# Fitting a shared frailty illness-death model to left-truncated semi-competing risks data to examine the impact of education level on incident dementia

Catherine Lee[1]* ![ORCID], Paola Gilsanz[1] and Sebastien Haneuse[2]

## Abstract

**Background:** Semi-competing risks arise when interest lies in the time-to-event for some non-terminal event, the observation of which is subject to some terminal event. One approach to assessing the impact of covariates on semi-competing risks data is through the illness-death model with shared frailty, where hazard regression models are used to model the effect of covariates on the endpoints. The shared frailty term, which can be viewed as an individual-specific random effect, acknowledges dependence between the events that is not accounted for by covariates. Although methods exist for fitting such a model to right-censored semi-competing risks data, there is currently a gap in the literature for fitting such models when a flexible baseline hazard specification is desired and the data are left-truncated, for example when time is on the age scale. We provide a modeling framework and openly available code for implementation.

**Methods:** We specified the model and the likelihood function that accounts for left-truncated data, and provided an approach to estimation and inference via maximum likelihood. Our model was fully parametric, specifying baseline hazards via Weibull or B-splines. Using simulated data we examined the operating characteristics of the implementation in terms of bias and coverage. We applied our methods to a dataset of 33,117 Kaiser Permanente Northern California members aged 65 or older examining the relationship between educational level (categorized as: high school or less; trade school, some college or college graduate; post-graduate) and incident dementia and death.

**Results:** A simulation study showed that our implementation provided regression parameter estimates with negligible bias and good coverage. In our data application, we found higher levels of education are associated with a lower risk of incident dementia, after adjusting for sex and race/ethnicity.

**Conclusions:** As illustrated by our analysis of Kaiser data, our proposed modeling framework allows the analyst to assess the impact of covariates on semi-competing risks data, such as incident dementia and death, while accounting for dependence between the outcomes when data are left-truncated, as is common in studies of aging and dementia.

**Keywords:** Illness-death, Left-truncation, Semi-competing risks, Multistate models, B-splines, Dementia

*Correspondence: catherine.lee@kp.org
[1] Kaiser Permanente Northern California, Division of Reseach, 2000 Broadway, Oakland, CA, US
Full list of author information is available at the end of the article

# Background

Semi-competing risks refers to the setting where interest lies in the time-to-event for some so-called *non-terminal* event, the observation of which is subject to some *terminal* event [1]. In contrast to standard competing risks, where each of the outcomes under consideration is typically terminal (e.g. death due to some cause or another), in the semi-competing risks setting it is possible to observe both events on the same study unit, so that there is at least partial information on their joint distribution [1, 2]. Take as an example the study of dementia among the elderly [3], a complex neurocognitive condition that is estimated to affect nearly 6 million individuals aged 65 and older in the US [4], a number that has been projected increase to 13.9 million by 2060 [5]. It is known that the risk of death is higher among those who are diagnosed with dementia [6]. As such, studies seeking to investigate risk factors for dementia must contend with death as a competing risk, which precludes the subsequent observation of dementia. However, it is possible to observe both outcomes among individuals who die following a diagnosis of dementia. This information can potentially increase efficiency of results and be used to assess the dependence between the nonterminal and terminal events.

Towards the analysis of semi-competing risks data, the statistical literature has focused on three broad frameworks that seek to exploit the joint information on the time to the non-terminal event and the time to the terminal event [7]: those based on copulas [1, 8–10]; those framed from the perspective of causal inference [11, 12]; and, those based on the illness-death multi-state model [2, 13–16]. In this paper, we focus on the last of these approaches, for which the philosophical underpinning is that patients begin in some initial state at time zero and may transition into the non-terminal and/or terminal state [14, 16–18]. Analyses typically proceed through the development of models for transition-specific hazard functions (which dictate the rate at which patients experience the events), often with the use of subject-specific frailties, which can be viewed as individual-specific random effects that acknowledge the heterogeneity across individuals that is not accounted for by covariates [19, 20]. Moreover, the shared frailty accounts for dependence between the nonterminal and terminal events, which can be quantified from an estimable frailty variance parameter.

In the analysis of time-to-event outcomes, data are subject to left-truncation or delayed entry when subjects are enrolled into a study after the time origin of interest. Left-truncation is common in the study of aging and dementia, where the age scale is commonly taken to be the time scale [21–23]. In this setting, sampling is biased toward longer follow-up times since patients are typically only included in the study if they are dementia-free at study entry. The analysis of left-truncated time-to-event data should apply statistical methods that account for this bias. Although current methods exist for analyzing left-truncated semi-competing risks data via a standard illness-death model (without a shared frailty) [24, 25] and have been applied to Alzheimer's disease [26–29], to our knowledge, there are no published methods in the literature for fitting an illness-death model with shared frailty to left-truncated semi-competing risks data. The purpose of this paper is to fill this gap by providing a modeling framework and openly available code for implementation.

In this paper, we provide methods for fitting an illness-death model with shared frailty to left-truncated semi-competing risks data. In "Methods" section, we present the model specification ("Model specification: Illness-death model" section), methods for estimation and inference ("Estimation and inference" section), a brief simulation study ("Simulation study" section), and an analysis of data from Kaiser Permanente Northern California examining the relationship between educational level and incident dementia and death ("Assessing the impact of education level on incident dementia in a large US cohort" section). We present results in "Results" section. We conclude with a discussion in "Discussion" section and conclusions in "Conclusions" section.

# Methods

## Model specification: Illness-death model

Semi-competing risks refers to the setting where interest lies in a nonterminal event that is subject to dependent censoring by a terminal event [1]. This paper focuses on modeling semi-competing risks data using the illness-death multistate model, where in the study of dementia we take the nonterminal and terminal events to be dementia diagnosis and death. Figure 1 illustrates the possible individual trajectories following time of study accrual. Let $T_1$ and $T_2$ denote the nonterminal and terminal event times, respectively, following a well-defined time origin. Following the multistate modeling literature, the illness-death model is characterized by the transition hazards:

$$\lambda_1(t_1) = \lim_{\Delta \to 0} P(T_1 \in [t_1, t_1 + \Delta)|T_1 \geq t_1, T_2 \geq t_1)/\Delta, \text{ for } t_1 > 0 \quad (1)$$

$$\lambda_2(t_2) = \lim_{\Delta \to 0} P(T_2 \in [t_2, t_2 + \Delta)|T_1 \geq t_2, T_2 \geq t_2)/\Delta, \text{ for } t_2 > 0 \quad (2)$$

$$\lambda_3(t_2|T_1 = t_1) = \lim_{\Delta \to 0} P(T_2 \in [t_2, t_2 + \Delta)|T_1 = t_1, T_2 \geq t_2)/\Delta, \text{ for } 0 < t_1 < t_2. \quad (3)$$

Although a variety of hazard regression models can be adopted, including Cox-type multiplicative hazard regression models [30], additive models [31] and accelerated

**Fig. 1** Schematic of an illness-death model
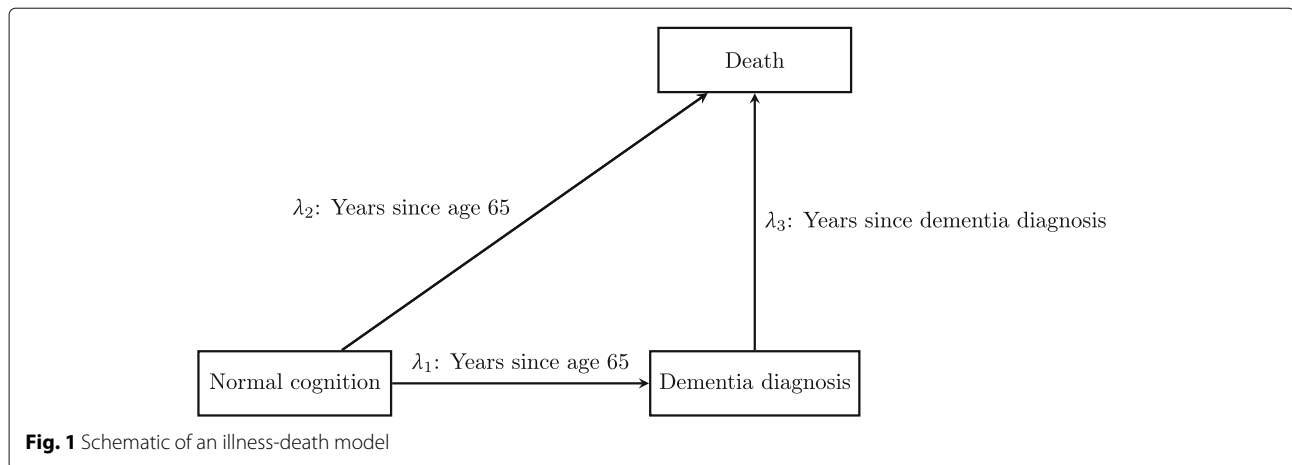
failure time models [32], this paper assumes the first of these which is prevalent in the multistate modeling literature:

$$\lambda_1(t_1|\mathbf{X}_1) = \gamma \cdot \lambda_{1,0}(t_1) \exp\{\beta_1\mathbf{X}_1\}, \quad (4)$$
$$\text{for } t_1 > 0$$

$$\lambda_2(t_2|\mathbf{X}_2) = \gamma \cdot \lambda_{2,0}(t_2) \exp\{\beta_2\mathbf{X}_2\}, \quad (5)$$
$$\text{for } t_2 > 0$$

$$\lambda_3(t_2|T_1 = t_1, \mathbf{X}_3) = \gamma \cdot \lambda_{3,0}(t_2|T_1 = t_1) \exp\{\beta_3\mathbf{X}_3\}, (6)$$
$$\text{for } 0 < t_1 < t_2,$$

where $\lambda_{k,0}(\cdot)$ are baseline hazard functions and $\beta_k$ are log-hazard ratio regression parameters, for $k = 1, 2, 3$. We allow for two common specifications of $\lambda_{3,0}(t_2|T_1 = t_1)$, a Markov model defined by $\lambda_{3,0}(t_2|T_1 = t_1) = \lambda_{3,0}(t_2)$ or a semi-Markov ('clock-reset') model defined by $\lambda_{3,0}(t_2|T_1 = t_1) = \lambda_{3,0}(t_2 - t_1)$. In each of expressions (4)-(6), $\gamma$ is a common subject-specific frailty with mean 1.0 and variance $\theta$, which serves two related purposes. First, similar to a random intercept in a generalized linear mixed model [33], the frailties serve to accommodate between-subject heterogeneity that is not accounted for by the covariates included in the linear predictors. Second, the frailty induces dependence between the non-terminal and terminal events since a patient with frailty larger than one will be at higher risk of both the nonterminal and terminal events than the population average (conditional on covariates).

To complete the specification of the model, the baseline hazard functions and frailty distribution must be specified. For the latter, we adopt a Gamma distribution, which is commonly used because closed-form expressions for the marginal likelihood contributions can be obtained. It is important to note that although the shared frailty term can only account for or represent positive dependence between the hazard functions [20], it is possible to have

negative correlation between the hazard functions corresponding to $T_1$ and $T_2$, e.g. if $\lambda_{1,0}(t)$ is monotonically increasing and $\lambda_{2,0}(t)$ is monotonically decreasing.

The baseline hazard functions can be specified parametrically [34], semi-parametrically [15, 26], or non-parametrically [2]. In this paper, we consider Weibull baseline hazards of the form $\lambda_{k,0}(t) = \alpha_k \kappa_k t^{\alpha_k - 1}$, for $k = 1, 2, 3$, which are commonly used in survival analysis, and flexible B-spline baseline hazard functions that satisfy $\log \lambda(t) = B(t)$, where $B(t)$ is a polynomial B-spline function of degree $d$ with unique knots at $t_0 < t_1 < \cdots < t_K < t_{K+1}$ and defined for $t \in [t_0, t_{K+1}]$. For continuous time-to-event outcomes with support on the positive real line, we let $t_0 = 0$ and $t_{K+1}$ to be the largest follow-up time. Note that the B-spline function $B(t)$ is parametrically defined as a linear combination of B-spline basis functions $B_{b,d}(t)$ of degree $d$,

$$B(t) = \sum_{b=0}^{K+d} \eta_b B_{b,d}(t),$$

where $\eta_b$ are parameters, known as control points, and the B-spline basis functions are defined for $t \in [t_0, t_{K+1}]$ [35].

## Estimation and inference
### Observed data
Let $T_{i1}$ and $T_{i2}$ denote the nonterminal and terminal event times and $\mathbf{X}_i$ a vector of patient-specific covariates that includes $\mathbf{X}_{i1}$, $\mathbf{X}_{i2}$ and $\mathbf{X}_{i3}$ for individual $i$ used in (4)-(6). We assume that the observed data are subject to right-censoring and left-truncation. Let $C_i$ and $L_i$ denote the right-censoring and left-truncation times which we assume are independent of $T_{i1}$ and $T_{i2}$ conditional on $\mathbf{X}_i$. Note that we assume the convention of Xu and colleagues [2], by setting $T_1 = \infty$ for individuals who experience the terminal event in the absence of the nonterminal event. In the analysis of dementia on the age scale, where the

nonterminal and terminal events are dementia diagnosis and death and the left-truncation time is study entry, prevalent dementia cases are typically excluded from the analysis since the primary outcome of interest is dementia [27], i.e. we require that $T_{i1} > L_i$. The observed data for the $i^{th}$ individual is $D_i = \{L_i, Y_{i1}, \delta_{i1}, Y_{i2}, \delta_{i2}, \mathbf{X}_i\}$, where $Y_{i1} = \min(T_{i1}, T_{i2}, C_i)$ with nonterminal event indicator $\delta_{i1} = I\{T_{i1} \leq \min(T_{i2}, C_i)\}$, $Y_{i2} = \min(T_{i2}, C_i)$ with terminal event indicator $\delta_{i2} = I\{T_{i2} \leq C_i\}$, and $L_i < Y_{i1}$.

### Likelihood

Towards developing a form of the likelihood, we first present the joint density, $(T_1, T_2)$, in the absence of left-truncation below [15, 25]:

$$g_{T_1,T_2}(t_1, t_1|X)$$
$$= \begin{cases} \lambda_1(t_1|X)\lambda_3(t_2|t_1, X)S_1(t_1|X)S_2(t_1|X)S_3(t_2|t_1, X), \\ \quad \text{for } 0 < t_1 \leq t_2 \\ \lambda_2(t_2|X)S_1(t_2|X)S_2(t_2|X), \\ \quad \text{for } 0 < t_2 \leq t_1 = \infty, \end{cases} \quad (7)$$

where $S_k(t|X) = \exp\left\{-\int_0^t \lambda_k(u|X)\,du\right\}$, for $k = 1, 2$, and either $S_3(t_2|t_1, X) = \exp\left\{-\int_{t_1}^{t_2} \lambda_3(u|t_1, X)\,du\right\}$ if a Markov model is assumed for $\lambda_3(\cdot)$ or $S_3(t_2|t_1, X) = \exp\left\{-\int_0^{t_2-t_1} \lambda_3(u|t_1, X)\,du\right\}$ if a semi-Markov model is assumed.

Under the assumption of independent truncation, the joint density of left-truncated semi-competing risks data, $(L, T_1, T_2)$, is:

$$\begin{aligned} f_{T_1,T_2,L}(t_1, t_2, l|X) &= f_{T_1,T_2,L}(t_1, t_2, l|L < T_1, X) \\ &= \frac{g_{T_1,T_2}(t_1, t_2|X)g_L(l)}{P(L < T_1|X)}I(l < t_1) \\ &= \frac{g_{T_1,T_2}(t_1, t_2|X)g_L(l)}{\int S_{T_1}(l|X)g_L(l)\,dl}I(l < t_1) \\ &= \frac{g_{T_1,T_2}(t_1, t_2|X)I(l < t_1)}{S_{T_1}(l|X)} \\ &\quad \cdot \frac{S_{T_1}(l|X)g_L(l)}{\int S_{T_1}(l|X)g_L(l)\,dl}, \end{aligned} \quad (8)$$

which is a product of the conditional density of $(T_1, T_2)$ given $L$ and the marginal density of $L$ [24, 25, 36] in (8), where $g_{T_1,T_2}$ is given in (7), $g_L(\cdot)$ is the density function corresponding to $L$ and $S_{T_1}(t|X) = P(T_1 > t|X) = S_1(t|X)S_2(t|X)$. When the distribution of left-truncation times is unknown, the maximum likelihood estimate can be obtained from the conditional likelihood, ignoring the marginal likelihood of the left-truncation times [24, 25, 36, 37].

Based on the observed data, $D_i$, there are four possible data scenarios and thus likelihood contributions corresponding to the distinct combinations of the nonterminal and terminal event indicators, $\delta_1$ and $\delta_2$. Using (7) and

the conditional likelihood expression (on the left) in (8), the observed data likelihood for model parameters $\phi = (\xi_1, \xi_2, \xi_3, \beta_1, \beta_2, \beta_3, \theta)$, where $\xi_k$ are the baseline hazard parameters and $\beta_k$ are the vectors of regression coefficients for transition $k = 1, 2, 3$ and $\theta$ is the Gamma frailty variance, is

$$\begin{aligned} &\mathcal{L}(\phi) \\ &= \prod_{i=1}^n \Big[ f_1(Y_{i1}, Y_{i2}|X_i)^{\delta_{i1}\delta_{i2}} \cdot f_2(Y_{i1}, Y_{i2}|X_i)^{\delta_{i1}(1-\delta_{i2})} \\ &\quad \cdot f_3(Y_{i1}, Y_{i2}|X_i)^{(1-\delta_{i1})\delta_{i2}} \cdot f_4(Y_{i1}, Y_{i2}|X_i)^{(1-\delta_{i1})(1-\delta_{i2})} \Big] \end{aligned}$$

where

$$f_k(Y_{i1}, Y_{i2}|X_i) = \int f_k(Y_{i1}, Y_{i2}|X_i, \gamma_i)f(\gamma_i)\,\partial\gamma_i,$$
$$\text{for } k = 1, \ldots, 4, \quad (9)$$

and

$$\begin{aligned} f_1(Y_{i1}, Y_{i2}|X_i, \gamma_i) &= \lambda_1(Y_{i1}|X_i, \gamma_i)\lambda_3(Y_{i2}|X_i, \gamma_i)S_1(Y_{i1}|X_i, \gamma_i) \\ &\quad S_2(Y_{i1}|X_i, \gamma_i)S_3(Y_{i2}|Y_{i1}, X_i, \gamma_i)/ \\ &\quad [S_1(L_i|X_i, \gamma_i)S_2(L_i|X_i, \gamma_i)] \\ f_2(Y_{i1}, Y_{i2}|X_i, \gamma_i) &= \lambda_1(Y_{i1}|X_i, \gamma_i)S_1(Y_{i1}|X_i, \gamma_i) \\ &\quad S_2(Y_{i1}|X_i, \gamma_i)S_3(Y_{i2}|Y_{i1}, X_i, \gamma_i)/ \\ &\quad [S_1(L_i|X_i, \gamma_i)S_2(L_i|X_i, \gamma_i)] \\ f_3(Y_{i1}, Y_{i2}|X_i, \gamma_i) &= \lambda_2(Y_{i2}|X_i, \gamma_i)S_1(Y_{i2}|X_i, \gamma_i) \\ &\quad S_2(Y_{i2}|X_i, \gamma_i)/[S_1(L_i|X_i, \gamma_i)S_2(L_i|X_i, \gamma_i)] \\ f_4(Y_{i1}, Y_{i2}|X_i, \gamma_i) &= S_1(Y_{i1}|X_i, \gamma_i)S_2(Y_{i1}|X_i, \gamma_i)/ \\ &\quad [S_1(L_i|X_i, \gamma_i)S_2(L_i|X_i, \gamma_i)] \end{aligned}$$

Since the individual-specific frailty terms, $\gamma_i$, are not observed, we marginalize the likelihood with respect to $\gamma_i$. Detailed derivations of the marginal likelihood components in (9) are included in Section A of the Additional file 1.

### Estimation and inference

We will use maximum likelihood for estimation and inference of model parameters using the marginal log-likelihood in Section A of the Additional file 1 [38]. Let $\mathcal{U}(\phi) = \partial/\partial\phi \left(\log \mathcal{L}(\phi)\right)$ denote the score function. Using standard arguments, under certain regularity conditions and a correctly specified model, the maximum likelihood estimator of $\phi$, denoted $\widehat{\phi}$, is the solution to $\mathcal{U}(\phi) = 0$ and is consistent for the true $\phi_0$ as $n \longrightarrow \infty$. In addition, $\sqrt{n}(\widehat{\phi} - \phi_0) \longrightarrow_d \text{MVN}(0, \Sigma)$ as $n \longrightarrow \infty$, where $\Sigma = I(\phi_0)^{-1}$ is the inverse of the expected information matrix:

$$I(\phi_0) = -E\left[\frac{\partial^2}{\partial\phi^2}\log L(\phi)\right]\Big|_{\phi=\phi_0}.$$

$\text{Var}[\widehat{\phi}]$ can be estimated via the inverse of the observed information matrix.

From our experience fitting the proposed model, optimization of the log-likelihood requires careful consideration of the numerical optimization algorithm used and the choice of starting values. For modeling fitting, we used a quasi-Newton non-linear numerical optimization algorithm [39] as implemented in the `optim` function in R [40] to maximize the log-likelihood. For Weibull baseline hazard models, starting values for the non-linear optimization were generated by fitting the univariate Weibull regression models for each transition (4)-(6). For B-spline parameterizations of the baseline hazard functions, we used the `bSpline` function in the `splines2` package [41] in R [40] to generate B-spline basis functions and obtained starting values for hazard parameters as follows: fit univariate Cox models for each transition hazard model; smoothed the estimated cumulative baseline hazard functions using linear interpolation; obtained smoothed baseline hazards function via numerical differentiation followed by loess; and found the control points that minimized the distance between the smoothed log-hazard functions and B-spline functions by least squares. The frailty variance $\theta$ was initialized to value 0.5.

### Simulation study

We performed a set of simulations to investigate the operating characteristics (e.g., bias and coverage) of the model implementation for both Weibull and B-spline parameterizations of the baseline hazard functions.

Data were generated from the model (4)-(6) assuming Weibull baseline hazard functions using `simID` from the `SemiCompRisks` package [42, 43] in R [40]. We set model parameters to those obtained from fitting the model to dementia data, where the nonterminal and terminal events were dementia diagnosis and death, the origin was age 65, the left-truncation time was study entry, and a dichotomous variable was included as a covariate in all three regression models. The baseline hazard and frailty parameter values used in the simulations were: $\log \alpha_1 = 1.05$, $\log \kappa_1 = -9.98$, $\log \alpha_2 = 1.15$, $\log \kappa_2 = -10.01$, $\log \alpha_3 = 0.92$, $\log \kappa_3 = -5.92$, and $\log \theta = -1.39$. The dichotomous covariate was drawn from a Bernoulli distribution with probability 0.57 and the regression coefficients for the three transitions were $\beta_1 = -0.03$, $\beta_2 = -0.33$ and $\beta_3 = -0.11$. Age at study entry was drawn from a uniform distribution with range between 65 and 78 years. We administratively censored at age 95 years.

Based on these set parameters, we simulated $R = 1,000$ datasets of sample size $n = 5,000$. For each simulated dataset, we fit both the model with Weibull and B-spline baseline hazard functions. For the fitted Weibull models, we reported the: mean of the parameter estimates; mean of the estimated analytical standard errors; standard deviation of the distribution of parameter estimates (empirical standard error); and the coverage probability

(the proportion of Wald-based 95% confidence intervals that contained the true parameter). For the fitted B-spline-based models, we reported the operating characteristics listed for the Weibull-based fitted models for the regression and frailty parameters only. To assess the B-spline fit of the transition baseline hazard functions, for each of the three transitions, we plotted the estimated baseline hazards from the fitted models, which we compared to the true baseline hazard. We also compared the estimated baseline hazard functions to kernel smoothed Nelson-Aalen baseline hazard estimates obtained from fitting separate Cox models to each transition using the `coxphHaz` function from the `biostat3` package [44] in R [40]. Note that the univariate Cox models for the 1- and 2-transitions accounted for left-truncation and the Cox model on the 3-transition was performed only among those who experienced the nonterminal event. To quantify the difference between the true baseline hazard function and those estimated from the B-spline model and the smoothed Nelson-Aalen, we calculated the median of the integrated squared error [26].

### Assessing the impact of education level on incident dementia in a large US cohort

We applied the methodological approach presented in "Model specification: Illness-death model" and "Estimation and inference" sections to data from Kaiser Permanente Northern California (KPNC) where the goal of the analysis was to examine the association of education level on incident dementia. The initial analytic cohort consisted of 36,134 individuals who were at least 65 years old, dementia-free by 1997 and KPNC members as of January 1, 1996 (the earliest date when electronic medical records were reliably available defined as study entry) and participated in the Multiphasic Health Checkups which began in the 1960s and continued through 1996 during which information on participant educational attainment was captured. We categorized the exposure of educational attainment level based on the highest level of completion in the following groups: elementary or high school; trade degree, any college or college graduate; and post-graduate study. We focused the analyses on the four most prevalent race/ethnic subgroups, which were White, Black, Asian and Latinx, resulting in a final analytic dataset comprised of 33,117 KPNC members.

We incorporated the strong force of mortality (of all causes) of individuals aged 65 or greater in the analysis via the illness-death model in (1)-(3), where the nonterminal event of interest was incident dementia diagnosis and the terminal event was death. We assumed the semi-Markov or 'clock-reset' approach for the 3-transition model (dementia → death). Follow-up was administratively censored at the earliest of age 90 or end of study defined by September 30, 2017. In our primary

analysis, we used a shared frailty to account for the dependence between incident dementia and death that is not accounted for by covariates. We assumed that time was on the age scale, starting at age 65, and thus left-truncated since members were not followed before age 65. For this time origin, left-truncation was accounted for using the approach in "Estimation and inference" section, assuming B-spline parameterizations of the transition baseline hazard functions. When B-spline baseline hazards are assumed, model fitting requires the specification of the internal knots (number and placement) and polynomial degree of the B-spline function. We fit a suite of models that varied the number of internal knots (we considered 1, 2 or 3 knots), which were placed at equally spaced percentiles of the observed event times, and varied the B-spline polynomial degree (linear, quadratic, cubic). The final model was chosen based on the largest log-likelihood. Time was scaled by a factor of five so that the interpretation of hazard ratios are in terms of 5-year increments. We controlled for sex (reference group: male) and race/ethnicity (reference group: White) in all hazard transition models via covariate adjustment.

In sensitivity analyses, we fit models: 1) without a shared frailty; 2) that defined study entry as the time origin (which we refer to as time-on-study models); and 3) assumed Weibull baseline hazard functions. When the origin was taken to be study entry, for the 1- and 2-transition hazard models we adjusted for age at study entry (centered at age 65); for the 3−transition hazard model, we adjusted for age at dementia (centered at the cohort mean which was 83 years). These models were fit using the `SemiCompRisks` package in R [42].

Note that this was an electronic medical record (EMR)-based study. KPNC is an integrated health care delivery system of approximately 4.3 million members that adopted an EMR system as early as 1993. All healthcare utilization, including care visits (inpatient, outpatient (ambulatory), emergency department, telephone and video visits), pharmacy fills, etc., are captured in the EMR. Dementia diagnoses and death dates were pulled from the EMR. Dementia diagnoses were based on

International Classification of Diseases (ICD) codes that were made during a member visit. Deaths were obtained from the KPNC mortality linkage file which includes data from multiple sources (internal reporting, California state death records, Social Security Administration). Follow-up was not prescribed within the cohort and could vary by individual. However, health system utilization is very high in this population with a median of 97 visits with a physician during the study period (minimum: 0.003 years, median: 15.4 years, maximum: 21.2 years). Details regarding the dataset, including outcome definition, exposure variable and covariates, can be found in the Additional file 1: Section C.1.

## Results

### Simulation study

The operating characteristics of the regression estimates and the frailty parameter are displayed in Table 1. Both modeling approaches resulted in regression estimates with little bias and good coverage. For the Weibull-based model, the estimates of the frailty variance parameter were slightly underestimated with conservative coverage. For the B-spline based model, the mean of the frailty variance parameter was upward downward. The simulation results for the baseline hazard parameters corresponding to the fitted Weibull-based models are displayed in Additional file 1: Table B.1 and exhibit negligible bias and good coverage. In Fig. 2, we display the true baseline hazard functions in red and the estimated baseline hazard functions from the fitted B-spline-based models in the left column and the corresponding kernel smooth Nelson-Aalen estimates from separate univariate Cox models in the right column. Deviations in the estimated baseline hazard functions for both approaches are likely due to data sparsity in the latter half of the follow-up time. The median integrated squared error comparing each of the true baseline hazard functions to the estimated baseline hazard functions over the range of observed event times are displayed in Fig. 2 and indicate that the B-spline-based model provides estimates of the baseline hazard functions that are close to the truth and are

**Table 1** Simulation results

| Parameter | Truth | Weibull-based model | | | | B-spline-based model | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Est. | $SE_a$ | $SE_e$ | Cover. | Est. | $SE_a$ | $SE_e$ | Cover. |
| $\log\theta$ | -1.39 | -1.56 | 0.27 | 0.27 | 0.99 | -0.97 | 0.30 | 0.34 | 0.53 |
| $\beta_1$ | -0.03 | -0.02 | 0.06 | 0.06 | 0.95 | -0.05 | 0.07 | 0.06 | 0.95 |
| $\beta_2$ | -0.33 | -0.33 | 0.04 | 0.04 | 0.95 | -0.36 | 0.05 | 0.05 | 0.93 |
| $\beta_3$ | -0.11 | -0.10 | 0.08 | 0.08 | 0.95 | -0.13 | 0.09 | 0.08 | 0.96 |

One thousand data sets were generated under a model with Weibull baseline hazard functions described in "Simulation study" section with $n = 5,000$ and were fit to models with Weibull and B-spline parameterized baseline hazard functions. Point estimates and analytical standard errors ($SE_a$) were averaged across estimates. Empirical standard errors ($SE_e$) correspond to the standard deviation of the parameter sampling distributions. Coverage was calculated as the proportion of estimated 95% Wald-based confidence intervals that contained the true parameter
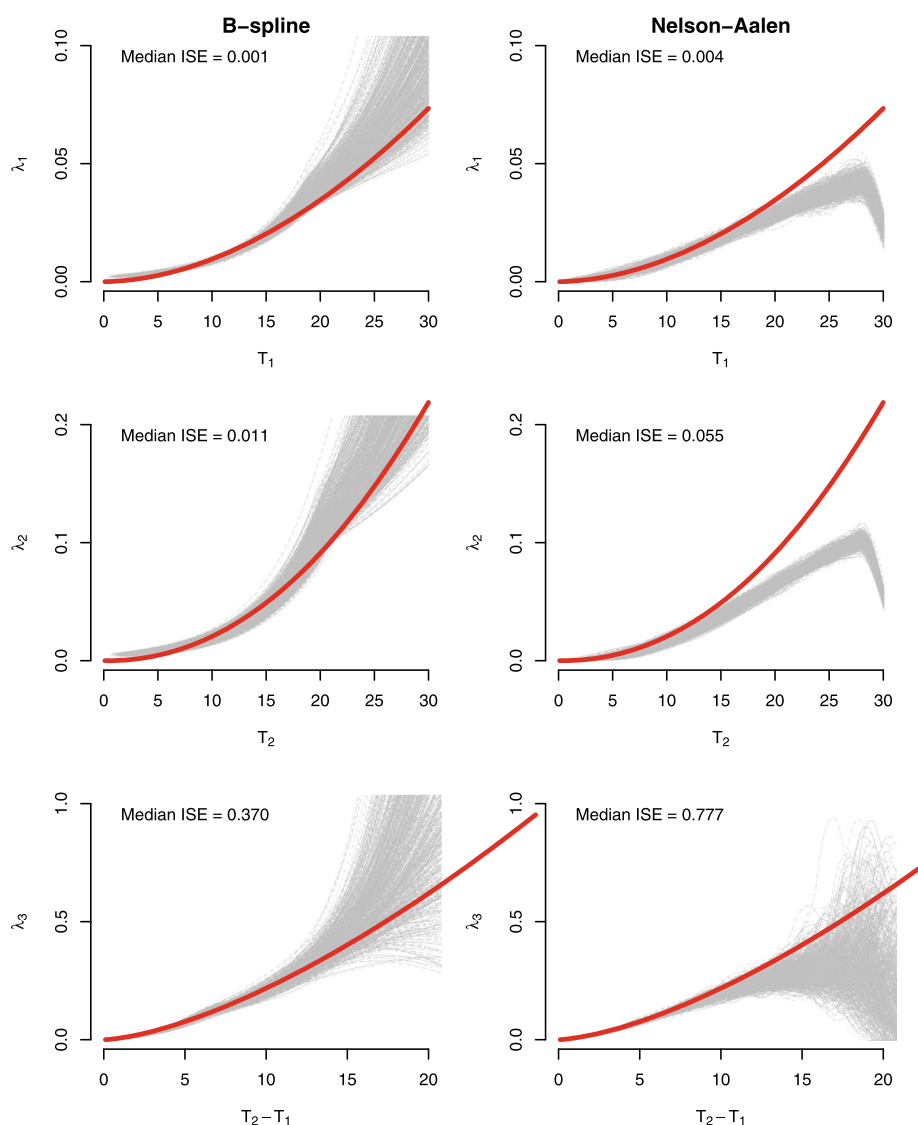
**Fig. 2** Simulation results corresponding to 1,000 simulated dataset. Presented are plots of true Weibull-based baseline hazards (red) for the three transition hazards compared to fitted baseline hazards from the B-spline parameterization and kernel smoothed Nelson-Aalen estimators on three separate Cox models. Median integrated squared error (ISE) comparing each of the true baseline hazard functions to the estimated baseline hazard functions are included

comparable, if not slightly better, than the kernel smooth Nelson-Aalen estimates corresponding to separate Cox models.

### Assessing the impact of education level on incident dementia in a large US cohort

A description of the semi-competing risks data and demographic variables used in the applied analysis are presented in Table 2. Of the 33,117 members in our analytic dataset, elementary or high school was the highest level of education in 43% of members, followed by trade school, some college or a college degree in 40%, and post-graduate study in 16.7%. More than half (56.6%)

were female and 72.0% were of White race. By the end of the study, nearly half (47.1%) of members died without a diagnosis of dementia, with 19.1% censored before dementia or death, 5.0% alive with dementia, and 28.9% died carrying a diagnosis of dementia. In examining the outcome prevalences among the three education level groups, there were more dementia diagnoses and all-cause deaths among members in the elementary and high school group.

The regression estimates from the fitted models are presented in Table 3 and Table C.1 of the Additional file 1. Across all models and time origins, increasing education level was associated with a decreased risk of dementia. In

**Table 2** Description of semi-competing risks data and demographic variables in the analytic cohort from Kaiser Permanente Northern California

| | Total | | Cens before dementia or death | | Alive with dementia | | Died without dementia | | Died with dementia | |
|---|---|---|---|---|---|---|---|---|---|---|
| | n | (%) | n | (%) | n | (%) | n | (%) | n | (%) |
| Total | 33,117 | (100.0) | 6,310 | (19.1) | 1,641 | (5.0) | 15,593 | (47.1) | 9,573 | (28.9) |
| Education | | | | | | | | | | |
| Elementary, High School | 14,325 | (43.3) | 2,330 | (16.3) | 711 | (5.0) | 6,858 | (47.9) | 4,426 | (30.9) |
| Trade degree, College | 13,273 | (40.1) | 2,625 | (19.8) | 666 | (5.0) | 6,259 | (47.2) | 3,723 | (28.0) |
| Graduate | 5,519 | (16.7) | 1,355 | (24.6) | 264 | (4.8) | 2,476 | (44.9) | 1,424 | (25.8) |
| Sex | | | | | | | | | | |
| Female | 18,748 | (56.6) | 3,880 | (20.7) | 1,152 | (6.1) | 7,918 | (42.2) | 5,798 | (30.9) |
| Male | 14,369 | (43.4) | 2,430 | (16.9) | 489 | (3.4) | 7,675 | (53.4) | 3,775 | (26.3) |
| Race/ethnicity | | | | | | | | | | |
| White | 23,830 | (72.0) | 4,381 | (18.4) | 973 | (4.1) | 11,646 | (48.9) | 6,830 | (28.7) |
| Black | 5,172 | (15.6) | 881 | (17.0) | 355 | (6.9) | 2,237 | (43.3) | 1,699 | (32.8) |
| Asian | 2,327 | (7.0) | 630 | (27.1) | 178 | (7.6) | 981 | (42.2) | 538 | (23.1) |
| Latinx | 1,788 | (5.4) | 418 | (23.4) | 135 | (7.6) | 729 | (40.8) | 506 | (28.3) |

the primary analysis (on the age scale with the inclusion of a shared frailty presented in the first two columns of Table 3), the estimated hazard ratio (HR) and 95% confidence interval (CI) was 0.87 (0.83, 0.92) comparing those with a trade degree, some college or a college degree to those who had an elementary or high school education. This protective effect was amplified when comparing those with post-graduate study to those who had an elementary or high school education (HR: 0.76, 95% CI: 0.71, 0.81). There was some variation in the estimated HR across different time origins and models. The hazard ratio comparing those with post-graduate study to those who had an elementary or high school education for the time-on-study model with shared frailty (third column of Table 3) was closer to the null with value 0.80 and to a greater degree in the time-on-study model without shared frailty (last two columns of Table 3) with an estimated HR of 0.83.

Similarly, across all models and time origins, an education level of some college or more compared with a trade degree or less was associated with a decreased risk of death, with estimated HR and 95% CI in the primary model of 0.84 (0.81, 0.88) comparing those with a trade degree, some college or a college degree to those who had an elementary or high school education and 0.68 (0.65, 0.72) comparing those with post-graduate education to those who had an elementary or high school education. In our primary model, increased education level was associated with a reduced risk of death following dementia diagnosis. However, the time-on-study models (columns 3, 4, 7 and 8 of Table 3) provide no evidence

of this association with point estimates that suggest an opposite effect.

In the models with a shared frailty and B-spline baseline hazard functions (first four columns of Table 3), the estimated frailty variance parameter, $\theta$ was 0.41 (0.37, 0.45) when age was the time scale and 0.25 (0.20, 0.31) for the time-on-study analysis, suggesting a greater degree of dependence between time-to-dementia and all-cause death when age is the time scale. When Weibull baseline hazards were assumed (presented in Table C.1 of the Additional file 1), the estimated frailty variance parameter, $\theta$ was comparable with value 0.36 (0.32, 0.40) when age was the time scale. However, the estimated frailty variance was null for the time-on-study analysis. This finding is difficult to explain, but an examination of the estimated baseline hazard functions from the frailty models indicate differences in the underlying baseline risk assumed in the various models. Figure 3 presents the estimated baseline hazard functions under B-spline or Weibull parameterizations and a shared frailty term for two time origins: 1) age 65 (age-scale, adjusted for left-truncation), and 2) study entry (time-on-study). After fitting a suite of models that varied the number of internal knots and the B-spline polynomial degree, the specification with the largest log-likelihood was defined by one internal knot with linear B-splines for the 1- and 2-transitions and two internals knots with cubic B-splines for the 3-transition. We observe that the baseline hazard for the 1- and 2-transitions were similar for the B-spline and Weibull specifications and increasing over time, more so when time was on the age scale (left panel) than time-on-study (right

**Table 3** Estimated regression parameters from analyses of Kaiser data based on an illness-death model with B-spline parameterized baseline hazard functions

| Parameter | With shared frailty | | | | | | Without shared frailty | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Years since age 65 | | | Years since study entry | | | Years since age 65 | | | Years since study entry | | |
| | HR | 95% CI | *p*-value* | HR | 95% CI | *p*-value* | HR | 95% CI | *p*-value* | HR | 95% CI | *p*-value* |
| Frailty variance, $\theta$ | 0.41 | (0.37,0.45) | <.001 | 0.25 | (0.2,0.31) | <.001 | | | | | | |
| Dementia | | | | | | | | | | | | |
| Trade degree, College | 0.87 | (0.83,0.92) | <.001 | 0.89 | (0.85,0.94) | <.001 | 0.88 | (0.84,0.92) | <.001 | 0.91 | (0.87,0.95) | <.001 |
| Post-graduate | 0.76 | (0.71,0.81) | | 0.80 | (0.74,0.85) | | 0.81 | (0.76,0.86) | | 0.83 | (0.78,0.88) | |
| Female | 1.01 | (0.96,1.05) | 0.40 | 1.02 | (0.98,1.07) | 0.17 | 1.00 | (0.95,1.04) | 0.44 | 1.03 | (0.99,1.08) | 0.06 |
| Black | 1.42 | (1.34,1.51) | <.001 | 1.42 | (1.34,1.51) | <.001 | 1.31 | (1.24,1.38) | <.001 | 1.38 | (1.31,1.46) | <.001 |
| Asian | 0.81 | (0.74,0.89) | | 0.90 | (0.82,0.98) | | 0.80 | (0.73,0.87) | | 0.90 | (0.83,0.98) | |
| Latinx | 0.99 | (0.89,1.09) | | 1.07 | (0.97,1.17) | | 1.07 | (0.98,1.16) | | 1.06 | (0.97,1.16) | |
| Age at study entry | | | | 2.03 | (1.98,2.08) | <.001 | | | | 1.97 | (1.93,2.02) | <.001 |
| Death | | | | | | | | | | | | |
| Trade degree, College | 0.84 | (0.81,0.88) | <.001 | 0.85 | (0.81,0.88) | <.001 | 0.84 | (0.81,0.87) | <.001 | 0.88 | (0.85,0.92) | <.001 |
| Post-graduate | 0.68 | (0.65,0.72) | | 0.69 | (0.65,0.73) | | 0.73 | (0.69,0.77) | | 0.75 | (0.71,0.79) | |
| Female | 0.62 | (0.60,0.65) | <.001 | 0.62 | (0.60,0.64) | <.001 | 0.62 | (0.60,0.65) | <.001 | 0.64 | (0.62,0.67) | <.001 |
| Black | 1.02 | (0.96,1.07) | <.001 | 1.05 | (0.99,1.11) | <.001 | 1.03 | (0.98,1.08) | <.001 | 1.01 | (0.96,1.06) | <.001 |
| Asian | 0.70 | (0.65,0.76) | | 0.77 | (0.71,0.83) | | 0.81 | (0.76,0.87) | | 0.76 | (0.70,0.82) | |
| Latinx | 0.72 | (0.65,0.79) | | 0.78 | (0.71,0.85) | | 0.76 | (0.70,0.82) | | 0.77 | (0.71,0.84) | |
| Age at study entry | | | | 1.62 | (1.59,1.66) | <.001 | | | | 1.57 | (1.55,1.60) | <.001 |
| Death following dementia | | | | | | | | | | | | |
| Trade degree, College | 0.90 | (0.85,0.96) | <.001 | 1.02 | (0.96,1.09) | 0.52 | 0.99 | (0.93,1.05) | 0.001 | 1.03 | (0.97,1.10) | 0.23 |
| Post-graduate | 0.77 | (0.71,0.84) | | 1.05 | (0.96,1.15) | | 0.86 | (0.79,0.93) | | 1.07 | (0.99,1.16) | |
| Female | 0.67 | (0.63,0.70) | <.001 | 1.02 | (0.95,1.08) | 0.31 | 0.68 | (0.64,0.71) | <.001 | 1.01 | (0.95,1.07) | 0.37 |
| Black | 0.86 | (0.79,0.92) | <.001 | 1.02 | (0.94,1.10) | 0.19 | 0.95 | (0.88,1.02) | <.001 | 0.99 | (0.92,1.06) | 0.90 |
| Asian | 0.69 | (0.60,0.78) | | 1.09 | (0.97,1.24) | | 0.69 | (0.61,0.78) | | 1.01 | (0.90,1.13) | |
| Latinx | 0.74 | (0.64,0.84) | | 1.08 | (0.95,1.23) | | 0.90 | (0.80,1.02) | | 1.00 | (0.89,1.13) | |
| Age at dementia | | | | 1.01 | (0.98,1.03) | 0.31 | | | | 1.00 | (0.98,1.03) | 0.37 |

Models were fit with and without a shared frailty term. Two time origins were considered: 1) age 65 (age scale, left-truncated data); and 2) study entry (time-on-study)
*p*-value for a categorical variable with $k > 2$ levels (educational attainment, race/ethnicity) is based on a $(k - 1)$-degree of freedom Wald test of linear hypotheses

panel). For the 3-transition (death following dementia), the estimated baseline hazard functions were quite different for the B-spline and Weibull parameterizations, as the Weibull can only accommodate baseline hazards of the form of a power function, while B-splines can approximate a range of functional forms. For both B-spline and Weibull parameterizations, the estimated baseline hazard for the 3-transition indicate that the risk of death is highest shortly after dementia diagnosis, which has been observed in a study of survival after dementia diagnosis in five racial/ethnic groups [45]. This is followed by either decreases then increases (based on the B-spline parameterization) or decreases over time (based on the Weibull parameterization).

## Discussion

For the analysis of left-truncated semi-competing risks data, we have provided methods and software for fitting an illness-death model with shared frailty assuming Weibull or B-spline baseline hazards. These methods were used to estimate the association of education level and dementia accounting for the competing risk of death in a cohort of 33,117 Kaiser Permanente Northern California members. We found a dose-response relationship between educational attainment and incident dementia, with a decreased risk of dementia associated with increasing levels of education, after adjusting for sex and race/ethnicity. The impact of education level and incident dementia is still not well understood, with
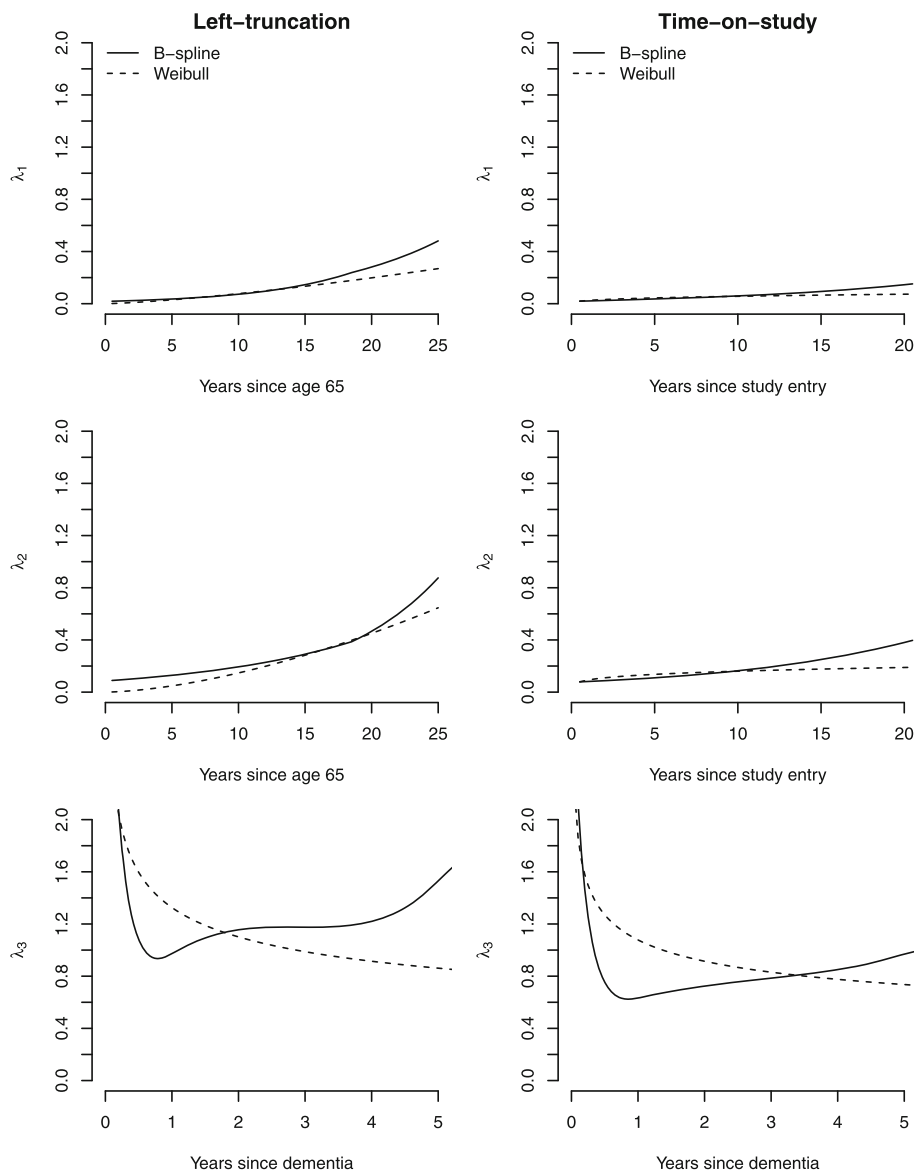
**Fig. 3** Estimated baseline hazard functions from analysis of Kaiser data assuming either B-spline or Weibull parameterized baseline hazard functions and shared frailty term for two time origins: 1) age 65 (age-scale, adjusted for left-truncation), and 2) study entry (time-on-study)

published studies reporting both protective and null effects [46, 47]. Our study supports that higher education is associated with a lower risk of dementia in a large US cohort.

Note that the conclusions drawn for the outcome of interest (dementia) from the illness-death model with shared frailty aligned with those from alternative models that we considered, which omitted the shared frailty or used study entry as the time origin (as shown in Table 3). We expect that there will be variation in the estimates from these models depending on the data at hand, as was observed in the estimates of education level on the risk of death following dementia diagnosis in our study.

In our examination of model fitting operating characteristics via simulated data, we found that the regression parameters were estimated with negligible bias and good coverage. However, on average the frailty variance parameter was slightly underestimated for the Weibull baseline hazard parameterization and overestimated for the B-spline baseline hazard parameterization. For the case of the Weibull baseline hazard parameterization, the coverage was conservative for a sample size 5,000, but was closer to 95% when the sample size was increased 10,000 (see Table B.2 in the Additional file 1). For the case of the B-spline baseline hazard parameterization, the coverage was lower than 95% due to the bias in the

frailty variance parameter. It is important to note that primary interest in the methods presented in this paper are the regression parameters. The frailty, and corresponding frailty variance parameter, allow us to further account for the dependence between the nonterminal and terminal events beyond covariate adjustment, analogous to a random effect in a random effects model. Similar to a random effects model, the primary interest lies in the mean outcome model and regression estimates; the variance parameter of the normally distributed random effects are typically of secondary interest.

In the analysis of dementia diagnoses such as those presented in this paper, prevalent cases at study entry are typically excluded. However, the likelihood in "Likelihood" section can be easily updated (see Additional file 1: Section D) to include prevalent nonterminal cases. Note that in the literature, there are two approaches for handling prevalent nonterminal cases in the analysis of left-truncated semi-competing risks data. The approach we take conditions on the history up to the left truncation time as in [25, 37] so that prevalent nonterminal cases only contribute to the estimation of $\lambda_3$. Estimation is straight-forward assuming an illness-death model with shared frailty since the frailty term, $\gamma$, can be easily integrated out. Alternatively, Saarela and colleagues [24] provided methods for estimation that conditions on the left-truncation time only, so that prevalent cases contribute to the estimation of all transition hazards. This approach is more efficient as it uses more of the data, but is computationally more intensive as it involves numerical integration. This approach does not accommodate an illness-death model with shared frailty well since the integration of the shared frailty term is not straight-forward.

In our approach to fitting an illness-death model with shared frailty to left-truncated semi-competing risks data, we have considered fully-parametric specifications of the baseline hazard functions, Weibull and B-spline. Both functional forms are flexible and can approximate a wide range of baseline hazard functions. At the time of submission, a pre-print by Gorfine et al. [48] proposed a semi-parametric approach to the illness-death model with shared frailty using a pseudo-likelihood approach to estimating the regression parameters and baseline hazard functions that accommodates left-truncated semi-competing risks data. While the illness-death model with shared frailty in this paper was formulated using hazard models that are conditional on the frailty in (4)-(6), Gorfine et al. [48] focused on marginal Cox hazard models. This is analogous to the conditional and marginal approaches to modeling mean outcomes in the presence of clustering via mixed models [49] and generalized estimating equations [50], respectively. Thus our approaches are complementary, filling a gap in the literature and allowing the analyst options for fitting an illness-death model with shared frailty to left-truncated semi-competing risks data.

One of the reviewers pointed out that dementia diagnoses may be subject to interval-censoring. To explore the possibility and/or extent of interval-censoring in the cohort, we looked at the patterns of inpatient and outpatient visits (during which dementia might be assessed) among two groups of members: those who were diagnosed with dementia during the study, and those who died without a dementia diagnosis. The concern is that long gaps between visits would lead to imprecise dementia diagnosis dates in the former and missed opportunities for dementia diagnoses in the latter. We found that among those who were diagnosed with dementia during the study, 81% had a visit with a physician within 60 days prior to the diagnosis date in the EMR. For those who died without a dementia diagnosis, 90% had a visit within 60 days prior to death. Plots of individual-level visit patterns over the study period (see Supplementary File Figures C1 and C2) among members of these two groups illustrate that utilization is high in this cohort.

Based on these data, we believe that interval-censoring is not of major concern in our data, as it is in prospective studies of Alzheimer's disease or dementia, such as PAQUID [51, 52] and the Adult Changes in Thought Study [53, 54], where dementia screening can be years apart. While analyses of data from those prospective studies are indeed complicated by interval-censoring, they were designed for the purpose of understanding incident dementia with identification of dementia cases based on a battery of neuropsychological testing and confirmation by a neurologist. An EMR-based analysis of a cohort of high care utilizers may avoid interval-censoring, however, may capture dementia cases with less rigor. At KPNC, a similar set of EMR code used to identify dementia diagnoses was shown to have a sensitivity of 77% and a specificity of 95% compared with a consensus dementia diagnosis utilizing a neuropsychiatric battery, structured interviews, physical examination, and medical records review. If interval-censoring were evident in our data, modeling should be updated to account for interval-censoring. This can be done by updating the likelihood function, as in Touraine, et al. 2017 [52]. In the setting of a shared frailty illness-death model presented in this paper, this is an avenue for future research.

It is important to mention that we provide methods for fitting a shared frailty illness-death model subject to left-truncated data when the covariate of interest is fixed with time. As one reviewer aptly pointed out, time-varying covariates may also be of interest to an analyst. We believe that the modeling framework specified in this paper can incorporate

time-dependent covariates. However, deriving the likelihood function for a shared frailty model requires marginalization (integration) of the frailty term, which is complicated when time-dependent covariates are used and, such, beyond the scope of this paper. We intend to explore the implementation of the proposed extension with time-varying covariates in future work.

## Conclusions

As illustrated by our analysis of Kaiser data, our proposed modeling framework allows the analyst to assess the impact of covariates on semi-competing risks data, such as incident dementia and death, while accounting for dependence between the outcomes when data are left-truncated, as is common in studies of aging and dementia. This approach has the potential to be applied to a wide range of settings beyond the field of aging.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-020-01203-8.

---

**Additional file 1:** Supplementary findings. File name: `Lee_Gilsanz_Haneuse-Additional file.pdf`. Includes: A: Marginal components of the likelihood referenced in "Likelihood" section; B: Additional simulation results; C: Additional information from the analysis of Kaiser data including: detailed description of the data from Kaiser Permanente Northern California and additional results from the analysis of Kaiser data; D: Likelihood expression when observed data include prevalent nonterminal cases referenced in "Discussion" section.

**Additional file:** r code for implementation. File names: `fit-models.txt` and `functions.txt` provide code for generating simulated semi-competing risks data and fitting the shared frailty illness-death model to left-truncated semi-competing risks data described in the manuscript.

---

### Abbreviations

KPNC: Kaiser Permanente Northern California; HR: Hazard ratio; CI: Confidence interval

### Authors' contributions

CL and SH developed the statistical theory and contributed to successful model implementation. PG curated the analytic dataset, directed the analysis, and interpreted the findings. All authors drafted, reviewed and approved the final manuscript.

### Availability of data and materials

The data used in this study contain Protected Health Information of KPNC members and cannot be shared. Code for generating simulated data used in the paper and that were similar to those analyzed in the data application are provided as additional files.

### Ethics approval and consent to participate

This study was approved by the Kaiser Permanente Northern California Institutional Review Board. This approval allowed for access to electronic health records and administrative and legacy databases for research purposes.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

[1]Kaiser Permanente Northern California, Division of Reseach, 2000 Broadway, Oakland, CA, US. [2]Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA, US.

## References

1. Fine J, Jiang H, Chappell R. On semi-competing risks data. Biometrika. 2001;88(4):907–19.
2. Xu J, Kalbfleisch J, Tai B. Statistical analysis of illness–death processes and semicompeting risks data. Biometrics. 2010;66(3):716–25.
3. Jacqmin-Gadda H, Blanche P, Chary E, Loubère L, Amieva H, Dartigues J-F. Prognostic score for predicting risk of dementia over 10 years while accounting for competing risk of death. Am J Epidemiol. 2014;180(8): 790–8.
4. Alzheimer's Association, et al. 2018 alzheimer's disease facts and figures. Alzheimers Dement. 2018;14(3):367–429.
5. Matthews KA, Xu W, Gaglioti AH, Holt JB, Croft JB, Mack D, McGuire LC. Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged≥65 years. Alzheimers Dement. 2019;15(1):17–24.
6. Brodaty H, Seeher K, Gibson L. Dementia time to death: a systematic literature review on survival time and years of life lost in people with dementia. Int Psychogeriatr. 2012;24(7):1034–45.
7. Varadhan R, Xue Q-L, Bandeen-Roche K. Semicompeting risks in aging research: methods, issues and needs. Lifetime Data Anal. 2014;20(4): 538–62.
8. Peng L, Fine J. Regression modeling of semicompeting risks data. Biometrics. 2007;63(1):96–108.
9. Hsieh J-J, Wang W, Ding A. Regression analysis based on semicompeting risks data. J R Stat Soc B. 2008;70(1):3–20.
10. Lakhal L, Rivest L-P, Abdous B. Estimating survival and association in a semicompeting risks model. Biometrics. 2008;64(1):180–8.
11. Egleston B, Scharfstein D, Freeman E, West S. Causal inference for non-mortality outcomes in the presence of death. Biostatistics. 2007;8(3): 526–45.
12. Tchetgen Tchetgen E. Identification and estimation of survivor average causal effects. Stat Med. 2014;33(21):3601–28.
13. Liu L, Wolfe R, Huang X. Shared frailty models for recurrent events and a terminal event. Biometrics. 2004;60(3):747–56.
14. Putter H, Fiocco M, Geskus R. Tutorial in biostatistics: competing risks and multi-state models. Stat Med. 2007;26(11):2389–430.
15. Lee KH, Haneuse S, Schrag D, Dominici F. Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. J R Stat Soc C. 2015;64(2):253–73.
16. Lee C, Lee SJ, Haneuse S. Time-to-event analysis when the event is defined on a finite time interval. Stat Methods Med Res. 2019;29(6): 1573–91.
17. Klein J, Keiding N, Copelan E. Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients. Stat Med. 1993;12(24):2315–32.
18. de Wreede L, Fiocco M, Putter H. mstate: An R Package for the Analysis of Competing Risks and Multi-State Models. J Stat Softw. 2011. 38(7).
19. Wienke A. Frailty Models in Survival Analysis. Boca Raton: CRC Press, Taylor & Francis Group; 2010.
20. Putter H, van Houwelingen H. Frailties in multi-state models: Are they identifiable? do we need them?. Stat Methods Med Res. 2015;24(6): 675–92.

21. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. Am J Epidemiol. 1997;145(1):72–80.
22. Liestol K, Andersen PK. Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. Stat Med. 2002;21(23):3701–14.
23. Lamarca R, Alonso J, Gomez G, Munoz A. J Gerontol A Biol Sci Med Sci. 1998;53(5):337–43.
24. Saarela O, Kulathinal S, Karvanen J. Joint analysis of prevalence and incidence data using conditional likelihood. Biostatistics. 2009;10(3): 575–87.
25. Vakulenko-Lagun B, Mandel M. Comparing estimation approaches for the illness–death model under left truncation and right censoring. Stat Med. 2016;35(9):1533–48.
26. Joly P, Commenges D, Letenneur L. A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. Biometrics. 1998;54(1):185–94.
27. Commenges D, Letenneur L, Joly P, Alioum A, Dartigues J-F. Modelling age-specific risk: application to dementia. Stat Med. 1998;17(17):1973–88.
28. Joly P, Commenges D, Helmer C, Letenneur L. A penalized likelihood approach for an illness–death model with interval-censored data: application to age-specific incidence of dementia. Biostatistics. 2002;3(3): 433–43.
29. Commenges D, Joly P, Letenneur L, Dartigues J-F. Incidence and mortality of alzheimer's disease or dementia using an illness-death model. Stat Med. 2004;23(2):199–210.
30. Cox D. Regression models and life tables (with discussion). J R Stat Soc Ser B (Methodol). 1972;34:187–220.
31. Aalen O. A linear regression model for the analysis of life times. Stat Med. 1989;8(8):907–25.
32. Wei L-J. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. Stat Med. 1992;11(14-15):1871–9.
33. McCulloch CE, Neuhaus JM. Generalized linear mixed models. Encycl Biostat. 2005;4:220–46.
34. Conlon A, Taylor J, Sargent D. Multi-state models for colon cancer recurrence and death with a cured fraction. Stat Med. 2014;33(10): 1750–66.
35. De Boor C. A Practical Guide to Splines. Applied Statistics, vol. 27. New York: Springer; 1978.
36. Wang M-C, Jewell NP, Tsai W-Y. Asymptotic properties of the product limit estimate under random truncation. Ann Stat. 1986;14(4):1597–605.
37. Guo G. Event-history analysis for left-truncated data. Sociol Methodol. 1993;23:217–243.
38. Ferguson T. A Course in Large Sample Theory. London: Chapman and Hall; 1996.
39. Nocedal J, Wright S. Numerical Optimization. 2nd ed. New York: Springer; 2006.
40. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2017. https://www.R-project.org/.
41. Wenjie W, Jun Y. splines2: Regression Spline Functions and Classes Too. 2017. R package version 0.2.5. https://CRAN.R-project.org/package=splines2. Accessed 15 June 2018.
42. Lee KH, Lee C, Alvares D, Haneuse S. SemiCompRisks: Hierarchical Models for Parametric and Semi-Parametric Analyses of Semi-Competing Risks Data. 2019. R package version 3.3. https://CRAN.R-project.org/package=SemiCompRisks. Accessed 30 Jan 2019.
43. Alvares D, Haneuse S, Lee C, Lee KH. SemiCompRisks: An R Package for the Analysis of Independent and Cluster-correlated Semi-competing Risks Data. R Journal. 2019;11(1):376–400. https://doi.org/10.32614/RJ-2019-038.
44. Karlsson A, Clements M. Biostat3: Utility Functions, Datasets and Extended Examples for Survival Analysis. 2019. R package version 0.1.4. https://CRAN.R-project.org/package=biostat3. Accessed 8 Nov 2019.
45. Mayeda ER, Glymour MM, Quesenberry CP, Johnson JK, Pérez-Stable EJ, Whitmer RA. Survival after dementia diagnosis in five racial/ethnic groups. Alzheimers Dement. 2017;13(7):761–9.
46. Sharp ES, Gatz M. The relationship between education and dementia an updated systematic review. Alzheimer Dis Assoc Disord. 2011;25(4):289.
47. Xu W, Tan L, Wang H-F, Tan M-S, Tan L, Li J-Q, Zhao Q-F, Yu J-T. Education and risk of dementia: dose-response meta-analysis of prospective cohort studies. Mol Neurobiol. 2016;53(5):3113–23.
48. Gorfine M, Keret N, Arie AB, Zucker D, Hsu L. Marginalized Frailty-Based Illness-Death Model: Application to the UK-Biobank Survival Data. J Am Stat Assoc. 2020. https://doi.org/10.1080/01621459.2020.1831922.
49. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982;38(4):963–74.
50. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73(1):13–22.
51. Dartigues J, Gagnon M, Michel P, Letenneur L, Commenges D, Barberger-Gateau P, Auriacombe S, Rigal B, Bedry R, Alperovitch A. The paquid research program on the epidemiology of dementia. methods and initial results. Rev Neurol. 1991;147(3):225–30.
52. Touraine C, Helmer C, Joly P. Predictions in an illness-death model. Stat Methods Med Res. 2016;25(4):1452–70.
53. Kukull WA, Higdon R, Bowen JD, McCormick WC, Teri L, Schellenberg GD, van Belle G, Jolley L, Larson EB. Dementia and alzheimer disease incidence: a prospective cohort study. Arch Neurol. 2002;59(11):1737–46.
54. Lee KH, Rondeau V, Haneuse S. Accelerated failure time models for semi-competing risks data in the presence of complex censoring. Biometrics. 2017;73(4):1401–12.

## Publisher's Note