



Published in final edited form as:

Biometrics. 2022 September ; 78(3): 922–936. doi:10.1111/biom.13480.

Modeling semi-competing risks data as a longitudinal bivariate process

Daniel Nevo¹, Deborah Blacker^{2,3}, Eric B. Larson⁴, Sebastien Haneuse⁵

¹ Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel

² Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA

³ Department of Psychiatry, Massachusetts General Hospital & Harvard Medical School, Boston, Massachusetts, USA

⁴ Kaiser Permanente Washington Health Research Institute, Seattle, Washington, USA

⁵ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

Abstract

As individuals age, death is a competing risk for Alzheimer's disease (AD) but the reverse is not the case. As such, studies of AD can be placed within the semi-competing risks framework. Central to semi-competing risks, and in contrast to standard competing risks, is that one can learn about the dependence structure between the two events. To-date, however, most methods for semi-competing risks treat dependence as a nuisance and not a potential source of new clinical knowledge. We propose a novel regression-based framework that views the two time-to-event outcomes through the lens of a longitudinal bivariate process on a partition of the time scales of the two events. A key innovation of the framework is that dependence is represented in two distinct forms, *local* and *global* dependence, both of which have intuitive clinical interpretations. Estimation and inference are performed via penalized maximum likelihood, and can accommodate right censoring, left truncation, and time-varying covariates. An important consequence of the partitioning of the time scale is that an ambiguity regarding the specific form of the likelihood contribution may arise; a strategy for sensitivity analyses regarding this issue is described. The framework is then used to investigate the role of gender and having $\epsilon 1$ apolipoprotein E (APOE) $\epsilon 4$ allele on the joint risk of AD and death using data from the Adult Changes in Thought study.

Correspondence Daniel Nevo, Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, 69978, Israel. danielnevo@gmail.com.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 1–6 are available with this paper at the Biometrics website on Wiley Online Library. All simulations were carried out using code available in the `LongitSemiComp` package for R.

Simulation code, seeds and results are all available online via the Github repository of the first author (<https://github.com/daniel258/CausalSemiCompReproduce>).

Keywords

alzheimer's disease; b-splines; discrete-time survival; longitudinal modeling; penalized maximum likelihood

1 | INTRODUCTION

Alzheimer's disease (AD) is a brain disorder characterized by progressive dementia that slowly destroys memory and cognitive function. In 2018, an estimated 5.7 Americans were living with AD (Alzheimer's Association, 2018). First described in 1906, factors that have been shown to be associated with AD include: age, family history, the apolipoprotein E (APOE) $\epsilon 4$ allele, midlife obesity, midlife hypertension, diabetes, education, and physical activity (Baumgart et al., 2015). Many of these factors are also strongly associated with mortality, suggesting that AD and mortality may be dependent within individuals and, furthermore, that this dependence may be influenced by a range of factors.

Practically, studies of risk factors for AD often focus on the timing of a diagnosis and thus use survival analysis methods. Such analyses typically treat death as a censoring mechanism. An alternative is the semi-competing risks paradigm, within which AD and mortality are considered *simultaneously*. In general terms, semi-competing risks refer to settings where interest lies in some nonterminal event, the occurrence of which is subject to a terminal event (Fine et al., 2001; Varadhan et al., 2014; Haneuse and Lee, 2016). Let T_1 and T_2 denote the time to the nonterminal and terminal events, respectively. Key to semi-competing risks is that one can potentially observe *both* T_1 and T_2 on individual study units. As such, in contrast to the standard competing risks setting (Tsiatis, 1975), there is partial information on the joint distribution of the two events (Jazi et al., 2016). This, in turn, provides an opportunity to learn about the dependence structure between T_1 and T_2 .

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 . In the statistical literature, numerous frameworks for the analysis of semi-competing risks data have been proposed, including: methods grounded in causal inference (Egleston et al., 2006; Tchetgen Tchetgen, 2014; Nevo and Gorfine, 2020); methods based on structuring dependence via a copula (Fine et al., 2001; Peng and Fine, 2007; Li and Peng, 2015); the use of illness-death models (Xu et al., 2010; Lee et al., 2017, 2015); and the recently proposed cross-quantile residual ratio (Yang and Peng, 2016). While additional review details are provided in Section A.1 of the Supporting Information, we note that these methods either: (i) view dependence as a statistical nuisance, and not a potential source of new clinical knowledge; or (ii) focus on the role of the nonterminal event as a risk factor for the terminal event, thereby reframing the nonterminal event away from being an outcome of interest. As such, collectively, these methods fail to take advantage of the opportunity to learn about dependence between the two outcomes that semi-competing risks data provide.

In this paper, we propose a novel regression-based framework for semi-competing risks data based on a partitioning of the time scales for the two events that simultaneously structures covariate effects on T_1 , T_2 , and the dependence between the two. A key innovation of the proposed framework is that dependence is represented in two distinct forms, termed the *local* and *global* dependence, both of which have intuitive clinical interpretations. Practically, since modeling within the framework is based upon probabilities conditional on survival, left truncation, right censoring, and time-varying covariates may be accommodated in a straightforward manner. An important consequence of the partitioning of the time scale is that an ambiguity regarding the specific form of the likelihood contribution may arise; a strategy for sensitivity analyses regarding this issue is described. Finally, in addition to a series of simulation studies, we illustrate the approach with a case study investigating the complex interplay between gender, APOE, AD diagnosis, and death using data from the Adult Changes in Thought (ACT) study (Kukull et al., 2002).

2 | A LONGITUDINAL BIVARIATE MODELING FRAMEWORK

This paper proposes a novel longitudinal bivariate modeling framework for semi-competing risks within which dependence between T_1 and T_2 is characterized in a meaningful and interpretable way, and is permitted to be a function of covariates. Prior to describing the framework, we note how semi-competing risks data are typically represented: letting L and C denote study entry and right censoring times, respectively, the observed outcome data for the i th study participant is $\mathcal{D}_i = (L, \tilde{T}_{1i}, \delta_{1i}, \tilde{T}_{2i}, \delta_{2i})$, where $\tilde{T}_{1i} = \min(T_{1i}, T_{2i}, C_i)$ and $\tilde{T}_{2i} = \min(T_{2i}, C_i)$. Furthermore, $\delta_{1i} = I\{\tilde{T}_{1i} = T_{1i}\}$ and $\delta_{2i} = I\{\tilde{T}_{2i} = T_{2i}\}$ indicate whether the nonterminal and terminal events, respectively, are observed.

2.1 | A novel representation of semi-competing risks data

The capacity of the proposed framework to investigate dependence between T_1 and T_2 hinges on a novel representation of semi-competing risks data that follows from a discretization or partitioning of the analysis time scale. Toward this, let $\tau = \{\tau_0, \dots, \tau_K\}$ be a set of user-specified points, with $\tau_{k-1} < \tau_k$ for $k = 1, \dots, K$, that define a partition. For example, in analyzing data from ACT one could choose age beyond 65 years as the time scale and $\tau = \{65, 70, 75, \dots, 100\}$. As will become clear, the specification of τ has important implications, and we return to this task in Section 3.

Given τ , one can query whether the nonterminal and/or the terminal event was observed to occur within or by the end of any given interval. Toward this, let $Y_{1i,k} = I\{T_{1i} \leq \tau_k\}$ and $Y_{2i,k} = I\{T_{2i} \leq \tau_k\}$ be indicators of whether the i th study participant experiences the nonterminal and terminal events by time τ_k , respectively. In practice, depending on when a study participant contributes person-time on the analysis time scale, they may or may not provide direct information about risk during each of the K intervals defined by τ . For example, an ACT participant enrolled at age 74 years cannot provide direct information about the $(65, 70]$ interval. Thus, the observed outcome data for the i th study participant on partition τ will consist of a longitudinal bivariate process on a subset of the K

intervals, specifically $\mathbf{Y}_i = \{(Y_{1i,k}, Y_{2i,k}); k = k_i^l, \dots, k_i^r\}$, where $k_i^l, k_i^r \in \{1, \dots, K\}$ and $k_i^l \leq k_i^r$. The determination of k_i^l and k_i^r represents a nontrivial task, and we return to it in Section 3.

Finally, let $\mathbf{X}_i(\tau_k)$ denote a vector of possibly time-varying covariates measured at time τ_{k-1} , and $\mathbf{X}_{i,k}^H = (\mathbf{X}_i(\tau_1), \dots, \mathbf{X}_i(\tau_k))$ the history of all such information at time τ_{k-1} .

2.2 | The joint distribution of the observed data

Under the novel representation, we assume that the joint distribution of the observed outcome data, $P(\mathbf{Y}_i = \mathbf{y}_i \mid \mathbf{X}_{i,k_i^r}^H)$, can be decomposed as the product:

$$\prod_{k=k_i^l}^{k_i^r} P(Y_{1i,k} = y_{1i,k}, Y_{2i,k} = y_{2i,k} \mid Y_{1i,k-1} = y_{1i,k-1}, Y_{2i,k-1} = y_{2i,k-1}, \mathbf{X}_{i,k}^H), \quad (1)$$

$y_{1i,k} \in \{0, 1\}$ and $y_{2i,k} \in \{0, 1\}$, and where we define $Y_{1i,0} = Y_{2i,0} = 0$. Underpinning this decomposition is a Markovtype assumption, specifically that, conditional on the totality of the covariate information to that point, the joint probability of the two events in a given interval depends on the history of the two events solely through the status at the start of the interval. As shown in Section A.2 of the Supporting Information, the components of expression (1) can be written in terms of:

$$\pi_{1i,k} = P(Y_{1i,k} = 1 \mid Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \mathbf{X}_{i,k}^H), \quad (2)$$

$$\pi_{2i,k}(y_1) = P(Y_{2i,k} = 1 \mid Y_{1i,k-1} = y_1, Y_{2i,k-1} = 0, \mathbf{X}_{i,k}^H), \quad (3)$$

$$\theta_{i,k} = \text{OR}(Y_{1i,k}, Y_{2i,k} \mid Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \mathbf{X}_{i,k}^H) \quad (4)$$

for $y_1 \in \{0, 1\}$ and where

$$\begin{aligned} & \text{OR}(Y_{1i,k}, Y_{2i,k} \mid \dots) \\ &= \frac{P(Y_{1i,k} = 1, Y_{2i,k} = 1 \mid \dots) \times P(Y_{1i,k} = 0, Y_{2i,k} = 0 \mid \dots)}{P(Y_{1i,k} = 0, Y_{2i,k} = 1 \mid \dots) \times P(Y_{1i,k} = 1, Y_{2i,k} = 0 \mid \dots)}. \end{aligned}$$

We emphasize that the interpretations of these three quantities are specific to the partition τ : $\pi_{1i,k}$ is the cumulative probability of experiencing the nonterminal event by time τ_k , given that neither event had occurred by time τ_{k-1} ; $\pi_{2i,k}(\cdot)$ is the cumulative probability of experiencing the terminal event by time τ_k , given that the individual is alive at time

τ_{k-1} ; finally, $\theta_{i,k}$ is the cross-sectional odds ratio for the 2×2 table corresponding to the four possible observed $(Y_{1i,k}, Y_{2i,k})$ outcome vectors at time τ_k , given that neither event had occurred by time τ_{k-1} . Note that both $\pi_{1i,k}$ and $\pi_{2i,k}$ are quantities that are modeled in discrete time survival analyses (Prentice and Gloeckler, 1978).

2.3 | Regression structure

We proceed by placing regression structure on $\pi_{1i,k}$, $\pi_{2i,k}(y_1)$ and $\theta_{i,k}$, specifically:

$$\pi_{1i,k} = g_1^{-1}\{\alpha_{1,k} + f_1(\mathbf{X}_{i1,k}^H; \boldsymbol{\beta}_1)\}, \quad (5)$$

$$\pi_{2i,k}(y_1) = g_2^{-1}\{\alpha_{2,k} + f_2(\mathbf{X}_{i2,k}^H, y_1; \boldsymbol{\beta}_2)\}, \quad (6)$$

$$\theta_{i,k} = g_\theta^{-1}\{\alpha_{\theta,k} + f_\theta(\mathbf{X}_{i\theta,k}^H; \boldsymbol{\beta}_\theta)\}, \quad (7)$$

where $g_1(\cdot)$, $g_2(\cdot)$, and $g_\theta(\cdot)$ are user-specified link functions (e.g., the logistic or log link); $\mathbf{X}_{i1,k}^H$, $\mathbf{X}_{i2,k}^H$, and $\mathbf{X}_{i\theta,k}^H$ are each subsets of $\mathbf{X}_{i,k}^H$; and, $f_1(\cdot)$, $f_2(\cdot)$, and $f_\theta(\cdot)$ are user-specified functions that characterize how the respective quantities depend on the covariates. For example, one may adopt a linear specification with no interactions between the components of $\mathbf{X}_{i2,k}^H$ and $Y_{1i,k-1}$ in model (6) by setting $f_2(\mathbf{X}_{i2,k}^H, Y_{1i,k-1}; \boldsymbol{\beta}_2) \equiv (\mathbf{X}_{i2,k}^H)^T \boldsymbol{\beta}_{2,X} + Y_{1i,k-1} \boldsymbol{\beta}_{2,Y}$.

While their precise interpretations will depend on the partition of the analysis time scale and the chosen link functions, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_\theta)$ characterize the impact of covariates; Section 2.4 considers the role and interpretation of components of $\boldsymbol{\beta}_2$ that correspond to $Y_{1i,k-1}$. Finally, $\boldsymbol{\alpha}_1 = (\alpha_{1,1}, \dots, \alpha_{1,K})$, $\boldsymbol{\alpha}_2 = (\alpha_{2,1}, \dots, \alpha_{2,K})$, and $\boldsymbol{\alpha}_\theta = (\alpha_{\theta,1}, \dots, \alpha_{\theta,K})$ represent baseline time trends in $\pi_{1i,k}$, $\pi_{2i,k}(\cdot)$, and $\theta_{i,k}$, respectively, with their precise interpretation again depending on the choice of τ , the link functions and covariates included in the models.

From a practical perspective, that $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_\theta)$ consists of $3K$ parameters may result in computational and/or convergence issues unless the observed data are rich (i.e., a large sample size or high event rates) or the initial partition is coarse. To mitigate such issues, we propose that a B-spline structure be adopted across the K components of each of $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, and $\boldsymbol{\alpha}_\theta$ (Eilers and Marx, 1996). Focusing on $\boldsymbol{\alpha}_1$, let $\{\tilde{\tau}_1^1, \dots, \tilde{\tau}_1^{J_1}\}$ denote a collection of J_1 user-specified knots on the interior of the range (τ_0, τ_K) and q_1 the user-specified degree of the local polynomial basis functions. Given these choices, we specify the k th component of $\boldsymbol{\alpha}_1$ as $\alpha_{1,k} = \sum_{j=1}^{\tilde{J}_1} \eta_{1,j} B_{j,k}$, where $B_{j,k}$ is the value of the j th B-spline at time τ_k and $\tilde{J}_1 = J_1 - 1 + q_1$ is total number of spline terms. Thus, this specification requires estimation of \tilde{J}_1 coefficient terms, specifically $\boldsymbol{\eta}_1 = (\eta_{1,1}, \dots, \eta_{1,\tilde{J}_1})$. Applying the same strategy to $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_\theta$

will result in $\tilde{J}_1 + \tilde{J}_2 + \tilde{J}_\theta$ unknown coefficients, collectively denoted as $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_\theta)$. Finally, to distinguish between specifications, we refer to the model based on $\boldsymbol{\alpha}$ with $3K$ unknowns as the *unstructured model* while that based on $\boldsymbol{\eta}$ as the *B-spline model*.

2.4 | Dependence

The central innovation of the framework in expressions (5)–(7) is that dependence between the nonterminal and terminal events is quantified in two distinct and yet complementary ways. The first is captured through $\theta_{i,k}$ which can be viewed as a measure of *local dependence* in that it quantifies the risk of co-occurrence of the two events, via the odds ratio (Lipsitz et al., 1991; Carey et al., 1993; Ten Have and Morabia, 1999), during the $(\tau_{k-1}, \tau_k]$ interval; a large positive value of $\theta_{i,k}$ indicates that if the nonterminal event (e.g., AD) occurs during $(\tau_{k-1}, \tau_k]$ then the terminal event (e.g., death) is likely to subsequently occur during the same interval. Because of the novel parameterization, we emphasize that the local dependence can vary over the time scale or as a function of covariates. As such, one could investigate whether local dependence between AD and death is weaker at younger ages relative to older ages. Furthermore, one could investigate whether the probability that the two events co-occur changes in response to key life events such as the death of a partner. The second quantification of dependence is through the components of $\boldsymbol{\beta}_2$ in expression (6) that correspond to how the status of the nonterminal event, $Y_{1i,k-1}$, influences $\pi_{2i,k}$. Labeling these components as $\beta_{2,y}$, analogous to how one interprets covariates effects in regression models, these parameters capture the extent to which whether the non-terminal event has occurred is associated with a change in risk of the terminal event; for this reason, we refer to $\beta_{2,y}$ as capturing *global dependence*. Note that global dependence is conceptually similar to the explanatory hazard ratio and cross-quantile residual ratio (see section A.1 of the Supporting Information) in that it concerns the role that the nonterminal event plays in modifying risk of the terminal event.

3 | KEY PRACTICAL TASKS

The development in Section 2 indicated two practical tasks that analysts must engage in, specifically the choice of $\boldsymbol{\tau}$ and the determination of k'_j and k'_i .

3.1 | Choice of partition, $\boldsymbol{\tau}$

The partition, $\boldsymbol{\tau}$, plays a critical role in the proposed framework in that it provides the foundation for being able to distinguish between local and global dependence, as we conceptualize them in Section 2.4, and for being able to investigate the role that covariates play. In Section A.2 of the Supporting Information, we show that the underlying distribution for (T_1, T_2) induces the quantities (2)–(4) for *any* $\boldsymbol{\tau}$. Thus, regardless of the choice of $\boldsymbol{\tau}$, the components of the proposed model are well-defined mathematical objects and are, therefore, valid targets for estimation and inference. An important consequence of this is that one cannot say that any given partition corresponds to the “truth.” As indicated in Section 2.2, however, the choice of $\boldsymbol{\tau}$ dictates the numerical values and interpretation of the quantities given by (2)–(4), and correspondingly (in part, at least) the numerical values and

interpretation of the parameters in the regression structure given by expressions (5)–(7). As such the choice of τ requires careful consideration.

In principle, one could approach choosing τ through consideration of the clinical condition under investigation such as the pace at which the disease progresses. In the ACT study, for example, the choice to schedule follow-up biennially was made in part for logistical considerations but also because AD is a slowly-developing condition. Alternatively, one may pursue a data-driven strategy where, for example, some goodness-of-fit criterion is specified and then optimized as a function of τ . Our perspective is that the decision should be based primarily, if not exclusively, on clinical considerations. Central to this position is that, in addition to their interpretation, the numerical values of (2)–(4) change with τ . To see this, we again note that the probabilities in each of (2)–(4) speak to the cumulative incidence of events during the interval $(\tau_{k-1}, \tau_k]$. If the length of the interval is decreased then the incidence will necessarily decrease. The corresponding new interpretation and numerical value will not be “wrong,” however, but just different. Put another way, the change in the interpretation and the numerical values of the model parameters that result from, say, adopting a finer partition of the time scale should, arguably, be viewed as a change in the question that is being answered. Thus, in our view, purely data-driven approaches, while they may have some initial intuitive appeal, should be avoided. Nevertheless, we do acknowledge that it may not be easy to elicit a single partition, from the literature or collaborators, on which to base the analyses and conclusions. If it is the case that there is no clear choice, analysts may opt to perform a range of analyses over different partitions, both in terms of how fine the partition is and in terms of where the cut-points are for a given (common) interval length. We pursue this strategy in conducting the analyses of the data from ACT in Section 6.

3.2 | Determination of k_i^l and k_i^r

While the act of partitioning the analysis time scale provides the basis for learning about dependence, there is a trade-off in that there is a loss of information that impacts how one approaches setting k_i^l and k_i^r . To illustrate this, and the framework more generally, Figure 1 provides a graphical representation of observed outcome data for two hypothetical study participants; an expanded figure, with five hypothetical participants, is provided in Section A.2 of the Supporting Information. From the top half of the figure, we see that the first participant was observed to enter the study in the $(\tau_1, \tau_2]$ interval, experience the nonterminal event in the $(\tau_3, \tau_4]$ interval and subsequently experience the terminal event in $(\tau_4, \tau_5]$. From the bottom half of the figure, the second participant was observed to enter the study in the $(\tau_2, \tau_3]$ interval and subsequently experience the nonterminal event in $(\tau_4, \tau_5]$. In contrast to participant $i = 1$, however, they are censored prior to being observed to experience the terminal event, specifically in $(\tau_5, \tau_6]$.

In considering the value of k_i^r , it seems clear that setting $k_i^r = 5$ for the first participant is appropriate, so that their outcome data in the $(\tau_4, \tau_5]$ interval are $(Y_{1i,5}, Y_{2i,5}) = (1, 1)$. For the second participant, one could follow suit and set $k_i^r = 6$ so that, acknowledging that they had previously been observed to experience the nonterminal event but not the terminal event,

their outcome data in the $(\tau_5, \tau_6]$ interval would be $(Y_{1i,6}, Y_{2i,6}) = (1, 0)$. Doing so, however, assumes that the terminal event did not occur in the subinterval $(C_i, \tau_6]$. Alternatively, one can interpret the choice as assuming complete person-time in the $(\tau_5, \tau_6]$ interval when this was not, in fact, the case. With this latter interpretation, there is an argument to be made that one should not attempt to learn about expressions (2)–(4) for $(\tau_5, \tau_6]$ and that, instead, one should adopt a more conservative stance by setting $k_i^r = 5$ and forgoing the partial information regarding this particular interval.

Figure 1 also helps illustrate similar considerations regarding k_i^l . Specifically, in considering which of the intervals the first participant can be said to first contribute to, since $L_i \in (\tau_1, \tau_2]$ setting $k_i^l = 2$ seems a natural choice and, coupled with $k_i^r = 5$, would result in observed outcome data in four intervals, specifically $Y_i = \{(0, 0), (0, 0), (1, 0), (1, 1)\}$. Problematic with setting $k_i^l = 2$, however, is that complete person-time for the $(\tau_1, \tau_2]$ interval was not observed since $L_i > \tau_1$. As such, there is, arguably, incomplete data regarding the quantities given by expressions (2)–(4) for this particular interval and a more conservative approach would be set $k_i^l = 3$ so that, again coupled with $k_i^r = 5$, the observed outcome data would be restricted to three intervals, specifically $Y_i = \{(0, 0), (1, 0), (1, 1)\}$, and that the partial information in $(\tau_1, \tau_2]$ is ignored.

To summarize, in setting both k_i^l and k_i^r analysts have an option that can be viewed as being *conservative*, in which partial information is ignored. In contrast, analysts have the option to be *anticonservative* by including the intervals with partial person-time as if they had complete person-time. One common feature of both of these strategies is that they are agnostic to the amount of partial information; that is how close L_i and C_i are to the right-hand limit of the interval in which they were observed. We therefore propose a third strategy, labeled the *nearest neighbor* strategy in which the partial information is acknowledged. Focusing initially on k_i^l , suppose $L_i \in (\tau_{k-1}, \tau_k]$. While the anti-conservative and conservative strategies would always set $k_i^l = \tau_k$ and $k_i^l = \tau_{k+1}$, respectively, the nearest neighbor strategy takes $k_i^l = \tau_k$ if $|\tau_k - L_i| > |\tau_k - \tau_{k-1}|/2$ and $k_i^l = \tau_{k+1}$ otherwise. Similarly, suppose $C_i \in (\tau_{k-1}, \tau_k]$. While the anticonservative and conservative strategies would always set $k_i^r = \tau_k$ and $k_i^r = \tau_{k-1}$, respectively, the nearest neighbor strategy takes $k_i^r = \tau_k$ if $|\tau_k - C_i| \leq |\tau_k - \tau_{k-1}|/2$ and $k_i^r = \tau_{k-1}$ otherwise.

To highlight the impact adopting each of the three proposed strategies have, Figure 1 provides the corresponding outcome information for the two hypothetical study participants. In practice, analysts will need to base the decision of which strategy to use by balancing the information gain/loss with their own conservatism regarding taking person-time to be complete in an interval when it is not. Since the nearest neighbor strategy represents, in a sense, a compromise, it is the primary strategy that we would recommend. Of course, analysts may consider other strategies or, as we do in Section 6, consider a range of strategies and examine their impact. Finally, we note that the extent to which this is a concern depends, in part at least, on how coarse the partition is; with finer and finer partitions, the information loss associated with the conservative strategy will decrease.

4 | ESTIMATION AND INFERENCE

4.1 | The observed data likelihood

Building on the notation developed in Section 2, the first two columns of Table 1 provide the six possible outcome data scenarios in the k th interval of the partition given by τ as a function of the outcome vector in the previous interval (i.e., as a function of $(Y_{1i,k-1}, Y_{2i,k-1})$). The third column provides the corresponding likelihood contributions, that is the interval-specific components in the decomposition given by expression (1), with

$$\begin{aligned} & \pi_{12i,k} \\ &= P(Y_{1i,k} = 1, Y_{2i,k} = 1 \mid Y_{1i,k-1} = 0, Y_{2i,k-1} = 0, \mathbf{X}_{i,k}^H) \\ &= \begin{cases} \pi_{1i,k} \pi_{2i,k}(0) & \theta_{i,k} = 1 \\ \frac{1}{2(\theta_{i,k} - 1)} \times [1 + a_{ij} - \sqrt{(1 + a_{i,j})^2 - 4\theta_{i,k}(\theta_{i,k} - 1)\pi_{1i,k}\pi_{2i,k}(0)}] & \theta_{i,k} \neq 1, \end{cases} \end{aligned}$$

where $a_{i,k} = (\pi_{1i,k} + \pi_{2i,k}(0))(\theta_{i,k} - 1)$ (Ten Have and Morabia, 1999).

Let ϕ denote the collection of unknown parameters in the specification of model (5)–(7). Note that if the unstructured form of the model is fit then $\phi \equiv \phi^\alpha = (\alpha, \beta)$ while $\phi \equiv \phi^\eta = (\eta, \beta)$ if the B-spline model is fit. For either specification, the observed data likelihood for a random sample of N study participants from the population of interest, $\mathcal{L}(\phi)$ is the product of N terms, each of the form:

$$\begin{aligned} \mathcal{L}_i(\phi) &= \prod_{k=k_i^l}^{k_i^r} P(Y_{1i,k} = y_{1i,k}, Y_{2i,k} = y_{2i,k} \mid Y_{1i,k-1} \\ &= y_{1i,k-1}, Y_{2i,k-1} = y_{2i,k-1}, \mathbf{X}_{i,k}^H), \\ &= \prod_{k=k_i^l}^{k_i^r} \left\{ [\pi_{12i,k}]^{y_{1i,k} y_{2i,k}} [\pi_{1i,k} - \pi_{12i,k}]^{y_{1i,k}(1-y_{2i,k})} \right. \\ &\quad \times [\pi_{2i,k}(0) - \pi_{12i,k}]^{(1-y_{1i,k})y_{2i,k}} \\ &\quad \times [1 - \pi_{1i,k} - \pi_{2i,k}(0) \\ &\quad \left. + \pi_{12i,k}]^{(1-y_{1i,k})(1-y_{2i,k})} \right\} (1-y_{1i,k-1})(1-y_{2i,k-1}) \\ &\quad \times \left\{ [\pi_{2i,k}(1)]^{y_{2i,k}} [1 - \pi_{2i,k}(1)]^{1-y_{2i,k}} \right\}^{y_{1i,k-1}(1-y_{2i,k-1})}. \end{aligned}$$

See Section A.2 of the Supporting Information for further details.

4.2 | Estimation

If the unstructured form of model (5)–(7) is adopted, estimation can proceed in a straightforward manner by finding the maximizer of $\ell(\phi^\alpha) = \log \mathcal{L}(\phi^\alpha)$, denoted by $\hat{\phi}^\alpha$. Under the proposed B-spline model, however, care is needed to avoid overfitting. To resolve this, we maximize the observed data likelihood subject to a penalty that imposes smoothness in resulting estimated functions. One such penalty is the integrated squared second derivative which, following Eilers and Marx (1996), can be approximated by penalizing coefficient

differences. Let Δ be the difference operator so that, for example, $\Delta\eta_{l,j} = \eta_{l,j} - \eta_{l,j-1}$. Then, letting $\lambda = (\lambda_1, \lambda_2, \lambda_\theta)$, consider the penalized likelihood:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\phi}^n; \lambda) = & \mathcal{L}(\boldsymbol{\phi}^n) - \lambda_1 \sum_{j=m+1}^{\tilde{J}_1} (\Delta^m \eta_{1,j})^2 \\ & - \lambda_2 \sum_{j=m+1}^{\tilde{J}_2} (\Delta^m \eta_{2,j})^2 - \lambda_\theta \sum_{j=m+1}^{\tilde{J}_\theta} (\Delta^m \eta_{\theta,j})^2, \end{aligned}$$

where Δ^m corresponds to the difference operator having been applied m times; for example, $\Delta^2 \eta_{l,j} = \Delta(\Delta \eta_{l,j}) = \eta_{l,j} - 2\eta_{l,j-1} + \eta_{l,j-2}$. Note that letting \mathbf{D}_m denote the matrix representation of Δ^m , such that, for example, $\sum_{j=m+1}^{\tilde{J}_1} (\Delta^m \eta_{1,j})^2$ can be written as $\boldsymbol{\eta}_1^T \mathbf{D}_m^T \mathbf{D}_m \boldsymbol{\eta}_1$ (Section A.2 of the Supporting Information), the penalized likelihood can be written as:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\phi}^n; \lambda) = & \mathcal{L}(\boldsymbol{\phi}^n) - \lambda_1 \boldsymbol{\eta}_1^T \mathbf{D}_m^T \mathbf{D}_m \boldsymbol{\eta}_1 - \lambda_2 \boldsymbol{\eta}_2^T \mathbf{D}_m^T \mathbf{D}_m \boldsymbol{\eta}_2 \\ & - \lambda_\theta \boldsymbol{\eta}_\theta^T \mathbf{D}_m^T \mathbf{D}_m \boldsymbol{\eta}_\theta. \end{aligned} \tag{8}$$

Let $\ell(\boldsymbol{\phi}^n; \lambda) = \log \mathcal{L}(\boldsymbol{\phi}^n; \lambda)$ be the log-penalized likelihood. Furthermore, let $\mathbf{U}(\boldsymbol{\phi}^n; \lambda) = \nabla_{\boldsymbol{\phi}^n} \ell(\boldsymbol{\phi}^n; \lambda)$ be the gradient of $\ell(\boldsymbol{\phi}^n; \lambda)$ with respect to $\boldsymbol{\phi}^n$ and $\boldsymbol{\mathcal{H}}(\boldsymbol{\phi}^n; \lambda) = \nabla_{\boldsymbol{\phi}^n \boldsymbol{\phi}^n} \ell(\boldsymbol{\phi}^n; \lambda)$ the corresponding matrix of second partial derivatives (i.e., the Hessian matrix). For a given value of λ , the penalized maximum likelihood estimator, which we denote by $\hat{\boldsymbol{\phi}}_s^n$, is the solution to $\mathbf{U}(\boldsymbol{\phi}^n; \lambda) = 0$. Note, since the penalty is quadratic in $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$, and $\boldsymbol{\eta}_\theta$, gradient-based maximization is relatively straightforward to carry out. This penalization induces smoothing in the estimated $\boldsymbol{\alpha}$ function as it approximates penalization of the m th derivatives of the $\boldsymbol{\alpha}$ functions (Eilers and Marx, 1996).

Finally, toward choosing the value of λ on which the final results are to be based, say λ^* , one can proceed using any standard model-selection criteria such as the Akaike Information Criterion (AIC) or cross-validation (Gray, 1992; Eilers and Marx, 1996). In our implementation of the methods, we follow Gray (1994) by taking the trace of $\boldsymbol{\mathcal{H}}(\boldsymbol{\phi}^n; 0) \boldsymbol{\mathcal{H}}^{-1}(\boldsymbol{\phi}^n; \lambda)$ as the effective degrees of freedom when calculating the AIC.

4.3 | Asymptotic properties

For the unstructured model, under standard regularity conditions, the asymptotic distribution of $\hat{\boldsymbol{\phi}}^\alpha$ follows from standard likelihood theory. That is, $\hat{\boldsymbol{\phi}}^\alpha$ is consistent and, letting $\tilde{\boldsymbol{\phi}}^\alpha$ denote the value of $\boldsymbol{\phi}^\alpha$ induced from the partition, $N^{1/2}(\hat{\boldsymbol{\phi}}^\alpha - \tilde{\boldsymbol{\phi}}^\alpha)$ is asymptotically normal with variance equal to the inverse of Fisher information matrix.

For the B-spline model, careful consideration of the asymptotic properties of $\hat{\boldsymbol{\phi}}^n$ requires specification of how the number of knots, $\tilde{\mathbf{J}} = (\tilde{J}_1, \tilde{J}_2, \tilde{J}_\theta)$, and the penalty, λ^* , change with the

sample size N . This has been the focus of a large body of work, much of which is concerned with error estimation for the nonparametric function(s) (e.g., Li and Ruppert, 2008). Here, however, primary interest lies with inference for $\boldsymbol{\phi}^\eta$. Given this, we consider the behavior of $\hat{\boldsymbol{\phi}}^\eta$ in settings where both $\tilde{\mathcal{J}}$ and λ are fixed. With this, henceforth, we denote the estimator as $\hat{\boldsymbol{\phi}}_\lambda^\eta$.

In Section A.3 of the Supporting Information, assuming standard regularity conditions, we show that $N^{1/2}(\hat{\boldsymbol{\phi}}_\lambda^\eta - \tilde{\boldsymbol{\phi}}_\lambda^\eta)$ is asymptotically normal with variance that can be consistently estimated by

$$\begin{aligned} \widehat{\mathcal{V}} &= N \mathcal{H}^{-1}(\hat{\boldsymbol{\phi}}^\eta; \lambda) \left(\frac{1}{N} \sum_{i=1}^N U_i(\hat{\boldsymbol{\phi}}^\eta; \lambda) U_i^T(\hat{\boldsymbol{\phi}}^\eta; \lambda) \right) \\ &\quad \times N \mathcal{H}^{-1}(\hat{\boldsymbol{\phi}}^\eta; \lambda), \end{aligned} \tag{9}$$

where $\tilde{\boldsymbol{\phi}}_\lambda^\eta$ is the solution of $E[U(\boldsymbol{\phi}^\eta; \lambda)] = 0$ and where $U_i(\cdot; \lambda)$ is the gradient of the log-penalized likelihood of a single observation i . For individual parameters, the appropriate entry in the diagonal of (9) is used for estimating the variance. For the baseline time trends, let $\widehat{\mathcal{V}}_{\eta_1}$ be the submatrix corresponding to the estimated variance matrix of $\hat{\boldsymbol{\eta}}_1$. Note that $\hat{\boldsymbol{\alpha}}_1$ can be written as $\hat{\boldsymbol{\alpha}}_1 = \hat{\boldsymbol{\eta}}_1^T \mathcal{B}$ with \mathcal{B} being a $\tilde{\mathcal{J}}_1 \times K$ matrix with elements $\mathcal{B}_{jk} = B_{j,k}$. Therefore, to construct a 95% confidence interval (CI) for the AD time-varying reference probability, $\pi_{1i,k}$, one can use $g^{-1} \left\{ \hat{\boldsymbol{\eta}}_1^1 \pm 1.96 [\text{diag}(\mathcal{B}^T \widehat{\mathcal{V}}_{\eta_1} \mathcal{B})]^{1/2} \right\}$ where here g^{-1} to be understood as operating entrywise on the vector, and $\text{diag}(\cdot)$ returns the diagonal of a matrix.

5 | SIMULATION STUDIES

We conducted a series of simulations to investigate: (i) finite-sample properties of the methods proposed in Section 4; and (ii) the potential bias-variance trade-off that analysts will have to contend with when choosing the degree of regularization in $\mathcal{L}(\boldsymbol{\phi}; \lambda)$. We note that we make no attempt to perform a comparison with existing methods since the proposed framework was developed to investigate dependence in a way that is distinct from how existing methods approach it. Due to space constraints, we present brief details and a summary of the main conclusions; see Section A.4 in the Supporting Information for full details.

Building on features of the observed data from the ACT study, we generated the data according to models (5)–(7) as a function of both time-fixed and time-dependent covariates under three scenarios for dependence: a null scenario, a simple dependence scenario, and a complex dependence scenario. A range of right-censoring rates (0–30%) and sample sizes (500–5000) were considered. For each scenario, 1000 data sets were generated. All simulations were carried out using code available in the `LongitSemiComp` package for R. Simulation code, seeds, and results are all available online via the Github repository of the first author.

Figures A.2 and A.3 in the Supporting Information summarize performance regarding estimation of time-varying functions. The time-varying terminal event probability function $\text{expit}(\alpha_2)$ was generally well-estimated. The time-varying nonterminal probability function $\text{expit}(\alpha_1)$ was also well-estimated except when the number of knots was small and a substantial amount of regularization were used (i.e., $\tilde{J} = 5$ and $\lambda \geq 5$); in this case the B-spline estimator was oversmoothed, resulting in bias for later time points where less information is available. The time-varying odds ratio function $\exp(\alpha_o)$ was well-estimated by the 10 knots B-spline estimator for large enough sample size, as well as by the five knots B-spline estimator when sample size was large enough and there were no oversmoothing. Similar to the nonterminal probability, the five knots over-smoothed estimator suffered from bias. A model with completely unrestricted α_1 , α_2 , and α_o performed well for α_1 , α_2 but not for α_o , for which considerable bias was found for low and moderate sample sizes. Instability of the undersmoothed estimators for time-varying odds ratio, and biased, yet stable, estimators when using excessive smoothing were found when the sample size was small. Using AIC to choose λ , more smoothing was desired for $\tilde{J} = 10$ compared with $\tilde{J} = 5$ (Table A.2).

Turning to the coefficients (Tables A.3–A.10), small finite-sample bias for β_1 and β_2 was mitigated under larger sample sizes. The global dependence parameter $\beta_{2,y}$ was well-estimated, with negligible bias for all sample sizes and censoring rates. A small finite-sample bias was observed for β_o in some of the scenarios, although it decreased as the sample size increased. Furthermore, the bias was more substantial when the sample size was low and the unrestricted model was used for the time-varying functions. Unpenalized estimation under B-spline representation also resulted in bias for β_o . However, penalization largely mitigated this and also reduced the standard error. For larger sample sizes, this bias disappeared. These results were consistent with the performance of the time-varying component estimator of the odds ratio. Finally, in most cases the proposed variance estimators performed well, with empirical coverage of confidence intervals close to the nominal level.

We conducted additional simulations to compare the different strategies in setting k'_i and k'_j (Section 3.2). Details are presented in Section A.4.4 of the Supporting Information.

6 | ANALYSIS OF DATA FROM THE ACT STUDY

The ACT study is an ongoing community-based prospective study of incident all-cause dementia and AD among the elderly in western Washington state (Kukull et al., 2002). Initiated in 1994, the goals of the study are to learn about how the brain ages and to identify risk factors for AD. In this paper, we consider data on $N = 4367$ ACT participants enrolled between 1994 and 2015, and who were aged 65 years or older and cognitively intact at the time of enrollment. Table A.12 in the Supporting Information summarizes key characteristics measured at study entry, including: age, gender, race, marital status, education, comorbid depression, and APOE- $\epsilon 4$ carrier status. Follow-up in ACT consists of biennial visits during which participants undergo a comprehensive neurological evaluation. For the purposes of this paper, follow-up time was administratively censored at the first of

December, 2016 or age 99 years. Based on these criteria, 205 (5%) were diagnosed with AD during follow-up but were censored prior to death, 818 (19%) were diagnosed with AD and died during follow-up, 1613 (37%) died during follow-up without a diagnosis of AD, and 1731 (39%) were censored prior to either a diagnosis of AD or experiencing death. Figure A.5 in the Supporting Information provides a summary of the observed person-time.

Having 1 APOE $\epsilon 4$ allele is well-established as a genetic risk factor for AD (Baumgart et al., 2015). The extent to which having 1 APOE $\epsilon 4$ allele is associated with mortality, however, is unclear (Helzner et al., 2008). Since, to the best of our knowledge, no previous studies have examined the role of the APOE $\epsilon 4$ allele through the lens of semicompeting risks, the opportunities that semi-competing risks analyses provide, particularly in terms of explicit acknowledgment of death as a competing risk and in terms of being able to learn about dependence, have not been taken advantage of.

With this backdrop, we present a case study with the goal to investigate the role of having 1 APOE $\epsilon 4$ allele on the joint risk of AD and death, and whether this varies by gender. Throughout, we use data from ACT with the time scale taken to be “time since age 65” or “time since a diagnosis of AD,” as appropriate. In reporting results, we focus on those that pertain to dependence; additional details, results, and sensitivity analyses are provided in Section A.6 of the Supporting Information. In addition, we performed a series of analyses based on a range of existing methods, including a standard illness-death model, the explanatory hazard ratio, and the cross-quantile residual ratio; Section A.6 of the Supporting Information provides details on the different analyses, results, and discussion.

6.1 | Analyses based on the proposed framework

As emphasized in Section 3.1, the choice of the partition τ is important. Toward illustrating the role of this choice, and recalling that participants in ACT underwent biennial visits, we considered two partitions of the time scale $[65, 100)$: $\tau^{2.5} = \{65.0, 67.5, 70.0, \dots, 97.5, 100.0\}$, for which $K = 14$; and, $\tau^{5.0} = \{65.0, 70.0, 75.0, \dots, 95.0, 100.0\}$, for which $K = 7$. See Tables A.13 and A.14 in the Supporting Information for the 2×2 outcome tables.

Toward specification of models (5)–(7), we used logit links for $\pi_{1i,k}$ and $\pi_{2i,k}(y_1)$, and a log link for the cross-sectional odds ratio, $\theta_{i,k}$. In these models, $\mathbf{X}_{i1,k}^H$, $\mathbf{X}_{i2,k}^H$, and $\mathbf{X}_{i\theta,k}^H$ were specified with the overarching goal of assessing the joint impact of having 1 APOE $\epsilon 4$ allele and gender on the joint risk of AD and death. As such, the models for $\pi_{1i,k}$ and $\theta_{i,k}$ included main effects for having 1 APOE $\epsilon 4$ allele and gender, and their interaction, while the model for $\pi_{2i,k}(\cdot)$ included main effects, two-way interactions and the three-way interaction between AD diagnosis, having 1 APOE $\epsilon 4$ allele and gender. For $\mathbf{X}_{i1,k}^H$, $\mathbf{X}_{i2,k}^H$, we additionally included all other variables available in data set.

To complete the models specifications, we considered unstructured and B-spline baseline parameters, α_1 , α_2 , and α_θ . For the latter, we used 10 knots for $\tau^{2.5}$ and five knots for τ^5 , with a cubic spline and a second-order difference penalty (i.e., $m = 2$ in Δ^m),

and considered varying degrees of penalty, specifically setting $\lambda_1 = \lambda_2 = \lambda_0 = \lambda$, with $\lambda \in \{0.0, 0.1, 0.5, 1.0, 2.5, 5.0\}$. Finally, following the discussion of Section 3.2, we considered all three strategies for determining k'_i and k'_r . Table A.19 in the Supporting Information reports AIC from the various fitted models from which we see that the optimal values of λ were $\lambda^* = 2.5$ and $\lambda^* = 0.0$ for $\tau^{2.5}$ and $\tau^{5.0}$, respectively, across all three strategies for determining k'_i and k'_r . While there were no convergence issues in the data analysis, in our experience with the simulation studies adding a small penalty may mitigate potential convergence issues (see Section A.4.2 in the Supporting Information).

6.1.1 | Baseline time trends—Figures 2 and 3 report estimated baseline time trends, with the latter focusing on the estimates obtained based on the nearest neighbor strategy. Note that the interpretation of these quantities is specific to a population of individuals who are cognitively intact at age 65 and have the following characteristics: male, nonwhite, non-college-educated, married without depression, and no APOE $\epsilon 4$ alleles.

From Figure 2, we find little overall sensitivity in the estimates across the three strategies for determining k'_i and k'_r , with the sole exception being in the tails of the conditional probability of death. That this is the case may not be surprising, however, given that overall probability is larger and that the sample size is smaller as the population ages into their 90s. As expected, however, the nearest neighbor strategy provides a reasonable compromise between the conservative and anticonservative strategies.

From Figure 3, the unstructured models, not surprisingly, exhibit greater uncertainty, especially in information-poor parts of the age scale (i.e., early on when there are relatively few AD events). From the top panel of Figure 3, under the B-spline analysis with $\lambda = 2.5$, the estimated baseline probability of an AD diagnosis during a given 2.5 year age interval, conditional on being AD-free and alive at the start of the interval, increases from 0.005 in (65.0, 67.5] to 0.11 in (97.5, 100.0]. From the bottom panel of Figure 3, the same increasing pattern emerges under partition $\tau^{5.0}$ with $\lambda = 0.0$, specifically from 0.01 in (65.0, 70.0] to 0.17 in (95.0, 100.0]. Note that this is as expected since the intervals are longer and, correspondingly, the cumulative number of events higher. Also from Figure 3, we see that the estimated baseline probability of death conditional on being event free during a given 2.5-year age interval increases from 0.01 in (65.0, 67.5] to 0.38 in (97.5, 100.0]. Again, the same general pattern is observed under $\tau^{5.0}$, with the probability increasing from 0.03 in (65.0, 70.0] to 0.59 in (95.0, 100.0].

6.1.2 | Covariate effects for AD—Table 2 reports estimated covariate associations from the B-spline fits for the main effects and interactions corresponding to gender, having 1 APOE $\epsilon 4$ allele and AD diagnosis (as appropriate). Results based on all three strategies regarding k'_i and k'_r are presented; complete results, including those from the unstructured models, are given in Tables A.20–A.24 of the Supporting Information.

From the first two columns of Table 2, the results regarding risk of AD are largely consistent between the two partitions as well as across the three strategies for determining k'_i and k'_r . Specifically, we find that, while having 1 APOE $\epsilon 4$ allele is a strong risk factor, there is no

evidence that gender is a risk factor nor that it is an effect modifier of the APOE $\epsilon 4$ effect. Note that the results under the unstructured model are similar (see the Supporting Information).

6.1.3 | Covariate effects for death and global dependence—The middle set of results in Table 2 report on how gender, having 1 APOE $\epsilon 4$ allele and having a diagnosis of AD jointly influence risk of mortality. Note the four components involving a diagnosis of AD correspond to $\beta_{2,y}$ in model (6) and, thus, jointly represent global dependence between AD and mortality. As with the results regarding the risk of AD, the results for death are largely consistent across the three strategies for determining k'_i and k''_i . Interestingly, there are some differences between those based on the $\tau^{2.5}$ partition and those based on the $\tau^{5.0}$ partition, although the overarching conclusion that there is an important interplay between the three factors in determining risk of mortality is consistent between the two.

To further facilitate discussion of the results, Table 3 reports odds ratio estimates and 95% confidence intervals for mortality, based on the B-spline model with the nearest neighbor strategy for determining k'_i and k''_i , across combinations of whether the patient has had an AD diagnosis, their gender, and whether they have 1 APOE $\epsilon 4$ allele. From the first four rows, in the absence of a diagnosis of AD, female gender is associated with approximately 40% lower odds of mortality while there is no evidence to indicate that the number of APOE $\epsilon 4$ alleles play a role.

From the lower half of Table 3, the odds of mortality among those patients with a diagnosis of AD diagnosis, and, thus, extent of global dependence, is substantially higher with the magnitude depending on the interplay between gender and the number of APOE $\epsilon 4$ alleles. Moreover, while having at least APOE $\epsilon 4$ alleles seems to be associated with higher odds, the increase is substantially larger for males.

6.1.4 | Local dependence—Turning to the assessment of local dependence, the third row in the two subfigures of Figure 3 suggest that there are meaningful time trends in the co-occurrence of AD and death, specifically that the risk of co-occurrence within a given interval are highest at the early ages. From the B-spline specification with $\lambda = 2.5$ applied to partition $\tau^{2.5}$, for example, the odds ratio among males with no APOE $\epsilon 4$ alleles decreases from 2.61 (95% CI: 0.89, 7.69) during the (65.0, 67.5] interval to 1.20 (95% CI: 0.57, 2.51) during the (97.5, 100.0]. From Table 2, although the confidence intervals all include 1.00, the point estimates for all three covariates in the model applied to the $\tau^{2.5}$ partition are indicative of clinically meaningful associations. For example, the local dependence odds ratio is estimated to be 44% smaller among males with 1 APOE $\epsilon 4$ allele compared to those without. The corresponding odds ratio for females is estimated to be approximately 31% smaller ($0.56 \times 1.24 \approx 0.69$) for those with 1 APOE $\epsilon 4$ allele compared to those without.

Finally, Tables 2 and 3 highlight that the choice of partition can have a meaningful impact on the conclusions regarding local. In particular, under $\tau^{5.0}$ the evidence regarding an interaction between gender and having 1 APOE $\epsilon 4$ allele on local dependence is weaker when the partition intervals are 5 years in length.

7 | DISCUSSION

Although less familiar than competing risks, semi-competing risks arise in a wide range of clinical settings, including: AD, as illustrated in this paper; hospital readmission (Lee et al., 2015, 2016); shock among patients with implanted cardiac devices (Reeder et al., 2019); and graft-versus-host disease (Jazi et al., 2020; Lee et al., 2020). One distinguishing feature of semi-competing risks is that there is partial information about the joint distribution between T_1 and T_2 . To leverage that information, we have proposed novel framework that seeks to gain interpretable insight into dependence between the two events. Because the interpretation of the model components, including the proposed notions of global and local dependence, are distinct from those obtained from existing methods, we view the framework as being scientifically complementary to existing methods. Moreover, we view the proposed framework as being in-line with recent work that seeks to better understand whether and how specific factors confer risk jointly on multiple outcomes, such as the *dual hazard rate* (Prentice and Zhao, 2020).

The foundation for the proposed framework is the discretization of the time scale. As discussed in Sections 2.2 and 3.1, one cannot say that a given choice of τ is the “truth” and yet the specific choice dictates the numerical values and interpretation of the results. This results in a tension that is illustrated in Table 2: one cannot claim that either $\tau^{2.5}$ or $\tau^{5.0}$ is the “right” choice and yet there are instances where the numerical results differ in meaningful ways. Our view of this dilemma is that consideration of multiple partitions should be viewed as an opportunity to obtain additional insight. Consider, for example, the main effect of female gender in the model for $\theta_{i,k}$ based on the B-spline specification: under $\tau^{2.5}$ the estimated impact of female gender is to reduce the local dependence odds ratio by 20% while the reduction is only 10% under $\tau^{5.0}$. Thus, ignoring (for the purpose of discussion) the lack of statistical significance, there is an indication that among individuals with no APOE $\epsilon 4$ alleles, gender plays a role in the co-occurrence of AD and death over relatively short time frames (i.e., 2.5 years) and over longer time frames (i.e., 5 years), although the magnitude of the effect attenuates over the longer time frame. Conceptually, this is analogous to the results that one might see in a Cox model for a univariate outcome if the effect of a covariate varies over time (i.e., nonproportional hazards) and yet proportional hazards is adopted; in such settings, the value of the common hazard ratio that is being estimated will depend on the interval over which data are available.

While the proposed framework provides researchers with a flexible approach to investigating dependence in semi-competing risks data, one trade-off is that, as discussed in detail in Section 3.2, a degree of ambiguity arises regarding the specific form of the contribution to the likelihood for select individuals. Our approach to addressing this challenge, one that is similarly faced in other time-to-event analyses based on a discretization of the time scale (Prentice and Gloeckler, 1978; D’Agostino et al., 1990; Hernán et al., 2000), is via sensitivity analyses. While the simulations we present and data application point to little sensitivity for the contexts they are based on, in other settings (e.g., when there is substantial censoring and/or when the intervals of the partition are wide) the results may indeed be sensitive. For these settings, researchers may opt to report all results and attempt to glean

insight from the specifics of the context or to view the sensitivity as an indicator that there is insufficient information on which to base definitive conclusions regarding dependence.

A number of extensions to the proposed framework are possible. The Markov assumption (1) could be relaxed, for example, by having $\pi_{2,i,k}(y_i)$ in (3) depend on the entire disease status history $Y_{i,1}, \dots, Y_{i,k-1}$. More broadly, a topic for future research is developing methods for testing of assumptions and assessment goodness of fit. From a theoretical perspective, an interesting alternative to having λ fixed is to consider $\lambda = \lambda_N = o(1)$; that is the amount of regularization decreases as more data become available. This framing of the asymptotics, however, leads to a variance expression that does not involve λ and thus has been perceived as of less useful in practice (Gray, 1992; Yu and Ruppert, 2002). Another issue concerns interval-censoring of the data. As mentioned in Section 6 follow-up in ACT consists of biennial visits. As such, while the date of death can be precisely ascertained through death records, AD is subject to interval censoring. Therefore, generalizing our approach to interval-censored data will be of interest. A related problem may arise in defining covariate values when time-dependent covariates are observed only intermittently (Nevo et al., 2020).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank the Associate Editor and Reviewer for helpful comments and gratefully acknowledge funding from National Institutes of Health Grant R-01 CA181360.

DATA AVAILABILITY STATEMENT

The data that were used in this paper are available from Kaiser Permanente Washington Health Research Institute. Restrictions may apply to the availability of these data. Further information can be found at <https://www.kpwashingtonresearch.org>.

REFERENCES

- Alzheimer's Association., (2018) 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 14, 367–429.
- Baumgart M, Snyder H, Carrillo M, Fazio S, Kim H and Johns H (2015) Summary of the evidence on modifiable risk factors for cognitive decline and dementia: a population-based perspective. *Alzheimer's & Dementia*, 11, 718–726.
- Carey V, Zeger S and Diggle P (1993) Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, 80, 517–526.
- D'Agostino R, Lee M-L, Belanger A, Cupples L, Anderson K and Kannel W (1990) Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham heart study. *Statistics in Medicine*, 9, 1501–1515. [PubMed: 2281238]
- Egleston B, Scharfstein D, Freeman E and West S (2006) Causal inference for non-mortality outcomes in the presence of death. *Biostatistics*, 8, 526–545. [PubMed: 16980696]
- Eilers P and Marx B (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–102.
- Fine J, Jiang H and Chappell R (2001) On semi-competing risks data. *Biometrika*, 88, 907–919.

- Gray R (1992) Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87, 942–951.
- Gray R (1994) Spline-based tests in survival analysis. *Biometrics*, 50, 640–652. [PubMed: 7981391]
- Haneuse S and Lee K (2016) Semi-competing risks data analysis: accounting for death as a competing risk when the outcome of interest is nonterminal. *Circulation: Cardiovascular Quality and Outcomes*, 9, 322–331. [PubMed: 27072677]
- Helzner E, Scarmeas N, Cosentino S, Tang M, Schupf N and Stern Y (2008) Survival in Alzheimer disease: a multi-ethnic, population-based study of incident cases. *Neurology*, 71, 1489–1495. [PubMed: 18981370]
- Hernán MÁ, Brumback B and Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11, 561–570. [PubMed: 10955409]
- Jazi I, Lee S, and Haneuse S (2020) Estimation and inference for semi-competing risks based on data from a nested case-control study. *Statistical Methods in Medical Research*, 29, 3326–3339. [PubMed: 32552435]
- Jazi I, Schrag D, Sargent D and Haneuse S (2016) Beyond composite endpoints analysis: semi-competing risks as an underutilized framework for cancer research. *Journal of the National Cancer Institute*, 108, djw154. [PubMed: 27381741]
- Kukull W, Higdon R, Bowen J, McCormick W, Teri L, Schellenberg G, et al. (2002) Dementia and Alzheimer disease incidence: a prospective cohort study. *Arch of Neurology*, 59, 1737–1746.
- Lee C, Lee SJ and Haneuse S (2020) Time-to-event analysis when the event is defined on a finite time interval. *Statistical Methods in Medical Research*, 29, 1573–1591. [PubMed: 31436136]
- Lee K, Dominici F, Schrag D and Haneuse S (2016) Hierarchical models for semi-competing risks data with application to quality of end-of-life care for pancreatic cancer. *Journal of the American Statistical Association*, 111, 1075–1095. [PubMed: 28303074]
- Lee K, Haneuse S, Schrag D and Dominici F (2015) Bayesian semiparametric analysis of semi-competing risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society Series C*, 64, 253–273.
- Lee K, Rondeau V and Haneuse S (2017) Accelerated failure time models for semi-competing risks data in the presence of complex censoring. *Biometrics*, 73, 1401–1412. [PubMed: 28395116]
- Li R and Peng L (2015) Quantile regression adjusting for dependent censoring from semi-competing risks. *Journal of the Royal Statistical Society Series B*, 77, 107–130.
- Li Y and Ruppert D (2008) On the asymptotics of penalized splines. *Biometrika*, 95, 415–436.
- Lipsitz S, Laird N and Harrington D (1991) Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, 78, 153–160.
- Nevo D and Gorfine M (2020) Causal inference for semi-competing risks data. arXiv preprint arXiv:2010.04485.
- Nevo D, Hamada T, Ogino S and Wang M (2020) A novel calibration framework for survival analysis when a binary covariate is measured at sparse time points. *Biostatistics*, 21, e148–e163. [PubMed: 30380012]
- Peng L and Fine J (2007) Regression modeling of semi-competing risks data. *Biometrics*, 63, 96–108. [PubMed: 17447934]
- Prentice R and Gloeckler L (1978) Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34, 57–67. [PubMed: 630037]
- Prentice R and Zhao S (2020) Regression models and multivariate life tables. *Journal of the American Statistical Association*, 1–16.
- Reeder H, Shen C, Buxton A, Haneuse S and Kramer D (2019) Joint shock/death risk prediction model for patients considering implantable cardioverter-defibrillators: a secondary analysis of the SCD-HeFT trial. *Circulation: Cardiovascular Quality and Outcomes*, 12, e005675. [PubMed: 31412732]
- Tchetgen Tchetgen EJ (2014) Identification and estimation of survivor average causal effects. *Statistics in Medicine*, 33, 3601–3628. [PubMed: 24889022]

- Ten Have T and Morabia A (1999) Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometrics*, 55, 85–93. [PubMed: 11318182]
- Tsiatis A (1975) A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72, 20–22.
- Varadhan R, Xue Q-L and Bandeen-Roche K (2014) Semi-competing risks in aging research: methods, issues and needs. *Lifetime Data Analysis*, 20, 538–562. [PubMed: 24729136]
- Xu J, Kalbfleisch J and Tai B (2010) Statistical analysis of illness–death processes and semi-competing risks data. *Biometrics*, 66, 716–725. [PubMed: 19912171]
- Yang J and Peng L (2016) A new flexible dependence measure for semi-competing risks. *Biometrics*, 72, 770–779. [PubMed: 26916804]
- Yu Y and Ruppert D (2002) Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97, 1042–1054.

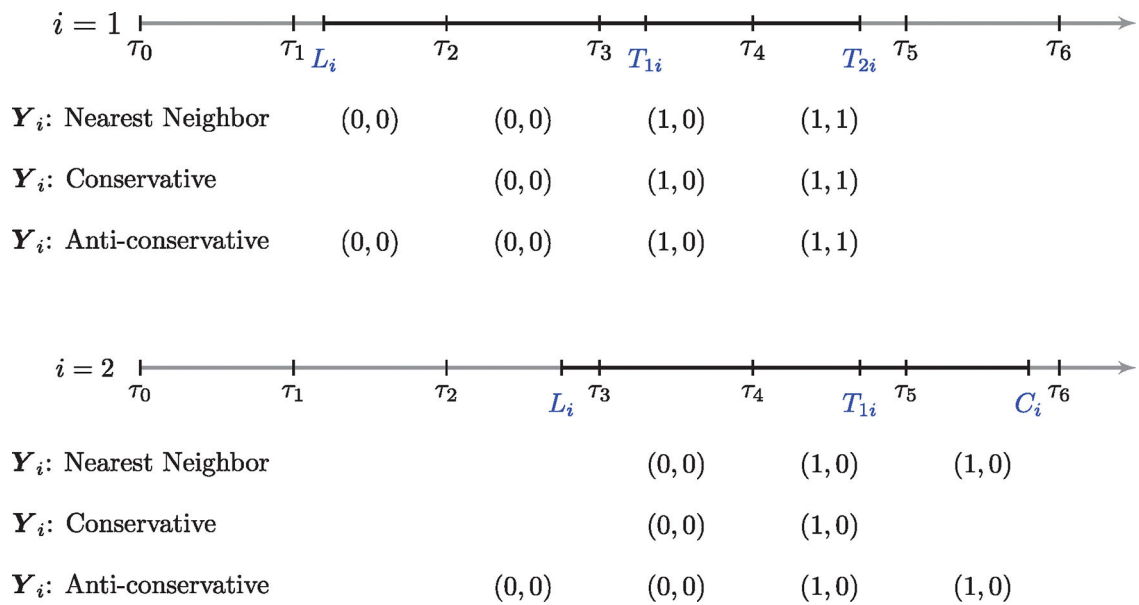


FIGURE 1. Graphical representation of the interplay between standard notation for semi-competing risks outcome data and the proposed bivariate longitudinal framework; see Sections 2 and 3.2 for details.

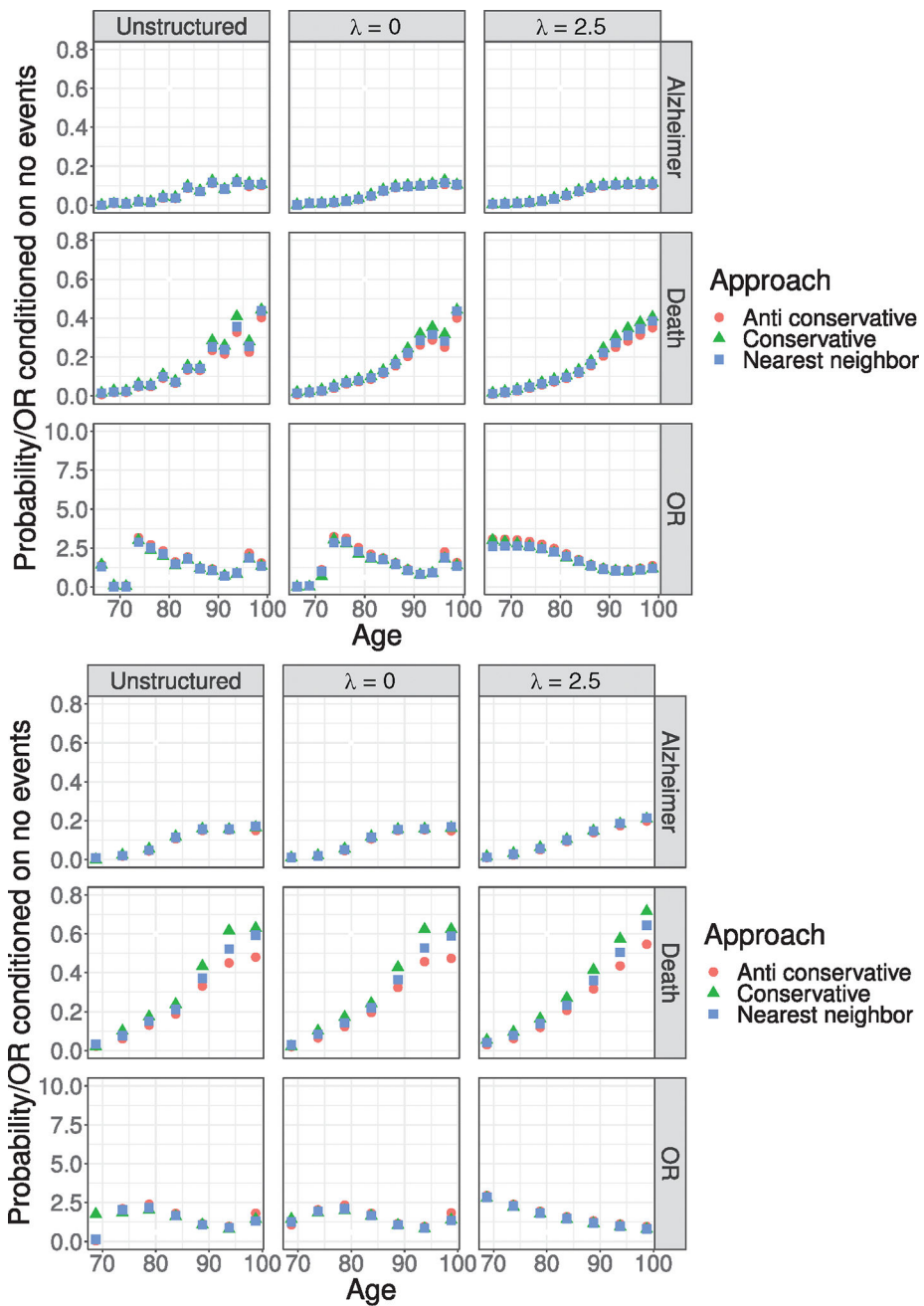


FIGURE 2. Comparison in the ACT data between the three approaches to overcome within-interval censoring and study entry times. The figures present the estimated baseline time trends, $\text{expit}(\hat{\alpha}_1)$, $\text{expit}(\hat{\alpha}_2)$, and $\text{exp}(\hat{\alpha}_\theta)$, under the unstructured specification and the B-spline specification with $\lambda \in \{0.0, 2.5\}$ and under the two partitions $\tau^{2.5}$ (top panel) and $\tau^{5.0}$ (bottom panel). Note that the y -axis has been truncated at 0.8 for $\text{expit}(\hat{\alpha}_i)$ and at 10 for $\text{exp}(\hat{\alpha}_\theta)$.

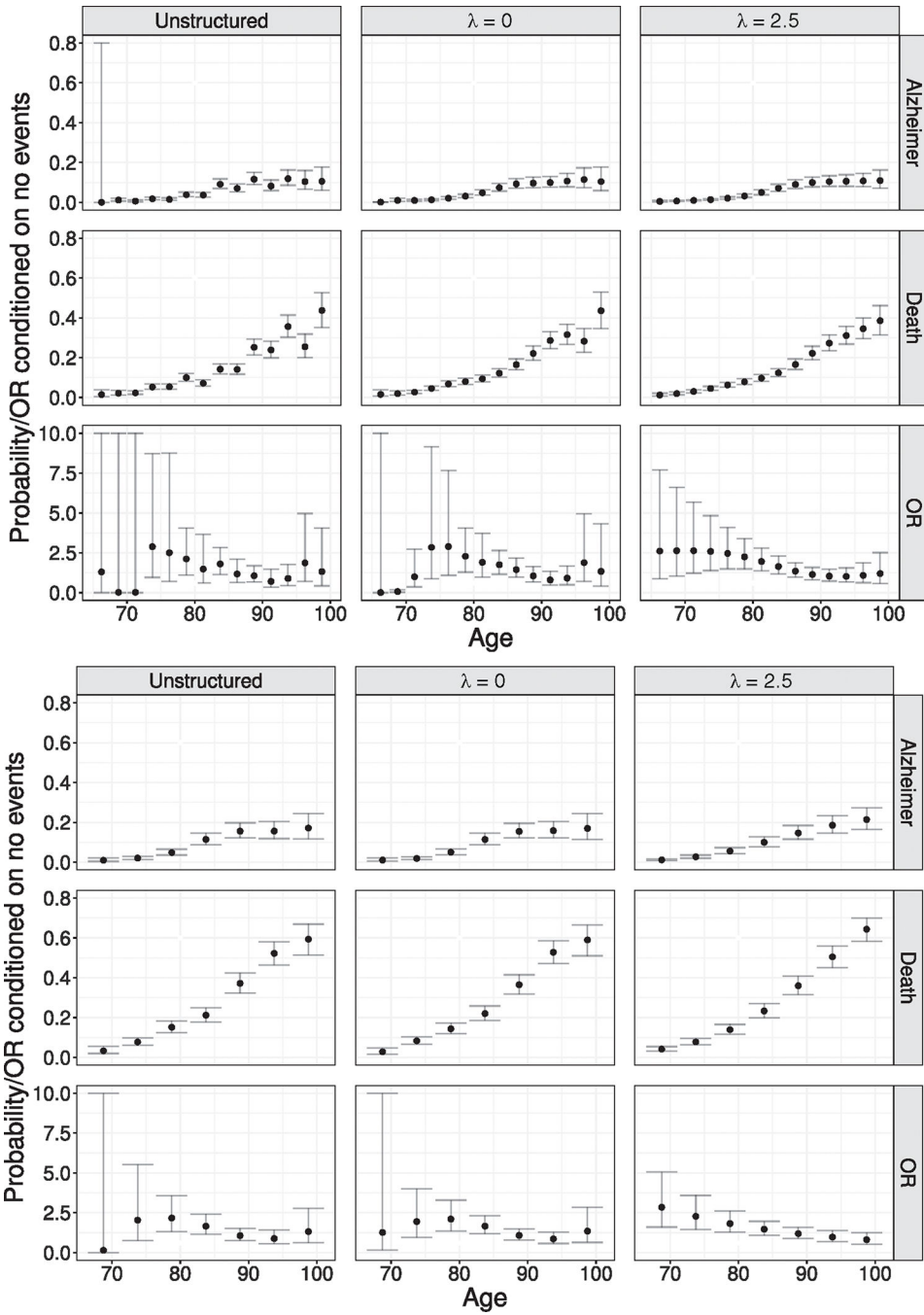


FIGURE 3. Estimated baseline time trends, $\text{expit}(\hat{\alpha}_1)$, $\text{expit}(\hat{\alpha}_2)$, and $\text{exp}(\hat{\alpha}_\theta)$ from a series of analyses to the ACT data, under the unstructured specification and the B-spline specification with $\lambda \in \{0.0, 2.5\}$ and under the two partitions $\tau^{2.5}$ (top panel) and $\tau^{5.0}$ (bottom panel). See Section 6.1 for details. Also shown are 95% confidence intervals. Note that the y-axis has been truncated at 0.8 for $\text{expit}(\hat{\alpha}_i)$ and at 10 for $\text{exp}(\hat{\alpha}_\theta)$. For results under more λ values,

see Supplementary Figures A.10 and A.11. Results presented under the nearest neighbor strategy

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 1Six possible data scenarios for the k th interval in the partition given by τ (see Section 2)

$(Y_{1i,k-1}, Y_{2i,k-1})$	$(Y_{1i,k}, Y_{2i,k})$	Likelihood contribution
(0, 0)	(0, 0)	$1 - \pi_{1i,k} - \pi_{2i,k}(0) + \pi_{12i,k}$
(0, 0)	(1, 0)	$\pi_{1i,k} - \pi_{12i,k}$
(0, 0)	(0, 1)	$\pi_{2i,k}(0) - \pi_{12i,k}$
(0, 0)	(1, 1)	$\pi_{12i,k}$
(1, 0)	(1, 0)	$1 - \pi_{2i,k}(1)$
(1, 0)	(1, 1)	$\pi_{2i,k}(1)$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Select results, specific to the role of gender and having 1 APOE $\epsilon 4$ allele on the joint risk of AD and death, based on the proposed framework applied to the data from the Adult Changes in Thought study

TABLE 2

	Alzheimer's Disease, $\pi_{1,k}$		Death, $\pi_{2,k}$		Cross-sectional odds ratio, $\theta_{i,k}$	
	$\exp(\hat{\beta}_1)$		$\exp(\hat{\beta}_2)$		$\exp(\hat{\beta}_0)$	
	$\tau^{2.5}$	$\tau^{5.0}$	$\tau^{2.5}$	$\tau^{5.0}$	$\tau^{2.5}$	$\tau^{5.0}$
Conservative ^a						
Female	0.98 (0.82, 1.16)	1.13 (0.94, 1.36)	0.58 (0.51, 0.65)	0.55 (0.48, 0.64)	0.76 (0.50, 1.17)	0.82 (0.55, 1.22)
APOE ^b	1.85 (1.44, 2.36)	1.96 (1.48, 2.58)	0.97 (0.80, 1.18)	0.97 (0.78, 1.22)	0.57 (0.28, 1.13)	0.72 (0.39, 1.34)
Female \times APOE ^b	0.98 (0.72, 1.34)	0.91 (0.64, 1.29)	1.13 (0.87, 1.46)	1.19 (0.89, 1.59)	1.24 (0.50, 3.04)	1.16 (0.53, 2.53)
AD ^c			2.54 (1.99, 3.24)	2.69 (1.88, 3.87)		
AD ^c \times Female			1.52 (1.12, 2.06)	1.95 (1.22, 3.11)		
AD ^c \times APOE ^b			1.55 (0.99, 2.44)	2.01 (0.99, 4.06)		
AD ^c \times Female \times APOE ^b			0.66 (0.38, 1.14)	0.55 (0.23, 1.29)		
Anti-conservative ^a						
Female	0.97 (0.82, 1.14)	0.99 (0.83, 1.17)	0.57 (0.51, 0.64)	0.55 (0.48, 0.62)	0.78 (0.51, 1.19)	0.96 (0.67, 1.38)
APOE ^b	1.84 (1.45, 2.32)	1.83 (1.44, 2.34)	0.97 (0.81, 1.17)	0.98 (0.81, 1.20)	0.58 (0.29, 1.14)	0.78 (0.44, 1.38)
Female \times APOE ^b	1.03 (0.77, 1.38)	1.03 (0.76, 1.41)	1.10 (0.85, 1.41)	1.13 (0.87, 1.46)	1.20 (0.49, 2.91)	1.01 (0.49, 2.09)
AD ^c			2.61 (2.06, 3.31)	2.77 (2.01, 3.82)		
AD ^c \times Female			1.56 (1.16, 2.09)	1.82 (1.21, 2.73)		
AD ^c \times APOE ^b			1.69 (1.09, 2.62)	2.33 (1.24, 4.41)		
AD ^c \times Female \times APOE ^b			0.62 (0.36, 1.06)	0.49 (0.23, 1.05)		
Nearest Neighbor ^a						
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 ^b	1.87 (1.48, 2.37)	1.91 (1.48, 2.45)	0.96 (0.79, 1.16)	0.98 (0.80, 1.19)	0.56 (0.28, 1.10)	0.70 (0.39, 1.26)
Female \times APOE ^b	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)

	Alzheimer's Disease, $\pi_{1i,k}$		Death, $\pi_{2i,k}$		Cross-sectional odds ratio, $\theta_{i,k}$	
	$\tau^{2.5}$	$\tau^{5.0}$	$\tau^{2.5}$	$\tau^{5.0}$	$\tau^{2.5}$	$\tau^{5.0}$
	$\exp(\hat{\beta}_1)$		$\exp(\hat{\beta}_2)$		$\exp(\hat{\beta}_b)$	
AD ^c			2.63 (2.07, 3.34)	2.80 (2.00, 3.92)		
AD ^c × Female			1.51 (1.12, 2.04)	1.94 (1.26, 2.99)		
AD ^c × APOE ^b			1.56 (1.00, 2.43)	2.34 (1.19, 4.58)		
AD ^c × Female × APOE ^b			0.66 (0.38, 1.15)	0.42 (0.19, 0.96)		

^b An indicator of having at least one APOE ε4 allele.

^c An indicator of whether the patient has previously had a diagnosis of Alzheimer's disease

^d Based on the optimal λ of $\lambda^*=2.5$ for $\tau^{2.5}$ and $\lambda^*=0$ for $\tau^{5.0}$.

Note Results are shown for the B-spline approach. Complete results are given in Tables A.20–A.24 of the Supporting Information

Odds ratio estimates and 95% confidence intervals for mortality across combinations of whether the patient has had an AD diagnosis, their gender, and whether they have 1 APOE $\epsilon 4$ allele

TABLE 3

AD status	Gender	# APOE $\epsilon 4$ alleles	Odds ratio (95% CI)	
			$\tau^{2.5}$	$\tau^{5.0}$
No AD	Male	0	1.00	1.00
No AD	Male	1	0.96 (0.79, 1.16)	0.98 (0.80, 1.19)
No AD	Female	0	0.57 (0.50, 0.64)	0.56 (0.49, 0.64)
No AD	Female	1	0.61 (0.52, 0.73)	0.61 (0.51, 0.74)
AD	Male	0	2.63 (2.07, 3.34)	2.80 (2.00, 3.92)
AD	Male	1	3.94 (2.79, 5.56)	6.42 (3.66, 11.28)
AD	Female	0	2.26 (1.87, 2.75)	3.04 (2.29, 4.04)
AD	Female	1	2.51 (1.97, 3.18)	3.28 (2.32, 4.65)

Note. Results are based on the B-spline models under the nearest neighbor strategy as reported in Table 2.