario (II): Simple Dependence

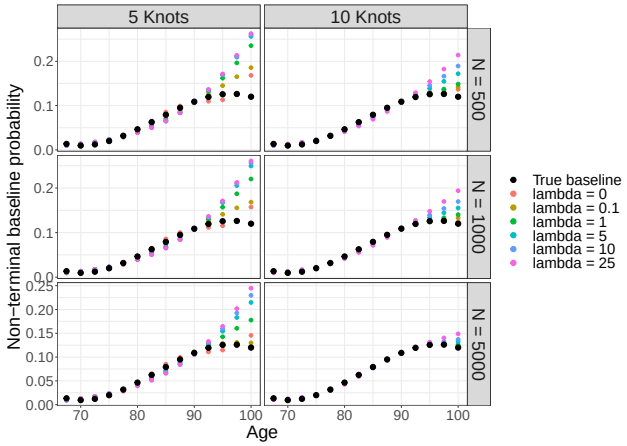


Scenario (III): Complex Dependence

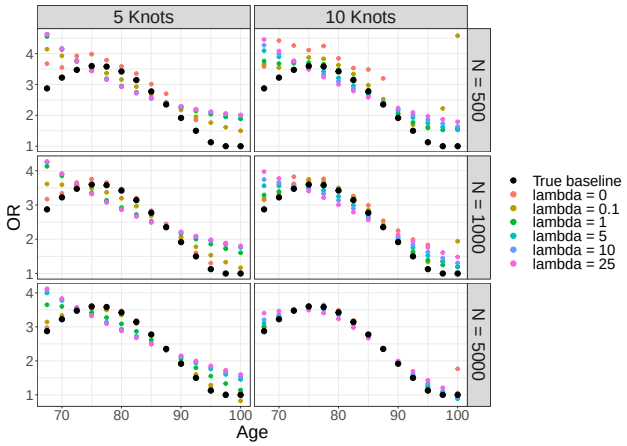
Figure A.2: True and mean estimated expit($\boldsymbol{\alpha}_1$), expit($\boldsymbol{\alpha}_2$) and exp($\boldsymbol{\alpha}_\theta$) under Scenario (I)–(III) with 30% censoring rate, and for different sample sizes. In the bottom row of each panel, values larger than 5.0 were not plotted



Non-terminal event probability



Terminal event probability



Odds ratio

Figure A.3: Impact of number of knots ($\widetilde{J}$) and penalty level (λ) on performance of the time-varying functions $\text{expit}(\boldsymbol{\alpha}_1), \text{expit}(\boldsymbol{\alpha}_2)$ and $\exp(\boldsymbol{\alpha}_\theta)$ for different sample sizes. Results presented for Scenario (II) under censoring rate 20%. The +'s represent mean estimated functions over 1000 simulation iterations per scenario.

Table A.2: Frequency of optimal λ in terms of AIC for different sample sizes, censoring rates, and choices of number of knots. Each row describes the proportion of times each λ value was chosen out of 1000 simulation iterations. The mode in each row is in bold.

N	Censoring	$\widetilde{J}$	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 5$	$\lambda = 10$	$\lambda = 25$
500	0%	5	0.26	**0.55**	0.12	0.02	0.00	0.06
		10	0.00	0.08	0.24	**0.28**	0.16	0.24
	30%	5	0.25	**0.32**	0.13	0.03	0.01	0.26
		10	0.01	0.11	0.27	0.18	0.09	**0.34**
1000	0%	5	0.43	**0.54**	0.02	0.00	0.00	0.00
		10	0.01	0.09	**0.35**	0.30	0.15	0.10
	30%	5	0.42	**0.43**	0.06	0.02	0.00	0.07
		10	0.01	0.12	**0.40**	0.26	0.09	0.13
5000	0%	5	**0.88**	0.11	0.00	0.00	0.00	0.00
		10	0.01	0.11	**0.69**	0.17	0.02	0.00
	30%	5	**0.90**	0.10	0.00	0.00	0.00	0.00
		10	0.01	0.11	**0.71**	0.17	0.01	0.00

A.4.3.2 Dependence parameters

Tables A.3, A.4 and A.5 present present estimation bias, mean estimated standard error, empirical standard error (over simulations) and empirical coverage rate of the 95% confidence intervals for the three dependence parameters $\beta_{2,y}$, $\beta_{\theta 1}$ and $\beta_{\theta 2}$ under dependence simulation scenarios (I), (II) and (III), respectively, and for different sample sizes and censoring rates. The results are presented for an estimator that used B-splines with 10 knots and penalty level $\lambda = 1$. The long-term dependence parameter $\beta_{2,y}$ was well-estimated, with negligible bias and with approximately the desired nominal coverage of the confidence interval, for all sample sizes and censoring rates. A small finite-sample bias was observed for $\beta_{\theta 1}$ (Scenario (II) only) and $\beta_{\theta 2}$ (Scenario (III) only), and it decreased with the increase in the sample size. We note that in most cases our variance estimators performed very well, and the empirical coverage (over simulations) of the confidence intervals was close to the desired nominal level.

We further studied of the impact of different strategies for estimating the time-varying functions on the the performance of the dependence parameter estimators (Table A.6, under scenario (II)). The performance of the long term dependence parameter was hardly affected by the time-varying function estimators. The bias remained negligible, although it was more substantial when only 5 knots were used (compared with 10 knots), for all penalty levels considered. The results for $\widehat{\beta}_{\theta 1}$ and $\widehat{\beta}_{\theta 2}$ were similar to those described in Table A.4, with the following conclusion emerging. Whenever estimation of the time-varying function was unstable, some bias was showed for relevant coefficients. For example, under low sample size ($N = 500$) using the non B-spline estimator or the unpenalized B-spline estimator have lead to more considerable bias in estimating $\beta_{\theta 1}$ and $\beta_{\theta 2}$. The bias generally decreased with the sample size. Impressively, the coverage rate of the confidence intervals remained relatively close to the desired 95% level.

Table A.3: Performance of dependence parameter estimators under the null scenario (I) for different sample sizes and censoring rates. Baseline functions were estimated using the penalized B-splines estimator with $\widetilde{J} = 10$ knots and penalty level $\lambda = 1$. EST.SE: mean estimated standard error; EMP.SE: empirical standard error across simulation; CP95: empirical coverage proportion the true value was included in the 95% Wald confidence interval.

	$N = 500$			$N = 1,000$			$N = 5,000$		
Censoring	0%	20%	30%	0%	20%	30%	0%	20%	30%
					$\widehat{\beta}_{2,y}$				
BIAS	-0.002	-0.001	-0.004	-0.004	0.000	0.000	0.001	0.001	0.001
EST.SE	0.150	0.185	0.223	0.106	0.130	0.156	0.047	0.058	0.069
EMP.SE	0.149	0.193	0.237	0.109	0.135	0.158	0.048	0.058	0.070
CP95	0.951	0.946	0.938	0.940	0.939	0.945	0.945	0.944	0.942
					$\widehat{\beta}_{\theta 1}$				
BIAS	0.000	-0.003	-0.011	0.007	0.003	0.004	0.000	0.000	0.005
EST.SE	0.201	0.239	0.273	0.140	0.165	0.188	0.063	0.073	0.082
EMP.SE	0.205	0.247	0.293	0.147	0.162	0.185	0.060	0.070	0.079
CP95	0.959	0.947	0.945	0.938	0.955	0.960	0.959	0.962	0.961
					$\widehat{\beta}_{\theta 2}$				
BIAS	0.009	-0.026	-0.027	-0.001	-0.004	-0.014	-0.001	0.000	-0.007
EST.SE	0.449	0.526	0.595	0.312	0.363	0.410	0.138	0.159	0.178
EMP.SE	0.466	0.552	0.706	0.314	0.383	0.427	0.142	0.163	0.180
CP95	0.947	0.958	0.939	0.952	0.939	0.947	0.945	0.950	0.951

Table A.4: Performance of dependence parameter estimators under scenario (II) for different sample sizes and censoring rates. Baseline functions were estimated using the penalized B-splines estimator with $\widetilde{J} = 10$ knots and penalty level $\lambda = 1$. EST.SE: mean estimated standard error; EMP.SE: empirical standard error across simulation; CP95: empirical coverage proportion the true value was included in the 95% Wald confidence interval.

	$N = 500$			$N = 1,000$			$N = 5,000$		
Censoring	0%	20%	30%	0%	20%	30%	0%	20%	30%
					$\widehat{\beta}_{2,y}$				
BIAS	-0.003	0.006	-0.004	0.000	0.001	0.000	0.002	0.001	0.001
EST.SE	0.138	0.165	0.192	0.097	0.116	0.135	0.043	0.052	0.060
EMP.SE	0.141	0.167	0.191	0.099	0.113	0.138	0.043	0.053	0.060
CP95	0.932	0.944	0.948	0.956	0.959	0.941	0.956	0.953	0.949
					$\widehat{\beta}_{\theta 1}$				
BIAS	-0.014	-0.019	-0.031	-0.005	-0.013	-0.015	-0.003	-0.005	-0.006
EST.SE	0.232	0.263	0.291	0.161	0.182	0.200	0.071	0.080	0.088
EMP.SE	0.244	0.275	0.308	0.161	0.186	0.212	0.070	0.076	0.086
CP95	0.944	0.953	0.944	0.950	0.945	0.931	0.951	0.961	0.948
					$\widehat{\beta}_{\theta 2}$				
BIAS	0.004	-0.003	-0.007	0.016	0.015	0.022	0.001	-0.006	-0.007
EST.SE	0.426	0.488	0.539	0.296	0.337	0.371	0.131	0.148	0.163
EMP.SE	0.438	0.496	0.566	0.288	0.334	0.374	0.135	0.150	0.163
CP95	0.943	0.951	0.957	0.957	0.952	0.947	0.950	0.949	0.956

Table A.5: Performance of dependence parameter estimators under scenario (III) for different sample sizes and censoring rates. Baseline functions were estimated using the penalized B-splines estimator with $\widetilde{J} = 10$ knots and penalty level $\lambda = 1$. EST.SE: mean estimated standard error; EMP.SE: empirical standard error across simulation; CP95: empirical coverage proportion the true value was included in the 95% Wald confidence interval.

	$N = 500$			$N = 1,000$			$N = 5,000$		
Censoring	0%	20%	30%	0%	20%	30%	0%	20%	30%
					$\widehat{\beta}_{2,y}$				
BIAS	0.002	-0.002	-0.004	-0.003	0.000	-0.001	0.002	0.003	0.001
EST.SE	0.136	0.162	0.188	0.096	0.114	0.132	0.043	0.051	0.059
EMP.SE	0.137	0.163	0.193	0.097	0.112	0.133	0.042	0.051	0.060
CP95	0.950	0.949	0.945	0.951	0.953	0.947	0.956	0.947	0.946
					$\widehat{\beta}_{\theta 1}$				
BIAS	-0.005	0.006	-0.005	0.000	0.003	-0.001	-0.002	-0.004	-0.002
EST.SE	0.236	0.270	0.303	0.197	0.188	0.209	0.073	0.083	0.092
EMP.SE	0.245	0.277	0.325	0.162	0.193	0.212	0.074	0.085	0.090
CP95	0.947	0.950	0.937	0.950	0.948	0.941	0.952	0.945	0.955
					$\widehat{\beta}_{\theta 2}$				
BIAS	-0.006	-0.024	-0.044	-0.017	0.004	0.001	-0.005	-0.012	-0.015
EST.SE	0.459	0.525	0.583	0.353	0.364	0.402	0.143	0.161	0.177
EMP.SE	0.478	0.539	0.604	0.324	0.387	0.407	0.140	0.156	0.177
CP95	0.941	0.938	0.946	0.951	0.938	0.953	0.941	0.958	0.946

Table A.6: Comparison of performance for the model parameters $\boldsymbol{\beta}_\theta$ and $\beta_{2,y}$ when different estimation strategies are used for the time-varying function. Results presented under the simple dependence scenario (II) and censoring level 20%. NBS: Non-B-spline estimator. EST.SE: mean estimated standard error; EMP.SE: empirical standard error across simulation; CP95: empirical coverage proportion the true value was included in the 95% Wald confidence interval.

	$N = 500$							$N = 1,000$						
	NBS	$\widetilde{J} = 5$			$\widetilde{J} = 10$			NBS	$\widetilde{J} = 5$			$\widetilde{J} = 10$		
		$\lambda = 0$	$\lambda = 1$	$\lambda = 5$	$\lambda = 0$	$\lambda = 1$	$\lambda = 5$		$\lambda = 0$	$\lambda = 1$	$\lambda = 5$	$\lambda = 0$	$\lambda = 1$	$\lambda = 5$
							$\widehat{\beta}_{2,y}$							
BIAS	-0.002	-0.005	-0.007	-0.010	0.001	-0.004	-0.004	0.001	-0.001	-0.002	-0.005	0.000	0.000	0.001
EST.SE	0.193	0.192	0.192	0.192	0.192	0.192	0.191	0.135	0.135	0.135	0.135	0.135	0.135	0.135
EMP.SE	0.193	0.191	0.192	0.192	0.189	0.191	0.191	0.138	0.138	0.138	0.138	0.138	0.138	0.138
CP95	0.948	0.950	0.948	0.946	0.953	0.948	0.947	0.939	0.939	0.937	0.936	0.938	0.941	0.940
							$\widehat{\beta}_{\theta 1}$							
BIAS	-0.114	-0.047	-0.024	-0.024	-0.092	-0.031	-0.027	-0.050	-0.021	-0.012	-0.012	-0.033	-0.015	-0.013
EST.SE	0.343	0.302	0.287	0.288	0.324	0.291	0.289	0.214	0.203	0.197	0.197	0.209	0.200	0.198
EMP.SE	0.382	0.321	0.301	0.301	0.352	0.308	0.303	0.234	0.215	0.208	0.208	0.224	0.212	0.209
CP95	0.939	0.943	0.947	0.948	0.950	0.944	0.946	0.928	0.931	0.931	0.932	0.931	0.931	0.931
							$\widehat{\beta}_{\theta 2}$							
BIAS	0.040	0.004	-0.015	-0.015	0.022	-0.007	-0.011	0.048	0.024	0.016	0.015	0.044	0.022	0.019
EST.SE	0.622	0.554	0.538	0.538	0.588	0.539	0.541	0.395	0.375	0.370	0.371	0.387	0.371	0.370
EMP.SE	0.700	0.587	0.558	0.558	0.654	0.566	0.560	0.412	0.379	0.368	0.368	0.390	0.374	0.371
CP95	0.932	0.949	0.954	0.954	0.940	0.957	0.957	0.939	0.949	0.951	0.953	0.945	0.947	0.950

A.4.3.3 $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_{2,X}$

Tables A.7, A.8, and A.9 present estimation bias, mean estimated standard error, empirical standard error (over simulations) and empirical coverage rate of the 95% confidence intervals for the coefficients relating covariates to the non-terminal and terminal probabilities, i.e. $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_{2,X}$. Each table presents results for a single dependence scenario ((I), (II), or (III)) for different sample sizes and censoring rates, using the penalized B-spline estimator for the time-varying function ($\widetilde{J} = 10$ knots, $\lambda = 1$). In all three scenarios, the bias was practically zero, the standard error estimator performed well, and the empirical coverage rate was approximately the desired 95%. Table A.10 further presents the impact of different strategies for estimating the time-varying functions on the the performance of $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_{2,X}$ (under scenario (II) with 20% censoring). Generally, even for low sample size, the bias remained negligible for nearly all four parameters ($\beta_{11}, \beta_{12}, \beta_{21,X}, \beta_{22,X}$), although it was slightly biased when using the non-B-spline estimator or when taking 5 knots only for the B-spline estimator (compared with 10 knots). For $\widehat{\beta}_{12}$, the increased bias was more substantial compared to Table A.8, under low sample size ($N = 500$) using the non B-spline estimator or the unpenalized B-spline estimator. Nevertheless, this bias generally decreased with the increase in the sample size. The coverage rate of the confidence intervals remained relatively close to the desired 95% level.

Table A.7: Performance of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_{2,X}$ estimators under the null scenario (I) for different sample sizes and censoring rates. Baseline functions were estimated using the penalized B-splines estimator with $\widetilde{J} = 10$ knots and penalty level $\lambda = 1$. EST.SE: mean estimated standard error; EMP.SE: empirical standard error across simulation; CP95: empirical coverage proportion the true value was included in the 95% Wald confidence interval.

	$N = 500$			$N = 1,000$			$N = 5,000$		
Censoring	0%	20%	30%	0%	20%	30%	0%	20%	30%
					$\widehat{\beta}_{11}$				
BIAS	-0.002	-0.001	-0.001	0.000	0.000	-0.002	0.003	0.002	0.002
EST.SE	0.085	0.098	0.110	0.060	0.069	0.078	0.027	0.031	0.035
EMP.SE	0.086	0.102	0.109	0.059	0.069	0.078	0.027	0.031	0.035
CP95	0.951	0.943	0.953	0.956	0.945	0.946	0.944	0.947	0.940
					$\widehat{\beta}_{12}$				
BIAS	0.003	0.003	-0.011	0.005	-0.004	-0.008	0.001	-0.003	-0.001
EST.SE	0.187	0.214	0.237	0.132	0.151	0.167	0.059	0.067	0.074
EMP.SE	0.195	0.219	0.242	0.140	0.151	0.163	0.062	0.067	0.077
CP95	0.943	0.945	0.942	0.935	0.941	0.953	0.932	0.953	0.944
					$\widehat{\beta}_{21,X}$				
BIAS	-0.001	0.003	0.002	0.000	0.001	0.001	0.000	-0.001	-0.001
EST.SE	0.051	0.059	0.066	0.036	0.042	0.046	0.016	0.018	0.021
EMP.SE	0.052	0.062	0.068	0.036	0.041	0.046	0.017	0.018	0.020
CP95	0.947	0.939	0.931	0.947	0.947	0.942	0.934	0.965	0.965
					$\widehat{\beta}_{22,X}$				
BIAS	0.003	0.009	0.011	-0.002	0.000	-0.001	0.001	0.001	0.003
EST.SE	0.113	0.127	0.140	0.079	0.090	0.098	0.035	0.040	0.044
EMP.SE	0.114	0.128	0.138	0.081	0.090	0.098	0.036	0.042	0.045
CP95	0.939	0.952	0.958	0.946	0.949	0.954	0.956	0.937	0.942

Table A.8: Performance of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_{2,X}$ estimators under the simple dependence scenario (II) for different sample sizes and censoring rates. Baseline functions were estimated using the penalized B-splines estimator with $\widetilde{J} = 10$ knots and penalty level $\lambda = 1$. EST.SE: mean estimated standard error; EMP.SE: empirical standard error across simulation; CP95: empirical coverage proportion the true value was included in the 95% Wald confidence interval.

	$N = 500$			$N = 1,000$			$N = 5,000$		
Censoring	0%	20%	30%	0%	20%	30%	0%	20%	30%
					$\widehat{\beta}_{11}$				
BIAS	-0.001	-0.001	-0.005	-0.001	-0.003	-0.002	-0.001	0.000	0.000
EST.SE	0.084	0.094	0.102	0.059	0.066	0.072	0.026	0.029	0.032
EMP.SE	0.088	0.094	0.102	0.060	0.067	0.074	0.026	0.029	0.031
CP95	0.940	0.948	0.957	0.940	0.946	0.948	0.950	0.962	0.951
					$\widehat{\beta}_{12}$				
BIAS	0.017	0.020	0.020	0.004	0.006	0.008	0.003	0.002	0.005
EST.SE	0.167	0.189	0.208	0.117	0.133	0.146	0.052	0.059	0.065
EMP.SE	0.168	0.194	0.212	0.121	0.133	0.145	0.054	0.059	0.065
CP95	0.952	0.947	0.951	0.951	0.952	0.953	0.944	0.953	0.956
					$\widehat{\beta}_{21,X}$				
BIAS	-0.005	-0.006	-0.006	-0.004	-0.003	-0.002	0.000	0.000	0.000
EST.SE	0.061	0.069	0.075	0.043	0.048	0.052	0.019	0.021	0.023
EMP.SE	0.063	0.067	0.075	0.043	0.049	0.052	0.019	0.021	0.023
CP95	0.940	0.960	0.952	0.954	0.946	0.944	0.946	0.955	0.956
					$\widehat{\beta}_{22,X}$				
Bias	0.007	0.005	0.008	0.000	-0.002	-0.003	0.000	-0.001	0.000
EST.SE	0.117	0.131	0.142	0.082	0.092	0.100	0.037	0.041	0.045
EMP.SE	0.114	0.133	0.144	0.083	0.092	0.099	0.037	0.041	0.045
CP95	0.954	0.944	0.945	0.942	0.946	0.951	0.951	0.955	0.949

Table A.9: Performance of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_{2,X}$ estimators under the complex dependence scenario (III) for different sample sizes and censoring rates. Baseline functions were estimated using the penalized B-splines estimator with $\widetilde{J} = 10$ knots and penalty level $\lambda = 1$. EST.SE: mean estimated standard error; EMP.SE: empirical standard error across simulation; CP95: empirical coverage proportion the true value was included in the 95% Wald confidence interval.

	$N = 500$			$N = 1,000$			$N = 5,000$		
Censoring	0%	20%	30%	0%	20%	30%	0%	20%	30%
					$\widehat{\beta}_{11}$				
BIAS	-0.002	0.001	0.000	0.001	-0.001	-0.001	-0.001	-0.001	0.000
EST.SE	0.086	0.096	0.104	0.060	0.067	0.073	0.027	0.030	0.033
EMP.SE	0.088	0.098	0.108	0.060	0.069	0.074	0.027	0.030	0.032
CP95	0.938	0.950	0.944	0.951	0.941	0.948	0.943	0.952	0.954
					$\widehat{\beta}_{12}$				
BIAS	0.000	0.001	0.004	0.000	0.006	0.009	0.002	-0.001	0.001
EST.SE	0.169	0.191	0.209	0.120	0.135	0.147	0.053	0.060	0.066
EMP.SE	0.171	0.192	0.209	0.118	0.135	0.148	0.054	0.059	0.064
CP95	0.942	0.956	0.954	0.956	0.951	0.943	0.940	0.952	0.954
					$\widehat{\beta}_{21,X}$				
BIAS	-0.005	-0.005	-0.005	0.000	-0.002	-0.004	-0.001	-0.001	-0.001
EST.SE	0.061	0.069	0.075	0.043	0.048	0.053	0.019	0.022	0.023
EMP.SE	0.061	0.069	0.075	0.043	0.047	0.052	0.019	0.021	0.023
CP95	0.951	0.944	0.942	0.955	0.954	0.954	0.957	0.959	0.944
					$\widehat{\beta}_{22,X}$				
BIAS	0.006	0.002	0.001	0.003	0.005	0.004	-0.002	0.000	-0.001
EST.SE	0.117	0.131	0.143	0.083	0.092	0.100	0.037	0.041	0.045
EMP.SE	0.123	0.132	0.145	0.081	0.090	0.099	0.037	0.041	0.045
CP95	0.930	0.948	0.947	0.952	0.952	0.955	0.948	0.954	0.949

Table A.10: Comparison of performance for the model parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_{2,X}$ when different time-varying function estimators are used, including number of knots and penalty level for the B-spline estimator. Results presented under the simple dependence scenario (II) and censoring level 20%. NBS: Non-B-spline estimator. EST.SE: mean estimated standard error; EMP.SE: empirical standard error across simulation; CP95: empirical coverage proportion the true value was included in the 95% Wald confidence interval.

	$N = 500$							$N = 1,000$						
	NBS	$\widetilde{J} = 5$			$\widetilde{J} = 10$			NBS	$\widetilde{J} = 5$			$\widetilde{J} = 10$		
		$\lambda = 0$	$\lambda = 1$	$\lambda = 5$	$\lambda = 0$	$\lambda = 1$	$\lambda = 5$		$\lambda = 0$	$\lambda = 1$	$\lambda = 5$	$\lambda = 0$	$\lambda = 1$	$\lambda = 5$
							$\widehat{\beta}_{11}$							
BIAS	-0.006	-0.005	-0.006	-0.007	-0.005	-0.005	-0.004	-0.003	-0.003	-0.003	-0.004	-0.004	-0.002	-0.002
EST.SE	0.103	0.103	0.102	0.102	0.103	0.102	0.102	0.072	0.072	0.072	0.072	0.072	0.072	0.072
EMP.SE	0.104	0.102	0.101	0.101	0.105	0.102	0.101	0.074	0.074	0.074	0.074	0.074	0.074	0.074
CP95	0.954	0.958	0.958	0.957	0.946	0.957	0.957	0.944	0.947	0.948	0.948	0.948	0.948	0.949
							$\widehat{\beta}_{12}$							
BIAS	0.029	0.024	0.015	0.016	0.027	0.020	0.017	0.012	0.010	0.002	0.003	0.010	0.008	0.005
EST.SE	0.210	0.209	0.209	0.210	0.209	0.208	0.207	0.146	0.146	0.146	0.147	0.146	0.146	0.145
EMP.SE	0.215	0.213	0.212	0.213	0.217	0.212	0.211	0.146	0.145	0.146	0.147	0.147	0.145	0.145
CP95	0.952	0.947	0.950	0.950	0.951	0.951	0.951	0.955	0.954	0.954	0.955	0.951	0.953	0.951
							$\widehat{\beta}_{21,X}$							
BIAS	-0.008	-0.006	-0.007	-0.007	-0.009	-0.006	-0.006	-0.003	-0.002	-0.003	-0.004	-0.003	-0.002	-0.002
EST.SE	0.075	0.075	0.075	0.075	0.075	0.075	0.075	0.053	0.053	0.052	0.052	0.053	0.052	0.052
EMP.SE	0.076	0.075	0.075	0.075	0.077	0.075	0.075	0.052	0.052	0.052	0.052	0.053	0.052	0.052
CP95	0.950	0.951	0.953	0.954	0.948	0.952	0.953	0.943	0.946	0.944	0.944	0.942	0.944	0.943
							$\widehat{\beta}_{22,X}$							
BIAS	0.007	0.008	0.006	0.005	0.009	0.008	0.008	-0.003	-0.003	-0.004	-0.006	-0.003	-0.003	-0.003
EST.SE	0.143	0.142	0.142	0.142	0.143	0.142	0.142	0.100	0.100	0.100	0.100	0.100	0.100	0.100
EMP.SE	0.146	0.145	0.144	0.144	0.146	0.144	0.144	0.100	0.099	0.099	0.099	0.100	0.099	0.099
CP95	0.946	0.943	0.946	0.946	0.945	0.945	0.946	0.950	0.951	0.951	0.950	0.948	0.951	0.951

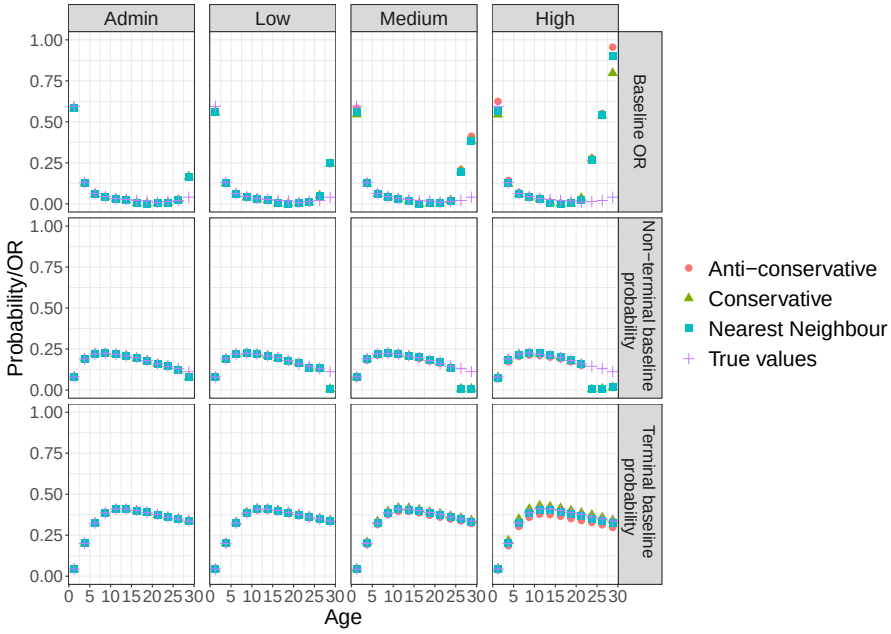
A.4.4 Impact of censoring on sensitivity approaches

To study the impact of censoring on the three different approaches to determine k_i^l and k_i^r, we have conducted additional simulation study. In this study, data were simulated from an illness-death model (see Section A.1.3) using the `SemiCompRisks` R package (Alvares et al., 2019). Data were first created from a Cox model with two covariates and Weibull baseline distributions for the three transition rates and a frailty Gamma random variable (variance 0.5) that induce further dependence, beyond additional two covariates included. All simulations are fully reproducible from repository in the first author's Github account. To calculate the true values induced by this data generating mechanism, we carried a single simulation with 10M observations and no censoring, under the two partitions $\boldsymbol{\tau}^{2.5}$ and $\boldsymbol{\tau}^5$ and focused on the time axis $(0, 30)$. As can be seen in Figure A.4 the baseline odds ratio were nearly zero for most time points, because, under this data generating mechanism, those diagnosed with the disease in later time points tend to live longer, and do not die at the same interval.

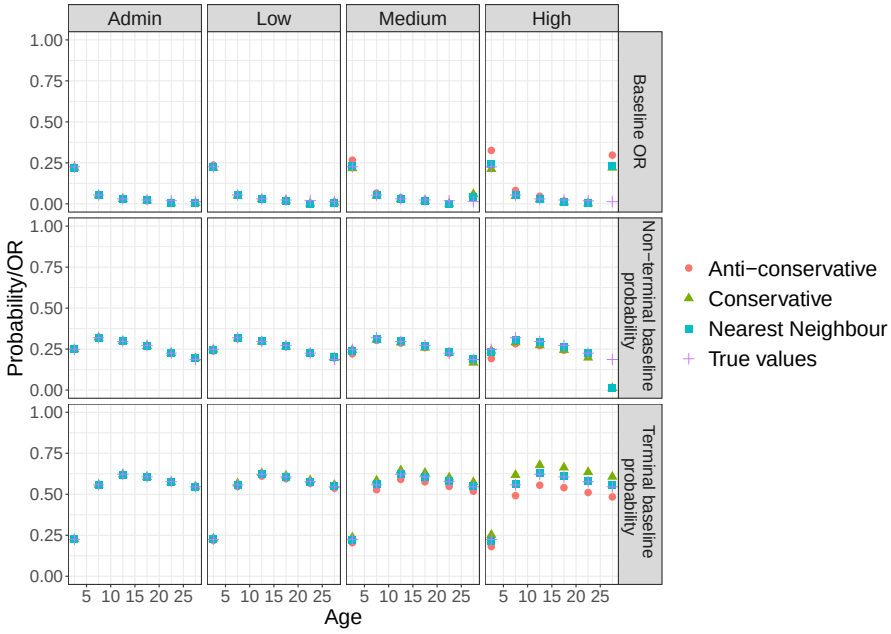
We considered four censoring scenarios. In the administrative censoring scenario, there was only censoring at time 30, mimicking end-of-study administrative censoring. Under this scenario, censoring rate as 6% and about 47% of the sample had the non-terminal event prior to this time (and prior to death). We then considered additional Exponential censoring, and determine the rates to achieve different censoring rates. The three additional scenarios are called Low (15% censoring), Medium (30%) censoring, and High (50% censoring). We compared between the three approaches to determine k_i^l and k_i^r, the nearest neighbor, the conservative and the anti-conservative approaches that were described in Section 3.2 of the main text. To reduce the dependence of the results on modeling assumptions and penalty specification, estimates were obtained by the unstructured model and for sample size of $N = 5,000$.

As seen in Figure A.4, for both partitions the median estimated time-varying functions $\text{expit}(\boldsymbol{\alpha}_1)$, $\text{expit}(\boldsymbol{\alpha}_2)$ and $\exp(\boldsymbol{\alpha}_\theta)$ the results were nearly identical in the Administrative and Low and censoring scenarios. Furthermore, the estimators are unbiased. For the Medium and High censoring rates some disagreement is shown, especially in either the early or later time points. Generally, the nearest neighbor approach performed the best, and produced unbiased estimators for nearly all scenarios and time points, with small number of exceptions.

The results for the regression coefficients (Table A.11) revealed similar trends. The nearest neighbor approach has shown little bias, even for the High censoring scenario. Disagreement level was more prominent for the coefficient of the odds-ratio sub model, namely $\boldsymbol{\beta}_\theta$.



(a) $\boldsymbol{\tau}^{2.5}$ partition



(b) $\boldsymbol{\tau}^{5.0}$ partition

Figure A.4: True and median estimated expit($\boldsymbol{\alpha}_1$), expit($\boldsymbol{\alpha}_2$) and exp($\boldsymbol{\alpha}_\theta$) when data were simulated from an illness-death model under two partitions and different censoring rates. Results are based on 1,000 simulation iterations.

Table A.11: Comparison of performance for the model parameters $\boldsymbol{\beta}$ under different approaches to deal with censoring when determining k_i^l and k_i^r. Results presented for under the four censoring scenarios. Cens: censoring scenario. Admin: Administrative. anticons: anti-conservative. cons: conservative. NN: nearest neighbour.

$\boldsymbol{\tau}$			β_{11}	β_{12}	$\beta_{21,\boldsymbol{X}}$	$\beta_{22,\boldsymbol{X}}$	$\beta_{\theta 1}$	$\beta_{\theta 2}$	$\beta_{2,y}$
$\boldsymbol{\tau}^{2.5}$	True values		1.30	-0.27	0.21	-0.40	0.20	0.84	-0.69
	Cens	Approach	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
	Admin	anticons	1.30 (0.03)	-0.28 (0.05)	0.21 (0.02)	-0.40 (0.04)	0.23 (0.19)	0.87 (0.23)	-0.69 (0.05)
		cons	1.30 (0.03)	-0.28 (0.05)	0.21 (0.02)	-0.40 (0.04)	0.23 (0.19)	0.87 (0.23)	-0.69 (0.05)
		NN	1.30 (0.03)	-0.28 (0.05)	0.21 (0.02)	-0.40 (0.04)	0.23 (0.19)	0.87 (0.23)	-0.69 (0.05)
	Low	anticons	1.31 (0.04)	-0.28 (0.05)	0.22 (0.02)	-0.41 (0.04)	0.24 (0.19)	0.87 (0.24)	-0.67 (0.05)
		cons	1.31 (0.04)	-0.28 (0.05)	0.22 (0.02)	-0.41 (0.04)	0.23 (0.19)	0.87 (0.24)	-0.68 (0.05)
		NN	1.31 (0.04)	-0.28 (0.05)	0.22 (0.02)	-0.41 (0.04)	0.23 (0.19)	0.87 (0.24)	-0.67 (0.05)
	Medium	anticons	1.32 (0.04)	-0.29 (0.06)	0.22 (0.02)	-0.44 (0.04)	0.25 (0.19)	0.86 (0.25)	-0.64 (0.06)
		cons	1.33 (0.04)	-0.29 (0.06)	0.22 (0.02)	-0.44 (0.04)	0.23 (0.19)	0.86 (0.26)	-0.65 (0.06)
		NN	1.33 (0.04)	-0.29 (0.06)	0.22 (0.02)	-0.44 (0.04)	0.23 (0.19)	0.86 (0.26)	-0.65 (0.06)
	High	anticons	1.33 (0.04)	-0.30 (0.06)	0.24 (0.03)	-0.48 (0.05)	0.26 (0.20)	0.85 (0.28)	-0.58 (0.07)
		cons	1.35 (0.04)	-0.30 (0.06)	0.24 (0.03)	-0.49 (0.05)	0.22 (0.21)	0.87 (0.28)	-0.59 (0.07)
		NN	1.35 (0.04)	-0.30 (0.06)	0.24 (0.03)	-0.48 (0.05)	0.22 (0.21)	0.87 (0.28)	-0.59 (0.07)
$\boldsymbol{\tau}^{2.5}$	True values		1.32	-0.19	0.20	-0.42	0.16	0.88	-0.71
	Cens	Approx	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
	Admin	anticons	1.31 (0.04)	-0.19 (0.06)	0.20 (0.02)	-0.42 (0.04)	0.17 (0.13)	0.90 (0.17)	-0.71 (0.06)
		cons	1.31 (0.04)	-0.19 (0.06)	0.20 (0.02)	-0.42 (0.04)	0.17 (0.13)	0.90 (0.17)	-0.71 (0.06)
		NN	1.31 (0.04)	-0.19 (0.06)	0.20 (0.02)	-0.42 (0.04)	0.17 (0.13)	0.90 (0.17)	-0.71 (0.06)
	Low	anticons	1.31 (0.04)	-0.20 (0.06)	0.20 (0.02)	-0.43 (0.04)	0.22 (0.13)	0.87 (0.17)	-0.69 (0.06)
		cons	1.33 (0.04)	-0.19 (0.06)	0.21 (0.02)	-0.44 (0.04)	0.16 (0.14)	0.91 (0.18)	-0.70 (0.06)
		NN	1.33 (0.04)	-0.20 (0.06)	0.21 (0.02)	-0.44 (0.04)	0.16 (0.13)	0.91 (0.17)	-0.70 (0.06)
	Medium	anticons	1.31 (0.04)	-0.20 (0.06)	0.21 (0.02)	-0.45 (0.04)	0.30 (0.14)	0.82 (0.18)	-0.67 (0.06)
		cons	1.34 (0.04)	-0.19 (0.06)	0.22 (0.02)	-0.47 (0.04)	0.15 (0.14)	0.92 (0.19)	-0.70 (0.07)
		NN	1.34 (0.04)	-0.21 (0.06)	0.22 (0.02)	-0.47 (0.04)	0.17 (0.14)	0.90 (0.18)	-0.69 (0.07)
	High	anticons	1.29 (0.04)	-0.21 (0.06)	0.23 (0.03)	-0.49 (0.05)	0.40 (0.15)	0.76 (0.20)	-0.63 (0.08)
		cons	1.36 (0.05)	-0.19 (0.07)	0.25 (0.03)	-0.53 (0.05)	0.12 (0.16)	0.94 (0.21)	-0.68 (0.08)
		NN	1.37 (0.05)	-0.22 (0.07)	0.24 (0.03)	-0.52 (0.05)	0.17 (0.15)	0.90 (0.21)	-0.66 (0.08)

A.5 The ACT dataset

Figure A.5 provides a summary of the observed person-time for the patients. Within each panel, the patients have been ordered by: (i) their age at entry and (ii) the age at which their eventual outcome status is observed. Table A.12 provides a summary of the available baseline covariates in the ACT dataset, overall and stratified according to the four outcome types in the data:

- those censored prior to AD diagnosis and death,
- those diagnosed with AD and then censored prior before death,
- those who died without receiving AD diagnosis before death, and
- those who died after receiving AD diagnosis.

The following tables describe the outcome data in the ACT dataset in the proposed representation:

- Table A.13 presents the data after applying a 2.5 year interval partition according to our approach. For each interval, it presents a 2×2 outcome table (Alzheimer's disease, death, both or neither), along with the respective odds ratio, for those who were still at risk for both events at the beginning of the interval.
- Table A.13 also presents the distribution of $Y_{2,k}$ for each interval k, among those who are at risk of death only, that is, those who were diagnosed with Alzheimer' as some point prior to the interval and are still alive.
- Table A.14 presents the analogue outcome data table under a 5 year interval partition of the time scale.

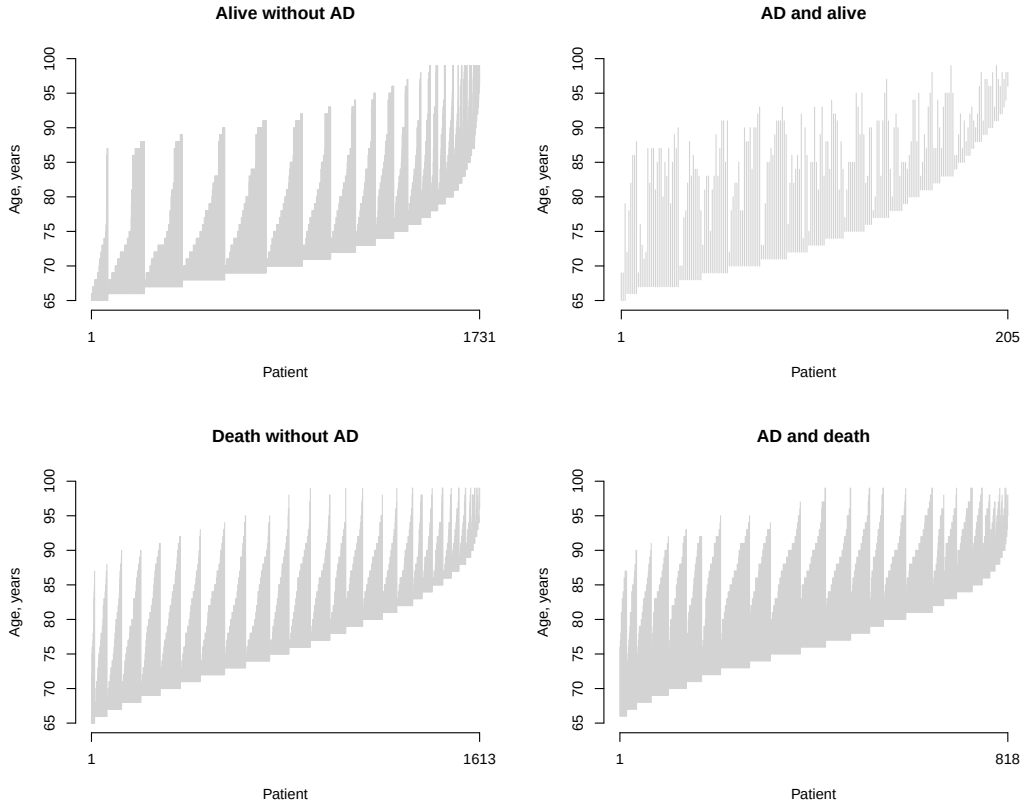


Figure A.5: Summary of person-time while on-study among N=4,367 participants in the ACT study, stratified by whether they had a diagnosis of AD and/or died during follow-up. Within each sub-figure, participants are ordered first by their age of enrollment and second by their observed event/censoring time.

Table A.12: Descriptive information about baseline covariates in the ACT dataset

	Total		Censored prior to AD or death		AD and then censored		Death without AD		AD diagnosis and death	
	n	(%†)	n	(%⋆)	n	(%⋆)	n	(%⋆)	n	(%⋆)
Total	4367	(100)	1,731	(39)	205	(5)	1,613	(37)	818	(19)
APOE ϵ-4 alleles										
None	3,313	(76)	1,323	(40)	141	(4)	1,285	(39)	564	(17)
At least one	1,054	(24)	408	(39)	64	(6)	328	(31)	254	(24)
Gender										
Female	2,513	(58)	1,025	(41)	134	(5)	849	(34)	505	(20)
Male	1,854	(42)	706	(38)	71	(4)	764	(41)	313	(17)
Race										
White	3,958	(91)	1,551	(39)	182	(5)	1,471	(37)	754	(19)
Non-white	409	(9)	180	(44)	23	(6)	142	(35)	64	(16)
Depression										
No	4,035	(92)	1,651	(41)	185	(5)	1,468	(36)	731	(18)
Yes	332	(8)	80	(24)	20	(6)	145	(44)	87	(26)
Education‡ (years)										
< 12	402	(9)	78	(19)	13	(3)	206	(51)	105	(26)
$12 - 15$	1,779	(41)	517	(29)	95	(5)	744	(42)	423	(24)
≥ 15	2,186	(50)	1,136	(52)	97	(4)	663	(30)	290	(13)
Marital status										
Married	2,463	(56)	1,064	(43)	114	(5)	844	(34)	441	(18)
Never Married	166	(4)	89	(54)	6	(4)	49	(30)	22	(13)
Divorced	605	(14)	284	(47)	36	(6)	203	(34)	82	(14)
Widowed	1,010	(23)	258	(26)	41	(4)	457	(45)	254	(25)
Other	123	(3)	36	(29)	8	(7)	60	(49)	19	(15)

† column%
⋆ row%
‡ In the analysis, education was treated as a continuous variable.
Mean: 14.75, SD: 3.22, (Q1,Q2,Q3): (12, 15, 17) Range: (3-21)

Table A.13: Outcome data at each interval according to a 2.5 interval partition. For each $k = 1, ..., 14$, we present the 2×2 table outcome table among those who are free of both events at the beginning of the interval, and the number of those died and survived among those already diagnosed with AD. Note, due to the possibility of an AD event followed by censoring at the same interval, not all those with AD will appear in the following interval as AD patients.

			$k=1$ $(65, 67.5]$		$k=2$ $(67.5, 70]$		$k=3$ $(70, 72.5]$		$k=4$ $(72.5, 75]$		$k=5$ $(75, 77.5]$		$k=6$ $(77.5, 80]$		$k=7$ $(80, 82.5]$	
			Y_1		Y_1		Y_1		Y_1		Y_1		Y_1		Y_1	
			0	1	0	1	0	1	0	1	0	1	0	1	0	1
$Y_{1,k} = 0$	Y_2	0	0	1522	1472	12	1834	11	2259	35	2272	36	2327	88	2168	90
		1	0	4	19	0	32	0	86	4	93	3	181	13	123	6
	OR		NA		0.00		0.00		3.00		2.04		1.90		1.18	
$Y_{1,k} = 1$	Y_2	0		0		0		9		15		34		46		96
		1		0		0		0		3		13		21		34

			$k=8$ $(82.5, 85]$		$k=9$ $(85, 87.5]$		$k=10$ $(87.5, 90]$		$k=11$ $(90, 92.5]$		$k=12$ $(92.5, 95]$		$k=13$ $(95, 97.5]$		$k=14$ $(97.5, 100]$	
			Y_1		Y_1		Y_1		Y_1		Y_1		Y_1		Y_1	
			0	1	0	1	0	1	0	1	0	1	0	1	0	1
$Y_{1,k} = 0$	Y_2	0	1864	183	1480	125	1012	140	624	68	328	53	180	18	73	8
		1	217	36	198	17	269	37	159	10	145	16	38	7	49	7
	OR		1.69		1.02		0.99		0.58		0.68		1.84		1.30	
$Y_{1,k} = 1$	Y_2	0		109		189		158		144		78		50		25
		1		71		87		133		112		105		51		32

Table A.14: Outcome data at each interval according to a 5 interval partition. For each $k = 1, ..., 7$, we present the 2×2 table outcome table among those who are free of both events at the beginning of the interval, and the number of those died and survived among those already diagnosed with AD. Note, due to the possibility ofa n AD event followed by censoring at the same interval, not all those with AD will appear in the following interval as AD patients.

			$k=1$ $(65,70]$ Y_1		$k=2$ $(70,75]$ Y_1		$k=3$ $(75,80]$ Y_1		$k=4$ $(80,85]$ Y_1		$k=5$ $(85,90]$ Y_1		$k=6$ $(90,95]$ Y_1		$k=7$ $(95,100]$ Y_1	
			0	1	0	1	0	1	0	1	0	1	0	1	0	1
$Y_{1,k}=0$	Y_2	0	2707	12	2407	45	2443	112	1988	242	1131	219	430	91	116	17
		1	23	0	118	5	274	28	340	73	467	100	304	56	87	23
	OR		0.00		2.27		2.23		1.76		1.11		0.87		1.80	
$Y_{1,k}=1$	Y_2	0		0		7		25		56		102		69		27
		1		0		2		22		74		174		187		74

A.6 Detailed results from the analyses of data from the Adult Changes in Thought study

A.6.1 Analyses based on existing methods

We consider in this section a series of analyses based on existing methods described in the Introduction of the main text and in Section A.1.

A.6.1.1 An illness-death model with a patient-specific frailty

At the outset we considered the illness-death framework, with the hazard functions for the three transitions (i.e. Healthy ⇒ AD, Healthy ⇒ Death and AD ⇒ Death) specified via the following Cox-type models:

$$\lambda_1(t_1; \boldsymbol{X}_i) = \gamma_i \lambda_{01}(t_1) \exp\{\boldsymbol{X}_i^T \boldsymbol{\xi}_1\}, \quad \text{(A.3)}$$

$$\lambda_2(t_2; \boldsymbol{X}_i) = \gamma_i \lambda_{02}(t_2) \exp\{\boldsymbol{X}_i^T \boldsymbol{\xi}_2\}, \quad \text{(A.4)}$$

$$\lambda_3(t_2|t_1; \boldsymbol{X}_i) = \gamma_i \lambda_{03}(t_2 - t_1) \exp\{\boldsymbol{X}_i^T \boldsymbol{\xi}_3\}, \quad \text{(A.5)}$$

where $\gamma_i \sim \text{Gamma}(\varphi^{-1}, \varphi^{-1})$ is a patient-specific frailty (Lee et al., 2015; Xu et al., 2010). Note, the model for $\lambda_3(t_2|t_1; \cdot)$ is *semi-Markov*; the time scale is time since diagnosis of AD. For all three transitions, the baseline hazard is modeled as a Weibull(ν_g, κ_g), such that $\lambda_{0g}(t) = \nu_g \kappa_g t^{\nu_g - 1}$. These choices were made because, as far as we are aware, there are no implementations of the illness-death model, as specified by expressions (A.3)-(A.5), that permit left truncation (a key feature of the ACT data) and anything other than Weibull baseline hazards.

Tables A.15 and A.16, and Figure A.6 report results for four illness-death models, that differ in whether a patient-specific frailty was incorporated and whether age at AD diagnosis was included in $\lambda_3(t_2|t_1; \cdot)$. A key observation from these analyses is that there is little evidence that the frailties, as they are included in models (A.3)-(A.5), serve to account for any of the dependence between the two events above and beyond how dependence is structured through the interplay of the remaining components of the illness-death model: the point estimates for $\log(\varphi)$ are -14.34 and -13.10, and the log-likelihood at the maximum likelihood estimates are the same whether one includes the frailties or not (Table A.16). Moreover, that the point estimates for θ are (essentially) on the boundary of the parameter space results in the hessian evaluated at the maximum likelihood estimates not being invertible.

Table A.15: Results from four semi-Markov illness-death models, each with Weibull baseline hazards, for the ACT data. Note, for both models that included a frailty, the estimate of the variance was very small: $\widehat{\log(\varphi)} = -14.34$ for Model 1 and $\widehat{\log(\varphi)} = -13.10$ for Model 2. Note, that the point estimates for φ are (essentially) on the boundary of the parameter space results in the hessian evaluated at the maximum likelihood estimates not being invertible. Hence the (NA, NA) reported for the confidence intervals.

	Model 1[a]				Model 2[a]			
	No frailty		W/ frailty		No frailty		W/ frailty	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Healthy ⇒ AD								
Gender: Female	0.94	(0.80, 1.10)	0.94	(NA, NA)	0.94	(0.80, 1.10)	0.94	(NA, NA)
Race: White	1.06	(0.85, 1.32)	1.06	(NA, NA)	1.06	(0.85, 1.32)	1.06	(NA, NA)
College graduate	0.84	(0.73, 0.96)	0.84	(NA, NA)	0.84	(0.73, 0.96)	0.84	(NA, NA)
Marital status:								
Never	0.99	(0.68, 1.45)	0.99	(NA, NA)	0.99	(0.68, 1.45)	0.99	(NA, NA)
Divorced	1.02	(0.83, 1.25)	1.02	(NA, NA)	1.02	(0.83, 1.25)	1.02	(NA, NA)
Widowed	0.97	(0.83, 1.13)	0.97	(NA, NA)	0.97	(0.83, 1.13)	0.97	(NA, NA)
Other	0.82	(0.56, 1.21)	0.82	(NA, NA)	0.82	(0.56, 1.21)	0.82	(NA, NA)
Depression	1.57	(1.28, 1.92)	1.57	(NA, NA)	1.57	(1.28, 1.92)	1.57	(NA, NA)
$\geq$ 1 APOE ϵ4 allele								
Main effect	1.76	(1.41, 2.19)	1.76	(NA, NA)	1.76	(1.41, 2.19)	1.76	(NA, NA)
× Gender: Female	1.03	(0.78, 1.36)	1.03	(NA, NA)	1.03	(0.78, 1.36)	1.03	(NA, NA)
Healthy ⇒ Death								
Gender: Female	0.60	(0.53, 0.67)	0.60	(NA, NA)	0.60	(0.53, 0.67)	0.60	(NA, NA)
Race: White	1.01	(0.85, 1.20)	1.01	(NA, NA)	1.01	(0.85, 1.20)	1.01	(NA, NA)
College graduate	0.92	(0.82, 1.02)	0.92	(NA, NA)	0.92	(0.82, 1.02)	0.92	(NA, NA)
Marital status:								
Never	1.15	(0.86, 1.54)	1.15	(NA, NA)	1.15	(0.86, 1.54)	1.15	(NA, NA)
Divorced	1.26	(1.08, 1.48)	1.26	(NA, NA)	1.26	(1.08, 1.48)	1.26	(NA, NA)
Widowed	1.18	(1.04, 1.33)	1.18	(NA, NA)	1.18	(1.04, 1.33)	1.18	(NA, NA)
Other	1.36	(1.04, 1.77)	1.36	(NA, NA)	1.36	(1.04, 1.77)	1.36	(NA, NA)
Depression	1.33	(1.12, 1.58)	1.33	(NA, NA)	1.33	(1.12, 1.58)	1.33	(NA, NA)
$\geq$ 1 APOE ϵ4 allele								
Main effect	1.01	(0.84, 1.21)	1.01	(NA, NA)	1.01	(0.84, 1.21)	1.01	(NA, NA)
× Gender: Female	1.05	(0.82, 1.34)	1.05	(NA, NA)	1.05	(0.82, 1.34)	1.05	(NA, NA)
AD ⇒ Death								
Gender: Female	0.91	(0.76, 1.10)	0.91	(NA, NA)	0.88	(0.73, 1.06)	0.88	(NA, NA)
Race: White	1.49	(1.15, 1.93)	1.49	(NA, NA)	1.45	(1.12, 1.87)	1.45	(NA, NA)
College graduate	0.97	(0.83, 1.14)	0.97	(NA, NA)	1.02	(0.87, 1.20)	1.02	(NA, NA)
Marital status:								
Never	1.09	(0.71, 1.69)	1.09	(NA, NA)	1.20	(0.78, 1.87)	1.20	(NA, NA)
Divorced	0.97	(0.76, 1.23)	0.97	(NA, NA)	1.14	(0.89, 1.45)	1.14	(NA, NA)
Widowed	1.12	(0.95, 1.33)	1.12	(NA, NA)	0.99	(0.84, 1.18)	0.99	(NA, NA)
Other	0.85	(0.53, 1.35)	0.85	(NA, NA)	0.82	(0.51, 1.31)	0.82	(NA, NA)
Depression	1.18	(0.94, 1.48)	1.18	(NA, NA)	1.22	(0.97, 1.53)	1.22	(NA, NA)
$\geq$ 1 APOE ϵ4 allele								
Main effect	0.98	(0.77, 1.26)	0.98	(NA, NA)	1.12	(0.88, 1.43)	1.12	(NA, NA)
× Gender: Female	0.83	(0.61, 1.13)	0.83	(NA, NA)	0.85	(0.62, 1.16)	0.85	(NA, NA)
Age at AD diagnosis[b]					1.37	(1.27, 1.47)	1.37	(NA, NA)

[a] Model 1 does not include adjustment for age at AD diagnosis in the AD ⇒ Death, while Model 2 does.

[b] Age at diagnosis was standardized so that the hazard ratio corresponds to a 5-year contrast

Table A.16: Log-likelihood at the maximized value for four semi-Markov illness-death models, each with Weibull baseline hazards, for the ACT data.

Include frailty	Adjust for AD diagnosis in AD ⇒ Death	log-Likelihood at maximum
No	No	-12914.81
Yes	No	-12914.81
No	Yes	-12876.65
Yes	Yes	-12876.65

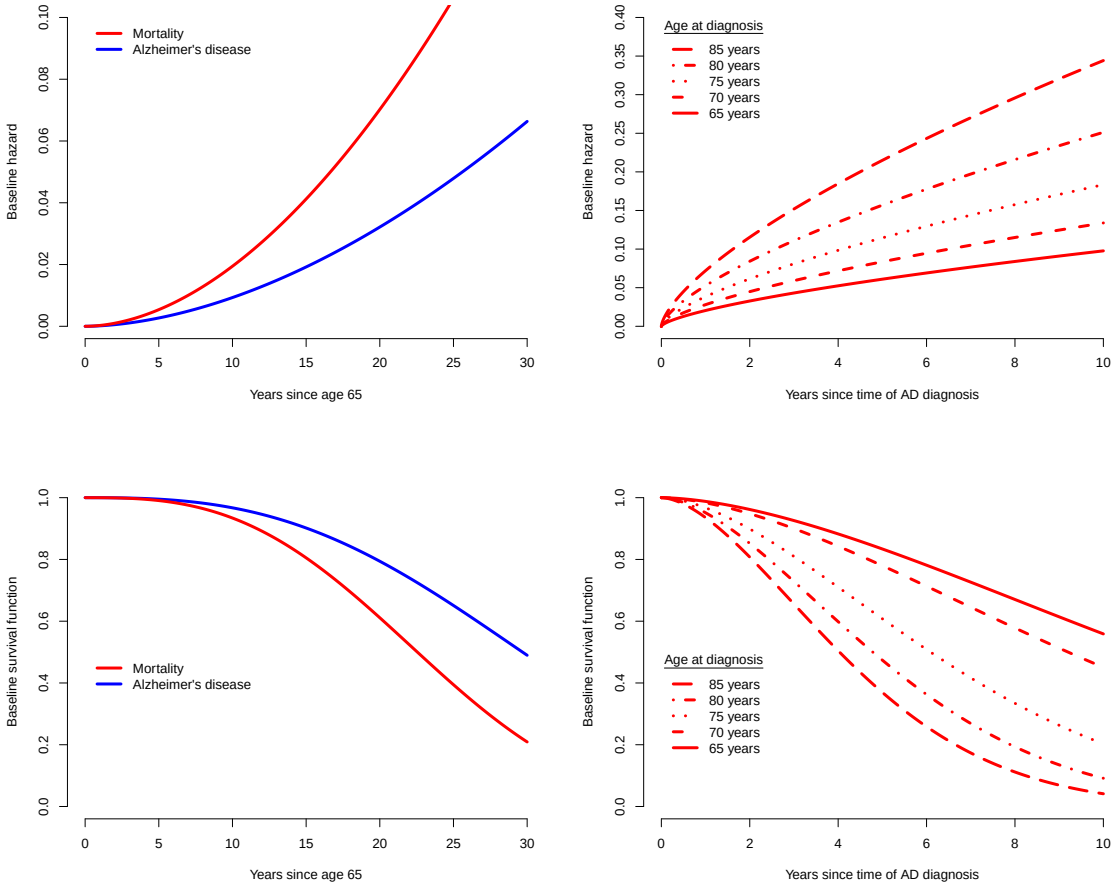


Figure A.6: Baseline hazards and corresponding baseline survivor functions from the semi-Markov Weibull illness-death model that adjusts for age at AD diagnosis but does not include a patient-specific frailty. For the AD ⇒ Death transition, baseline curves are provided for a series of ages at which the AD diagnosis was given.

A.6.1.2 An illness-death model with no patient-specific frailty

Motivated by the above results, we performed additional analyses based on Cox-type specifications for the three transition-specific hazards but without the γ_i frailties. Note, in removing the frailties one can estimate model components using partial likelihood methods for Cox models where the data are subject to left truncation. We fit two semi-Markov models, as in expressions (A.3)-(A.5), again differing in whether age at AD diagnosis was included. We also fit a *Markov* model in which the AD $\Rightarrow$ Death transition is modeled as:

$$\lambda_3(t_2|t_1;\boldsymbol{X}_i) \;=\; \lambda_{03}(t_2)\exp\{\boldsymbol{X}_i^T\boldsymbol{\xi}_3\},$$

so that the time scale is the same as for the other two transitions (i.e. time since age 65).

Tables A.17 and A.18 report estimated hazard ratios and 95% confidence intervals. Overall, the conclusions regarding the associations between the covariates and both Alzheimer's disease and mortality are consistent with the Weibull illness-death model fits. Also consistent is the evidence regarding the dependence between AD and mortality as quantified by the inclusion of age at AD diagnosis in the model for $\lambda_3(t_2|t_1;\boldsymbol{X}_i)$; the estimated hazard ratio for a 5-year contrast is 1.33 (95% CI: 1.24, 1.42).

From the Markov model, since Healthy $\Rightarrow$ Death and the AD $\Rightarrow$ Death transitions are modeled on the same time scale, one can report the explanatory hazard ratio for any given patient profile. The top-left sub-figure of Figure A.7 provides smoothed estimates of EHR(t_2; t_1) for four profiles (Figure A.8 provides additional detail). From Figure A.7 it is clear that, at any given age, the hazard for death is substantially higher for individuals who have been diagnosed with AD relative to those who have not, with the biggest differences among the relatively young. Intuitively, this suggests that, among individuals at least 65 years of age, a diagnosis of AD is more devastating (from a mortality perspective) for a young individual than for an older individual.

Table A.17: Results from univariate Cox regression analyses of the two transitions from out of the 'Healthy' state in an illness-death model for the ACT data.

	Healthy $\Rightarrow$ AD[a]		Healthy $\Rightarrow$ Death[a]	
	HR	95% CI	HR	95% CI
Gender: Female	0.94	(0.80, 1.11)	0.59	(0.52, 0.66)
Race: White	1.06	(0.85, 1.33)	1.01	(0.85, 1.20)
College graduate	0.84	(0.74, 0.96)	0.90	(0.81, 1.00)
Marital status:				
Never	0.99	(0.67, 1.44)	1.13	(0.84, 1.51)
Divorced	1.01	(0.83, 1.24)	1.27	(1.09, 1.49)
Widowed	0.97	(0.84, 1.14)	1.18	(1.04, 1.34)
Other	0.83	(0.56, 1.22)	1.37	(1.05, 1.78)
Depression	1.56	(1.28, 1.92)	1.37	(1.15, 1.62)
≥ 1 APOE $\epsilon 4$ allele:				
Main effect	1.76	(1.41, 2.19)	1.01	(0.84, 1.21)
$\times$ Gender: Female	1.03	(0.78, 1.35)	1.06	(0.83, 1.35)

[a] Death is treated is treated as a censoring mechanism
[b] AD is treated is treated as a censoring mechanism

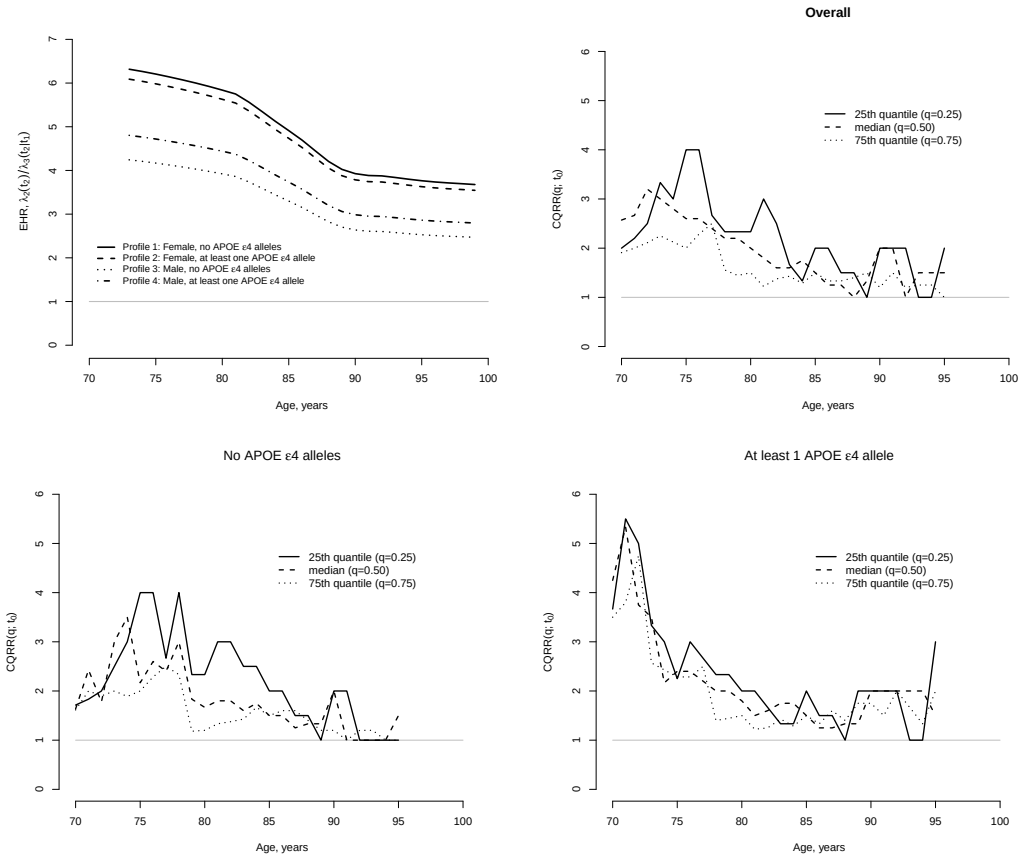


Figure A.7: Estimated EHRs for four patient profiles and CQRRs for the 25th quantile, the median and 75th quantile of residual time for death, overall and stratified by whether there was at least 1 APOE $\epsilon 4$ allele, based on analyses of the ACT data. See Section A.6.1.

Table A.18: Results from three univariate Cox regression analysis of the AD ⇒ Death transition in an illness-death model for the ACT data.

	Markov		Semi-Markov V1[a]		Semi-Markov V2[a]	
	HR	95% CI	HR	95% CI	HR	95% CI
Gender: Female	0.87	(0.73, 1.05)	0.90	(0.75, 1.08)	0.87	(0.73, 1.05)
Race: White	1.32	(1.02, 1.71)	1.51	(1.17, 1.96)	1.47	(1.14, 1.91)
College graduate	1.00	(0.86, 1.17)	0.97	(0.83, 1.13)	1.01	(0.86, 1.18)
Marital status:						
Never	1.17	(0.76, 1.81)	1.11	(0.72, 1.72)	1.21	(0.78, 1.87)
Divorced	1.07	(0.84, 1.37)	0.99	(0.78, 1.25)	1.12	(0.88, 1.43)
Widowed	0.96	(0.81, 1.14)	1.14	(0.96, 1.35)	1.01	(0.85, 1.20)
Other	0.80	(0.50, 1.28)	0.83	(0.52, 1.33)	0.80	(0.50, 1.28)
Depression	1.19	(0.95, 1.49)	1.15	(0.92, 1.45)	1.19	(0.95, 1.49)
≥ 1 APOE $\epsilon 4$ allele:						
Main effect	1.15	(0.89, 1.47)	0.98	(0.76, 1.25)	1.11	(0.87, 1.42)
× Gender: Female	0.90	(0.66, 1.23)	0.85	(0.62, 1.16)	0.85	(0.63, 1.16)
Age at AD diagnosis[b]					1.33	(1.24, 1.42)

[a] Semi-Markov V1 does not include adjustment for age at AD diagnosis in the AD in the model, while Semi-Markov V2 does.

[b] Age at diagnosis was standardized so that the hazard ratio corresponds to a 5-year contrast

A.6.1.3 The cross-quantile residual ratio

Finally, we report results based on the CQRR methodology[1]. Specifically, we calculated the $\text{CQRR}(q; t_0)$ at $q \in \{0.25, 0.5.0.75\}$ (i.e. for the 25th quantile, the median and 75th quantile of residual time for death) for ages $t_0 \in \{70, \ldots, 95\}$ years, stratifying by the number of APOE $\epsilon 4$ alleles (0 vs ≥ 1).

Figure A.7 provides estimates; see Figure A.9 for 95% confidence intervals. From Figure A.7, the estimated $\text{CQRR}(q; t_0)$ are greater than 1.0 at all ages, indicating that residual lifetime, at any given age, for individuals without a diagnosis of AD is estimated to be longer than that for individuals with an AD diagnosis. Comparing the bottom sub-figures of Figure A.7, we see that the spread in the lines is somewhat greater for patients with no APOE $\epsilon 4$ alleles. To interpret this, consider the population of patients with no APOE $\epsilon 4$ alleles and the population with at least one. In comparing patients without an AD diagnosis to those with such a diagnosis, the distribution of residual lifetime exhibits less variation, at any given age, in the second of these populations. Thus, a diagnosis of AD in patients with at least one APOE $\epsilon 4$ allele results in a relatively homogeneous decline whereas the decline associated with a diagnosis of AD in patients with no APOE $\epsilon 4$ alleles is more heterogeneous.

[1] Using code available at http://web1.sph.emory.edu/users/lpeng/Rpackage.html

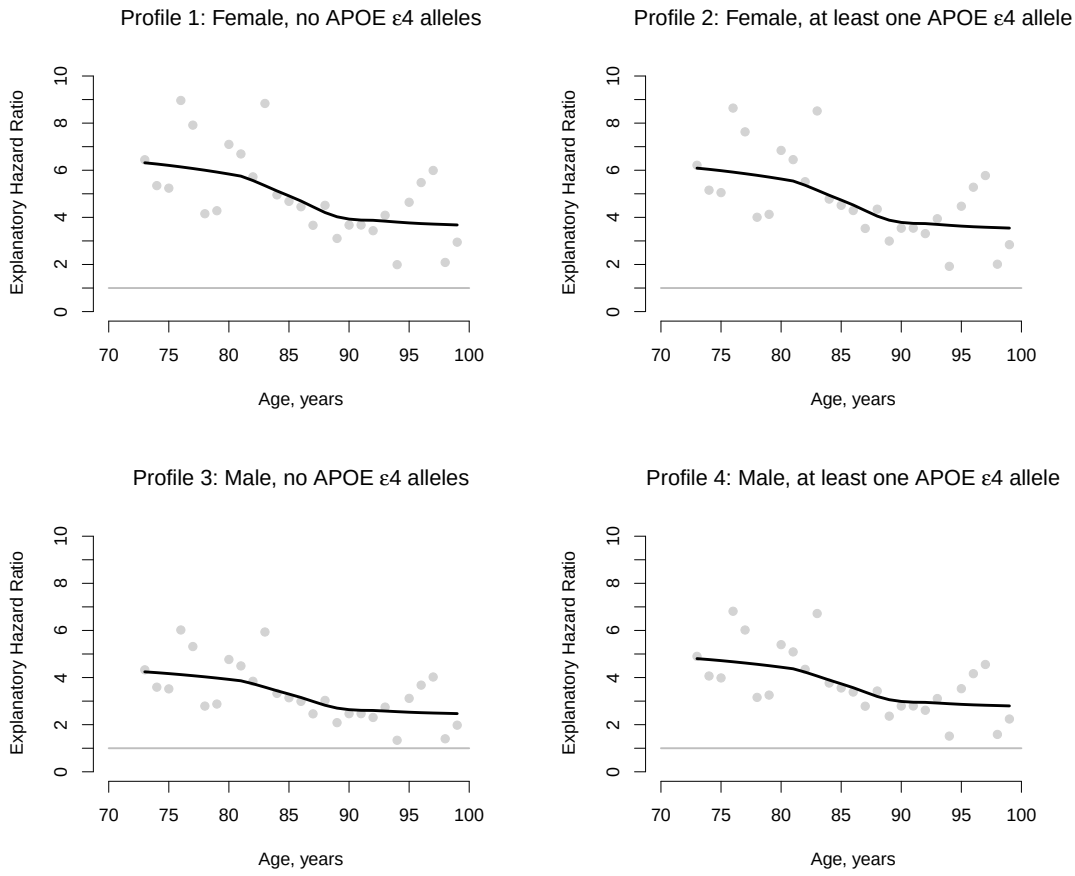


Figure A.8: Estimated age-specific explanatory hazard ratios for four patient profiles based on a Markov illness-death model with no frailty. Within each sub-figure, the line corresponds to the fit of a LOWESS smoother. Common to each profile is that the patients are non-white, married, have a college education, and no depression.

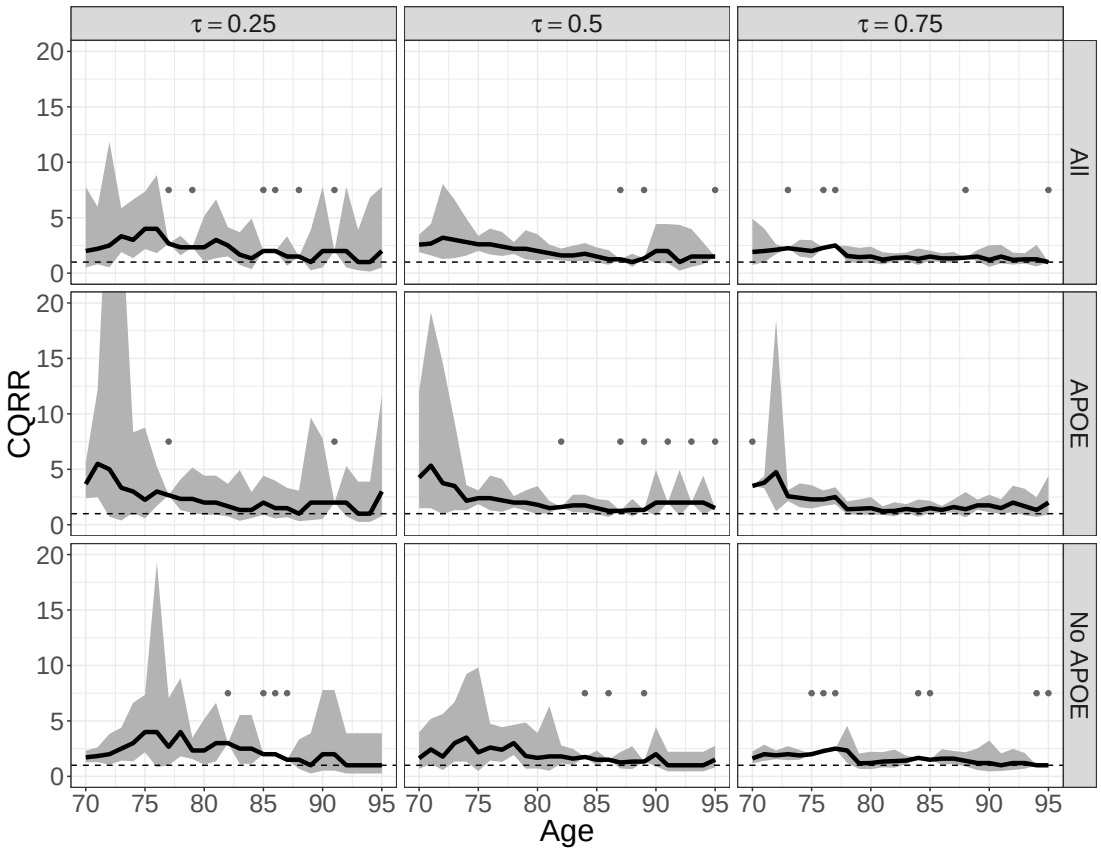


Figure A.9: Estimated CQRRs, with 95% confidence intervals, for $q \in \{0.25, 0.5.0.75\}$, overall and stratified by whether there was at least 1 APOE $\epsilon 4$ allele. The gray points mark time points in which there was no reliable variance estimate.

A.6.2 The proposed framework

We present here the results for six different models: three models under each of the partitions $\boldsymbol{\tau}^{2.5}$ and $\boldsymbol{\tau}^{5}$. All six models include the variables from Table 2 of the main text (Gender, APOE and their three-way interaction with AD). Additionally, they include all variables described in Table A.12. Because $\theta_{i,k}$ speaks to the joint-occurrence of AD and death, more information is needed for precise estimation of baseline trends and and coefficients. Therefore, for each partition, we considered three choices for the variables included in the local odds ratio $\theta_{i,k}$ model. The *full model* include the same variables as in the models for $\pi_{1i,k}$ and $\pi_{2i,k}$. The *reduced model* includes APOE only. The *intermediate* model, which is the model we reports results for in the main text (Table 2), includes Gender, having at least one APOEϵ4 allele and their interaction. Table A.19 reports AICs for the three models under each partitions, for $\lambda = (0, 0.1, 0.5, 1.0, 2.5, 5.0)$. Tables A.20 , A.21 and A.22 report the coefficient estimates and confidence intervals for the intermediate, reduced and full models, respectively. All the results in the aforementioned tables are reported for the NN approach to deal with right censoring and left truncation.

As for the time-varying functions, Figures A.10 and Figures A.11 are extended versions of Figure 3 in the main text. They present the estimated time-varying function for the reference level, taken from the intermediate model for the $\boldsymbol{\tau}^{2.5}$ partition (Figure A.10) and the $\boldsymbol{\tau}^{5.0}$ partition (Figure A.11).

Finally, Tables A.23 and A.23 present comparisons of the obtained point estimates and associated confidence intervals according the strategy to determine k_i^l and k_i^r and according to the model for the time-varying functions (unstructured vs. B-spline). Results are presented for the $\boldsymbol{\tau}^{2.5}$ (Table A.23) and for $\boldsymbol{\tau}^{5.0}$ (Table A.24) partitions.

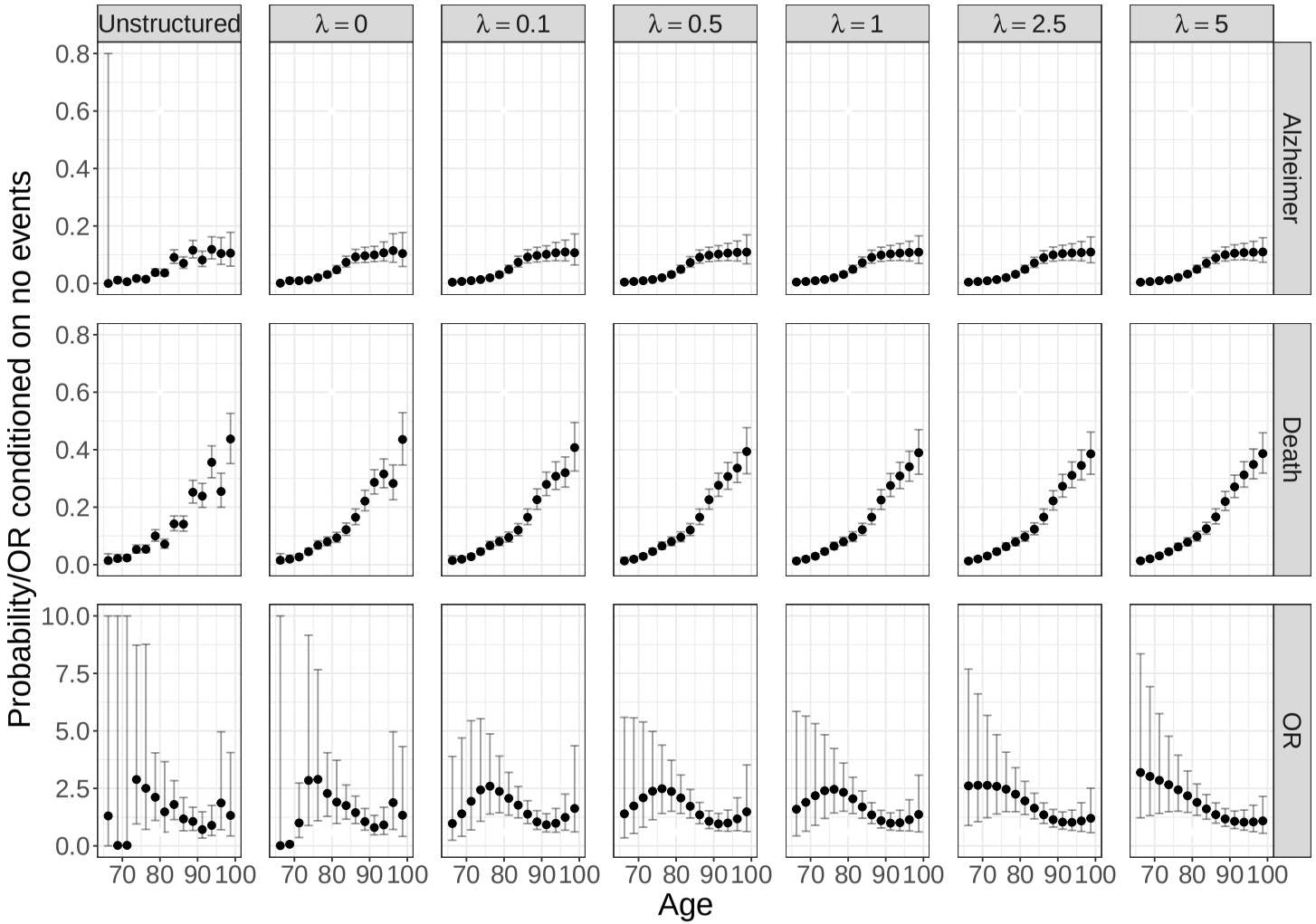


Figure A.10: Estimated baseline time trends, expit($\widehat{\boldsymbol{\alpha}}_1$), expit($\widehat{\boldsymbol{\alpha}}_2$) and exp($\widehat{\boldsymbol{\alpha}}_\theta$) from a series of analyses to the ACT data, under the unstructured specification and the B-spline specification with $\lambda \in \{0.0, 0.1, 0.5, 1, 2.5, 5\}$ and under the $\boldsymbol{\tau}^{2.5}$ partition. Also shown are 95% confidence intervals. Note, the y-axis has been truncated at 0.6 for expit($\widehat{\boldsymbol{\alpha}}_1$) and at 10 for exp($\widehat{\boldsymbol{\alpha}}_\theta$).

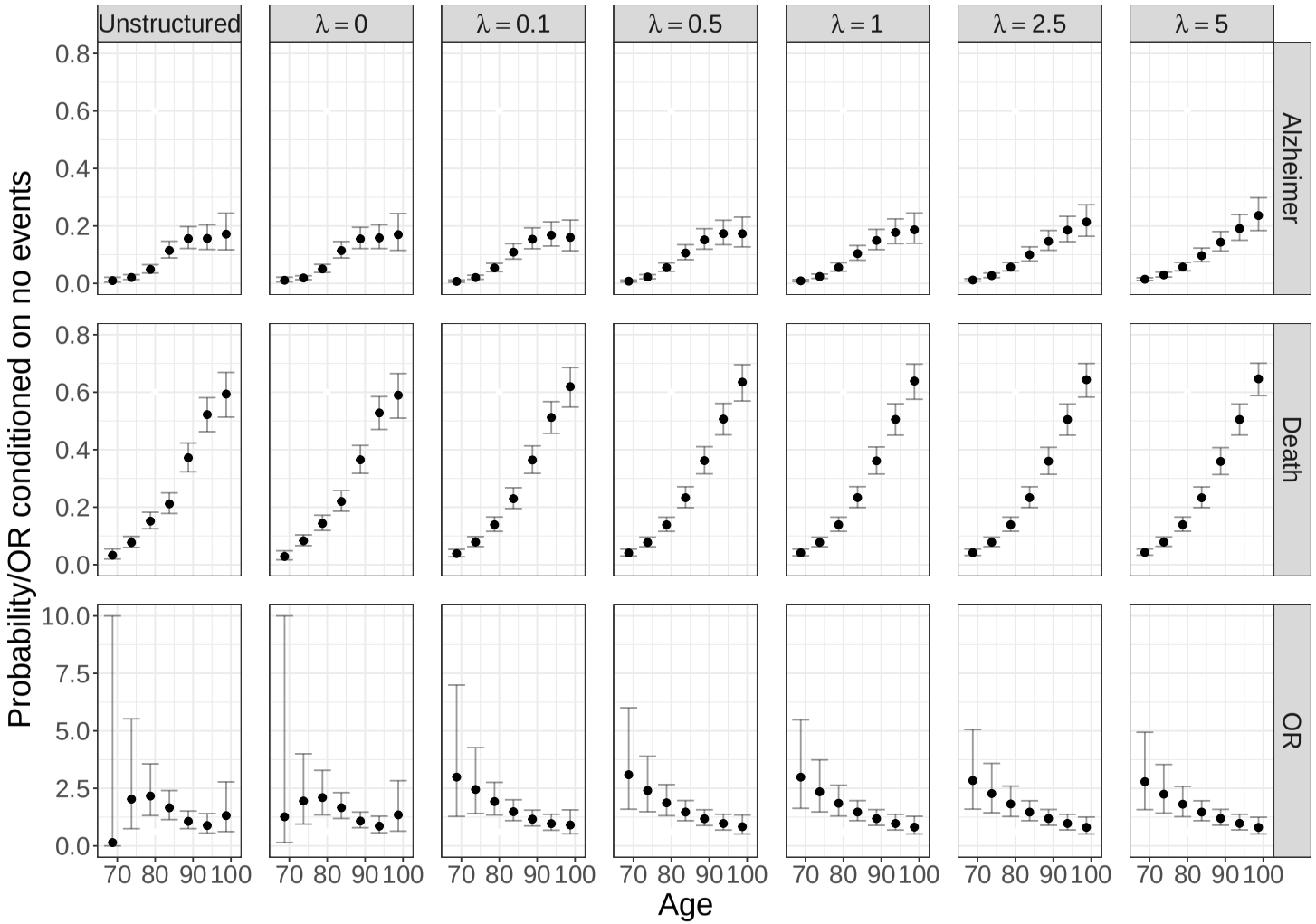


Figure A.11: Estimated baseline time trends, expit($\widehat{\boldsymbol{\alpha}}_1$), expit($\widehat{\boldsymbol{\alpha}}_2$) and exp($\widehat{\boldsymbol{\alpha}}_\theta$) from a series of analyses to the ACT data, under the unstructured specification and the B-spline specification with $\lambda \in \{0.0, 0.1, 0.5, 1, 2.5, 5\}$ and under the $\boldsymbol{\tau}^5$ partition. Also shown are 95% confidence intervals. Note, the y-axis has been truncated at 0.6 for expit($\widehat{\boldsymbol{\alpha}}_1$) and at 10 for exp($\widehat{\boldsymbol{\alpha}}_\theta$).

	Penalty, λ					
	0	0.1	0.5	1.0	2.5	5.0
Full model for $\theta_{i,k}$:						
Conservative						
$\boldsymbol{\tau}^{2.5}$	-18279	-18266	-18264	-18261	**-18260**	-18260
$\boldsymbol{\tau}^{5.0}$	**-12400**	-12408	-12417	-12424	-12436	-12448
Anti-conservative						
$\boldsymbol{\tau}^{2.5}$	-20340	-20334	-20333	-20331	**-20331**	-20331
$\boldsymbol{\tau}^{5.0}$	**-16418**	-16424	-16431	-16435	-16448	-16466
Nearest Neighbor						
$\boldsymbol{\tau}^{2.5}$	-19597	-19590	-19588	-19586	**-19586**	-19586
$\boldsymbol{\tau}^{5.0}$	**-14961**	-14965	-14969	-14973	-14984	-14997
Intermediate model for $\theta_{i,k}$:						
Conservative						
$\boldsymbol{\tau}^{2.5}$	-18270	-18257	-18254	-18252	**-18251**	-18251
$\boldsymbol{\tau}^{5.0}$	**-12397**	-12405	-12414	-12421	-12433	-12445
Anti-conservative						
$\boldsymbol{\tau}^{2.5}$	-20331	-20325	-20323	-20322	**-20322**	-20322
$\boldsymbol{\tau}^{5.0}$	**-16414**	-16419	-16426	-16431	-16444	-16461
Nearest Neighbor						
$\boldsymbol{\tau}^{2.5}$	-19589	-19581	-19578	-19577	**-19576**	-19577
$\boldsymbol{\tau}^{5.0}$	**-14958**	-14961	-14965	-14969	-14980	-14993
Reduced model for $\theta_{i,k}$:						
Conservative						
$\boldsymbol{\tau}^{2.5}$	-18267	-18254	-18251	-18249	**-18248**	-18248
$\boldsymbol{\tau}^{5.0}$	**-12394**	-12402	-12411	-12417	-12430	-12441
Anti-conservative						
$\boldsymbol{\tau}^{2.5}$	-20328	-20322	-20321	-20320	**-20319**	-20319
$\boldsymbol{\tau}^{5.0}$	**-16410**	-16416	-16422	-16427	-16440	-16458
Nearest Neighbor						
$\boldsymbol{\tau}^{2.5}$	-19585	-19578	-19576	-19574	**-19573**	-19574
$\boldsymbol{\tau}^{5.0}$	**-14954**	-14957	-14961	-14965	-14976	-14989

Table A.19: AIC values from the fit of the proposed framework to the ACT data, for different penalty levels under the two partitions considered and for all three approaches to overcome mid-interval censoring and truncation. See Section 6 of the main manuscript for details. The value indicated in **boldface** is the largest, with ties favoring models with lower penalty

	Alzheimer's Disease, $\pi_{1i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_1)$		**Death,** $\pi_{2i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_2)$		**Cross-sectional Odds Ratio,** $\theta_{i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_\theta)$	
	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$
Unstructured						
Race: White	1.07 (0.85, 1.35)	1.09 (0.85, 1.39)	1.12 (0.95, 1.32)	1.08 (0.90, 1.29)		
College graduate	0.84 (0.73, 0.97)	0.82 (0.70, 0.95)	0.91 (0.82, 1.01)	0.90 (0.80, 1.00)		
Marital status[d]:						
Never	1.03 (0.69, 1.53)	1.05 (0.69, 1.58)	1.17 (0.89, 1.54)	1.15 (0.85, 1.56)		
Divorced	0.99 (0.80, 1.23)	0.96 (0.77, 1.20)	1.25 (1.08, 1.45)	1.23 (1.04, 1.45)		
Widowed	0.98 (0.83, 1.15)	0.96 (0.81, 1.14)	1.12 (1.00, 1.26)	1.17 (1.02, 1.33)		
Other	0.84 (0.56, 1.27)	0.80 (0.52, 1.22)	1.19 (0.91, 1.56)	1.17 (0.87, 1.58)		
Depression	1.61 (1.29, 2.00)	1.59 (1.26, 2.01)	1.39 (1.18, 1.64)	1.51 (1.25, 1.81)		
Female	0.95 (0.81, 1.13)	1.03 (0.86, 1.23)	0.57 (0.50, 0.64)	0.56 (0.49, 0.64)	0.80 (0.52, 1.23)	0.90 (0.62, 1.31)
APOE[a]	1.87 (1.48, 2.37)	1.90 (1.49, 2.44)	0.96 (0.79, 1.16)	0.98 (0.80, 1.20)	0.59 (0.30, 1.18)	0.70 (0.40, 1.24)
Female × APOE[a]	1.00 (0.75, 1.35)	0.95 (0.70, 1.30)	1.13 (0.87, 1.45)	1.12 (0.86, 1.46)	1.21 (0.49, 2.98)	1.18 (0.57, 2.45)
AD[b]			2.71 (2.14, 3.42)	2.80 (2.02, 3.89)		
AD[b]× Female			1.50 (1.12, 2.01)	1.94 (1.27, 2.96)		
AD[b]× APOE[a]			1.56 (1.02, 2.38)	2.33 (1.24, 4.38)		
AD[b]× Female ×APOE[a]			0.66 (0.39, 1.13)	0.43 (0.20, 0.93)		
B-spline[c]						
Race: White	1.06 (0.84, 1.35)	1.09 (0.85, 1.40)	1.12 (0.94, 1.32)	1.08 (0.89, 1.30)		
College graduate	0.84 (0.73, 0.97)	0.82 (0.70, 0.95)	0.91 (0.83, 1.01)	0.90 (0.80, 1.00)		
Marital status[d]:						
Never	1.02 (0.68, 1.52)	1.05 (0.69, 1.59)	1.17 (0.89, 1.54)	1.15 (0.85, 1.57)		
Divorced	1.00 (0.80, 1.24)	0.96 (0.77, 1.20)	1.25 (1.08, 1.45)	1.23 (1.04, 1.44)		
Widowed	0.98 (0.83, 1.15)	0.96 (0.81, 1.14)	1.12 (0.99, 1.26)	1.17 (1.02, 1.33)		
Other	0.84 (0.56, 1.26)	0.80 (0.52, 1.22)	1.18 (0.90, 1.56)	1.17 (0.86, 1.59)		
Depression	1.61 (1.29, 2.00)	1.59 (1.26, 2.01)	1.40 (1.18, 1.65)	1.51 (1.24, 1.82)		
Female	0.95 (0.81, 1.12)	1.03 (0.87, 1.22)	0.57 (0.50, 0.64)	0.56 (0.49, 0.64)	0.79 (0.52, 1.21)	0.91 (0.62, 1.31)
APOE[a]	1.87 (1.48, 2.37)	1.91 (1.48, 2.45)	0.96 (0.80, 1.16)	0.98 (0.80, 1.19)	0.56 (0.28, 1.10)	0.70 (0.39, 1.26)
Female × APOE[a]	1.00 (0.74, 1.35)	0.95 (0.69, 1.30)	1.12 (0.87, 1.44)	1.12 (0.86, 1.47)	1.24 (0.51, 3.04)	1.18 (0.56, 2.47)
AD[b]			2.63 (2.07, 3.34)	2.80 (2.00, 3.92)		
AD[b]× Female			1.51 (1.12, 2.04)	1.94 (1.26, 2.99)		
AD[b]× APOE[a]			1.56 (1.00, 2.43)	2.34 (1.19, 4.58)		
AD[b]× Female ×APOE[a]			0.66 (0.38, 1.15)	0.42 (0.19, 0.96)		

[a] An indicator of having at least one APOE $\epsilon 4$ allele.
[b] An indicator of whether the patient has previously had a diagnosis of Alzheimer's disease
[c] Based on the optimal λ of λ^*=2.5 for $\boldsymbol{\tau}^{2.5}$ and λ^*=0.0 for $\boldsymbol{\tau}^{5.0}$.
[d] Reference level: married.

Table A.20: Coefficient results from the intermediate model based on the proposed framework applied to the data from the Adult Changes in Thought study.

	Alzheimer's Disease, $\pi_{1i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_1)$		**Death,** $\pi_{2i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_2)$		**Cross-sectional Odds Ratio,** $\theta_{i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_\theta)$	
	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$
Unstructured						
Race: White	1.07 (0.85, 1.35)	1.09 (0.85, 1.39)	1.12 (0.95, 1.32)	1.08 (0.90, 1.29)		
College graduate	0.84 (0.73, 0.97)	0.82 (0.70, 0.95)	0.91 (0.82, 1.01)	0.90 (0.80, 1.00)		
Marital status[d]:						
Never	1.03 (0.69, 1.53)	1.05 (0.69, 1.58)	1.17 (0.89, 1.54)	1.15 (0.85, 1.56)		
Divorced	0.99 (0.80, 1.23)	0.96 (0.77, 1.20)	1.25 (1.08, 1.45)	1.23 (1.04, 1.45)		
Widowed	0.98 (0.83, 1.15)	0.96 (0.81, 1.14)	1.12 (1.00, 1.26)	1.17 (1.02, 1.33)		
Other	0.85 (0.56, 1.27)	0.80 (0.52, 1.22)	1.19 (0.91, 1.56)	1.17 (0.87, 1.58)		
Depression	1.61 (1.29, 2.00)	1.59 (1.26, 2.01)	1.39 (1.18, 1.64)	1.51 (1.25, 1.82)		
Female	0.95 (0.81, 1.13)	1.03 (0.86, 1.23)	0.57 (0.50, 0.64)	0.56 (0.49, 0.64)		
APOE[a]	1.87 (1.48, 2.37)	1.91 (1.49, 2.44)	0.96 (0.79, 1.16)	0.98 (0.80, 1.20)	0.65 (0.41, 1.02)	0.77 (0.54, 1.11)
Female × APOE[a]	1.00 (0.75, 1.35)	0.95 (0.70, 1.30)	1.13 (0.87, 1.45)	1.12 (0.86, 1.46)		
AD[b]			2.71 (2.14, 3.42)	2.80 (2.02, 3.89)		
AD[b]× Female			1.50 (1.12, 2.01)	1.94 (1.27, 2.96)		
AD[b]× APOE[a]			1.56 (1.02, 2.38)	2.33 (1.24, 4.38)		
AD[b]× Female ×APOE[a]			0.66 (0.39, 1.13)	0.43 (0.20, 0.93)		
B-spline[c]						
Race: White	1.06 (0.84, 1.35)	1.09 (0.85, 1.40)	1.11 (0.94, 1.32)	1.08 (0.89, 1.30)		
College graduate	0.84 (0.73, 0.97)	0.82 (0.70, 0.95)	0.91 (0.83, 1.01)	0.90 (0.80, 1.00)		
Marital status[d]:						
Never	1.02 (0.68, 1.52)	1.05 (0.69, 1.59)	1.17 (0.89, 1.54)	1.15 (0.85, 1.57)		
Divorced	1.00 (0.80, 1.23)	0.96 (0.77, 1.20)	1.25 (1.08, 1.45)	1.23 (1.04, 1.44)		
Widowed	0.98 (0.83, 1.15)	0.96 (0.81, 1.14)	1.12 (0.99, 1.26)	1.17 (1.02, 1.33)		
Other	0.84 (0.56, 1.26)	0.80 (0.52, 1.23)	1.18 (0.90, 1.56)	1.17 (0.86, 1.59)		
Depression	1.61 (1.29, 2.00)	1.59 (1.26, 2.01)	1.40 (1.18, 1.65)	1.51 (1.25, 1.82)		
Female	0.95 (0.81, 1.12)	1.03 (0.87, 1.22)	0.57 (0.50, 0.64)	0.56 (0.49, 0.64)		
APOE[a]	1.87 (1.48, 2.37)	1.90 (1.48, 2.45)	0.96 (0.80, 1.16)	0.98 (0.80, 1.20)	0.62 (0.40, 0.97)	0.77 (0.54, 1.11)
Female × APOE[a]	1.00 (0.74, 1.35)	0.95 (0.69, 1.31)	1.12 (0.87, 1.44)	1.12 (0.86, 1.46)		
AD[b]			2.63 (2.07, 3.34)	2.81 (2.00, 3.93)		
AD[b]× Female			1.51 (1.12, 2.04)	1.93 (1.25, 2.98)		
AD[b]× APOE[a]			1.56 (1.00, 2.43)	2.32 (1.18, 4.55)		
AD[b]× Female ×APOE[a]			0.66 (0.38, 1.14)	0.43 (0.19, 0.97)		

[a] An indicator of having at least one APOE $\epsilon 4$ allele.
[b] An indicator of whether the patient has previously had a diagnosis of Alzheimer's disease
[c] Based on the optimal λ of λ^*=2.5 for $\boldsymbol{\tau}^{2.5}$ and λ^*=0.0 for $\boldsymbol{\tau}^{5.0}$.
[d] Reference level: married.

Table A.21: Coefficient results from the reduced model based on the proposed framework applied to the data from the Adult Changes in Thought study.

	Alzheimer's Disease, $\pi_{1i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_1)$		**Death,** $\pi_{2i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_2)$		**Cross-sectional Odds Ratio,** $\theta_{i,k}$ $\exp(\widehat{\boldsymbol{\beta}}_\theta)$	
	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$	$\boldsymbol{\tau}^{2.5}$	$\boldsymbol{\tau}^{5.0}$
Unstructured						
Race: White	1.07 (0.84, 1.35)	1.09 (0.85, 1.39)	1.12 (0.95, 1.32)	1.07 (0.90, 1.29)	1.47 (0.70, 3.11)	1.54 (0.85, 2.80)
College graduate	0.84 (0.72, 0.96)	0.81 (0.70, 0.95)	0.91 (0.82, 1.01)	0.90 (0.80, 1.00)	1.12 (0.75, 1.70)	1.13 (0.80, 1.61)
Marital status[d]:						
Never	1.03 (0.69, 1.53)	1.05 (0.70, 1.59)	1.17 (0.89, 1.54)	1.15 (0.85, 1.56)	1.43 (0.48, 4.29)	1.47 (0.57, 3.77)
Divorced	0.99 (0.80, 1.23)	0.96 (0.77, 1.20)	1.25 (1.08, 1.45)	1.23 (1.04, 1.45)	1.08 (0.59, 1.99)	1.25 (0.75, 2.10)
Widowed	0.98 (0.83, 1.15)	0.97 (0.81, 1.15)	1.12 (1.00, 1.26)	1.17 (1.02, 1.33)	0.95 (0.59, 1.52)	1.22 (0.82, 1.81)
Other	0.84 (0.56, 1.27)	0.79 (0.52, 1.21)	1.19 (0.91, 1.55)	1.17 (0.86, 1.57)	0.37 (0.08, 1.73)	0.27 (0.07, 1.03)
Depression	1.61 (1.30, 2.00)	1.59 (1.26, 2.01)	1.39 (1.18, 1.64)	1.50 (1.25, 1.81)	0.92 (0.51, 1.68)	0.98 (0.59, 1.64)
Female	0.95 (0.80, 1.13)	1.02 (0.86, 1.22)	0.57 (0.50, 0.64)	0.56 (0.49, 0.64)	0.82 (0.52, 1.32)	0.86 (0.57, 1.28)
APOE[a]	1.87 (1.48, 2.37)	1.90 (1.48, 2.44)	0.96 (0.79, 1.16)	0.98 (0.80, 1.19)	0.57 (0.28, 1.14)	0.68 (0.38, 1.20)
Female × APOE[a]	1.01 (0.75, 1.35)	0.96 (0.71, 1.31)	1.13 (0.87, 1.46)	1.12 (0.86, 1.47)	1.29 (0.52, 3.19)	1.26 (0.61, 2.63)
AD[b]			2.71 (2.14, 3.42)	2.80 (2.02, 3.89)		
AD[b]× Female			1.50 (1.12, 2.01)	1.94 (1.27, 2.96)		
AD[b]× APOE[a]			1.56 (1.02, 2.38)	2.33 (1.24, 4.38)		
AD[b]× Female ×APOE[a]			0.66 (0.39, 1.13)	0.42 (0.19, 0.92)		
B-spline[c]						
Race: White	1.06 (0.84, 1.34)	1.09 (0.85, 1.40)	1.11 (0.94, 1.32)	1.07 (0.89, 1.30)	1.47 (0.72, 3.00)	1.54 (0.85, 2.81)
College graduate	0.84 (0.73, 0.97)	0.81 (0.70, 0.95)	0.91 (0.83, 1.01)	0.90 (0.80, 1.00)	1.16 (0.77, 1.74)	1.13 (0.79, 1.60)
Marital status[d]:						
Never	1.02 (0.68, 1.52)	1.05 (0.70, 1.60)	1.17 (0.89, 1.53)	1.15 (0.85, 1.57)	1.21 (0.42, 3.55)	1.48 (0.58, 3.76)
Divorced	1.00 (0.81, 1.24)	0.96 (0.77, 1.21)	1.25 (1.08, 1.45)	1.23 (1.04, 1.45)	1.13 (0.61, 2.06)	1.27 (0.76, 2.12)
Widowed	0.98 (0.83, 1.15)	0.97 (0.82, 1.14)	1.12 (0.99, 1.26)	1.17 (1.02, 1.33)	0.94 (0.59, 1.49)	1.22 (0.83, 1.81)
Other	0.84 (0.56, 1.27)	0.79 (0.52, 1.21)	1.18 (0.90, 1.55)	1.17 (0.86, 1.58)	0.36 (0.07, 1.86)	0.28 (0.06, 1.23
Depression	1.61 (1.29, 2.01)	1.59 (1.26, 2.01)	1.40 (1.18, 1.65)	1.50 (1.24, 1.82)	0.98 (0.54, 1.75)	0.98 (0.59, 1.62)
Female	0.95 (0.80, 1.12)	1.02 (0.86, 1.22)	0.57 (0.50, 0.64)	0.56 (0.49, 0.64)	0.81 (0.51, 1.29)	0.86 (0.57, 1.28)
APOE[a]	1.87 (1.48, 2.37)	1.90 (1.48, 2.44)	0.96 (0.79, 1.16)	0.97 (0.80, 1.19)	0.54 (0.27, 1.07)	0.68 (0.38, 1.23)
Female × APOE[a]	1.01 (0.75, 1.35)	0.96 (0.70, 1.32)	1.12 (0.87, 1.45)	1.13 (0.87, 1.47)	1.31 (0.54, 3.22)	1.25 (0.59, 2.64)
AD[b]			2.62 (2.07, 3.33)	2.80 (2.00, 3.93)		
AD[b]× Female			1.51 (1.12, 2.04)	1.94 (1.26, 2.98)		
AD[b]× APOE[a]			1.56 (1.00, 2.44)	2.33 (1.19, 4.58)		
AD[b]× Female ×APOE[a]			0.66 (0.38, 1.14)	0.43 (0.19, 0.96)		

[a] An indicator of having at least one APOE $\epsilon 4$ allele.
[b] An indicator of whether the patient has previously had a diagnosis of Alzheimer's disease
[c] Based on the optimal λ of λ^*=1 for $\boldsymbol{\tau}^{2.5}$ and λ^*=0.0 for $\boldsymbol{\tau}^{5.0}$.
[d] Reference level: married.

Table A.22: Coefficient results from the full model based on the proposed framework applied to the data from the Adult Changes in Thought study.

Table A.23: For ACT dataset and 2.5 years interval partition, comparison of main coefficient results under the three approaches to deal with right censoring and left truncation.

		Conservative		Anti-conservative		Nearest Point	
		exp(β)	CI	exp(β)	CI	exp(β)	CI
Unstructured model							
AD	Female	0.98	(0.83, 1.17)	0.97	(0.82, 1.14)	0.95	(0.81, 1.13)
	APOE	1.86	(1.45, 2.37)	1.85	(1.46, 2.33)	1.87	(1.48, 2.37)
	Female×APOE	0.98	(0.72, 1.33)	1.03	(0.77, 1.37)	1.00	(0.75, 1.35)
Death	Female	0.58	(0.51, 0.65)	0.57	(0.50, 0.64)	0.57	(0.50, 0.64)
	APOE	0.97	(0.80, 1.18)	0.97	(0.81, 1.18)	0.96	(0.79, 1.16)
	Female×APOE	1.13	(0.87, 1.47)	1.10	(0.85, 1.41)	1.13	(0.87, 1.45)
	AD	2.63	(2.07, 3.34)	2.68	(2.13, 3.37)	2.71	(2.14, 3.42)
	AD×Female	1.51	(1.12, 2.04)	1.56	(1.17, 2.08)	1.50	(1.12, 2.01)
	AD×APOE	1.54	(1.00, 2.38)	1.68	(1.10, 2.56)	1.56	(1.02, 2.38)
	AD×Female×APOE	0.66	(0.38, 1.13)	0.61	(0.36, 1.04)	0.66	(0.39, 1.13)
OR	Female	0.78	(0.50, 1.20)	0.80	(0.52, 1.23)	0.80	(0.52, 1.23)
	APOE	0.60	(0.30, 1.21)	0.61	(0.31, 1.22)	0.59	(0.30, 1.18)
	Female×APOE	1.19	(0.48, 2.96)	1.18	(0.48, 2.90)	1.21	(0.49, 2.98)
B-spline model							
AD	Female	0.98	(0.82, 1.16)	0.97	(0.82, 1.14)	0.95	(0.81, 1.12)
	APOE	1.85	(1.44, 2.36)	1.84	(1.45, 2.32)	1.87	(1.48, 2.37)
	Female×APOE	0.98	(0.72, 1.34)	1.03	(0.77, 1.38)	1.00	(0.74, 1.35)
Death	Female	0.58	(0.51, 0.65)	0.57	(0.51, 0.64)	0.57	(0.50, 0.64)
	APOE	0.97	(0.80, 1.18)	0.97	(0.81, 1.17)	0.96	(0.80, 1.16)
	Female×APOE	1.13	(0.87, 1.46)	1.10	(0.85, 1.41)	1.12	(0.87, 1.44)
	AD	2.54	(1.99, 3.24)	2.61	(2.06, 3.31)	2.63	(2.07, 3.34)
	AD×Female	1.52	(1.12, 2.06)	1.56	(1.16, 2.09)	1.51	(1.12, 2.04)
	AD×APOE	1.55	(0.99, 2.44)	1.69	(1.09, 2.62)	1.56	(1.00, 2.43)
	AD×Female×APOE	0.66	(0.38, 1.14)	0.62	(0.36, 1.06)	0.66	(0.38, 1.15)
OR	Female	0.76	(0.50, 1.17)	0.78	(0.51, 1.19)	0.79	(0.52, 1.21)
	APOE	0.57	(0.28, 1.13)	0.58	(0.29, 1.14)	0.56	(0.28, 1.10)
	Female×APOE	1.24	(0.50, 3.04)	1.20	(0.49, 2.91)	1.24	(0.51, 3.04)

Table A.24: For ACT dataset and 5 years interval partition, comparison of main coefficient results under the three approaches to deal with right censoring and left truncation.

		Conservative		Anti-conservative		Nearest Point	
		exp(β)	CI	exp(β)	CI	exp(β)	CI
Unstructured model							
AD	Female	1.13	(0.93, 1.37)	0.99	(0.83, 1.17)	1.03	(0.86, 1.23)
	APOE	1.96	(1.49, 2.58)	1.83	(1.44, 2.34)	1.90	(1.49, 2.44)
	Female×APOE	0.91	(0.65, 1.28)	1.03	(0.76, 1.40)	0.95	(0.70, 1.30)
Death	Female	0.55	(0.48, 0.64)	0.56	(0.49, 0.64)	0.56	(0.49, 0.64)
	APOE	0.98	(0.78, 1.22)	0.98	(0.81, 1.19)	0.98	(0.80, 1.20)
	Female×APOE	1.19	(0.89, 1.59)	1.12	(0.87, 1.45)	1.12	(0.86, 1.46)
	AD	2.70	(1.89, 3.84)	2.77	(2.03, 3.77)	2.80	(2.02, 3.89)
	AD×Female	1.95	(1.24, 3.07)	1.82	(1.23, 2.70)	1.94	(1.27, 2.96)
	AD×APOE	2.00	(1.04, 3.86)	2.33	(1.28, 4.22)	2.33	(1.24, 4.38)
	AD×Female×APOE	0.55	(0.24, 1.25)	0.49	(0.23, 1.01)	0.43	(0.20, 0.93)
OR	Female	0.82	(0.55, 1.23)	0.96	(0.67, 1.39)	0.90	(0.62, 1.31)
	APOE	0.73	(0.40, 1.33)	0.78	(0.45, 1.37)	0.70	(0.40, 1.24)
	Female×APOE	1.16	(0.54, 2.49)	1.01	(0.49, 2.07)	1.18	(0.57, 2.45)
B-spline model							
AD	Female	1.13	(0.94, 1.36)	0.99	(0.83, 1.17)	1.03	(0.87, 1.22)
	APOE	1.96	(1.48, 2.58)	1.83	(1.44, 2.34)	1.91	(1.48, 2.45)
	Female×APOE	0.91	(0.64, 1.29)	1.03	(0.76, 1.41)	0.95	(0.69, 1.30)
Death	Female	0.55	(0.48, 0.64)	0.55	(0.48, 0.62)	0.56	(0.49, 0.64)
	APOE	0.97	(0.78, 1.22)	0.98	(0.81, 1.20)	0.98	(0.80, 1.19)
	Female×APOE	1.19	(0.89, 1.59)	1.13	(0.87, 1.46)	1.12	(0.86, 1.47)
	AD	2.69	(1.88, 3.87)	2.77	(2.01, 3.82)	2.80	(2.00, 3.92)
	AD×Female	1.95	(1.22, 3.11)	1.82	(1.21, 2.73)	1.94	(1.26, 2.99)
	AD×APOE	2.01	(0.99, 4.06)	2.33	(1.24, 4.41)	2.34	(1.19, 4.58)
	AD×Female×APOE	0.55	(0.23, 1.29)	0.49	(0.23, 1.05)	0.42	(0.19, 0.96)
OR	Female	0.82	(0.55, 1.22)	0.96	(0.67, 1.38)	0.91	(0.62, 1.31)
	APOE	0.72	(0.39, 1.34)	0.78	(0.44, 1.38)	0.70	(0.39, 1.26)
	Female×APOE	1.16	(0.53, 2.53)	1.01	(0.49, 2.09)	1.18	(0.56, 2.47)

References

Alvares, D., S. Haneuse, C. Lee, and K. H. Lee (2019). Semicomprisks: An R package for the analysis of independent and cluster-correlated semi-competing risks data. *The R Journal 11*(1), 376–400.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65*(1), 141–151.

Egleston, B. L., D. O. Scharfstein, E. E. Freeman, and S. K. West (2006). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics 8*(3), 526–545.

Fine, J. P., H. Jiang, and R. Chappell (2001). On semi-competing risks data. *Biometrika 88*(4), 907–919.

Hsieh, J.-J., W. Wang, and A. Adam Ding (2008). Regression analysis based on semicompeting risks data. *JRSS-B 70*(1), 3–20.

Lee, K. H., S. Haneuse, D. Schrag, and F. Dominici (2015). Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *JRSS-C 64*(2), 253–273.

Lee, K. H., V. Rondeau, and S. Haneuse (2017). Accelerated failure time models for semicompeting risks data in the presence of complex censoring. *Biometrics 73*(4), 1401–1412.

Li, R. and L. Peng (2015). Quantile regression adjusting for dependent censoring from semicompeting risks. *JRSS-B 77*(1), 107–130.

Peng, L. and J. P. Fine (2007). Regression modeling of semicompeting risks data. *Biometrics 63*(1), 96–108.

Tchetgen Tchetgen, E. J. (2014). Identification and estimation of survivor average causal effects. *Statistics in medicine 33*(21), 3601–3628.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Xu, J., J. D. Kalbfleisch, and B. Tai (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics 66*(3), 716–725.

Yang, J. and L. Peng (2016). A new flexible dependence measure for semi-competing risks. *Biometrics 72*(3), 770–779.

Zhang, J. L. and D. B. Rubin (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics 28*(4), 353–368.