



HHS Public Access

Author manuscript

Comput Stat Data Anal. Author manuscript; available in PMC 2017 December 01.

Published in final edited form as:

Comput Stat Data Anal. 2016 December ; 104: 197–208. doi:10.1016/j.csda.2016.07.003.

Cause-Specific Hazard Regression for Competing Risks Data Under Interval Censoring and Left Truncation

Chenxi Li

Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824, U.S.A

Abstract

Inference for cause-specific hazards from competing risks data under interval censoring and possible left truncation has been understudied. Aiming at this target, a penalized likelihood approach for a Cox-type proportional cause-specific hazards model is developed, and the associated asymptotic theory is discussed. Monte Carlo simulations show that the approach performs very well for moderate sample sizes. An application to a longitudinal study of dementia illustrates the practical utility of the method. In the application, the age-specific hazards of AD, other dementia and death without dementia are estimated, and risk factors of all competing risks are studied.

Keywords

Competing risks; Cause-specific hazard; Interval censoring; Left truncation; Penalized likelihood; Smoothing parameter selection

1. Introduction

Interval censored failure time data arise widely from longitudinal studies where the occurrence of the failure-defining event can only be detected at periodic study visits. In many such kind of longitudinal studies, multiple types of events could occur to a subject. If those types of events are dependent but preclude each other or only time to the first event is of interest, competing risks issue comes up for the time-to-event analysis. Furthermore, if a subject's follow-up starts later than the time origin of the time-to-event analysis, there is also left truncation that needs to be accounted for. Clinical studies of elder people often give rise to left truncated and interval censored competing risks data. For instance, when studying age to onset of a chronic disease like diabetes, osteoporosis and Alzheimer's disease (AD), the event time is interval censored between two consecutive assessments, death is a competing risk that precludes those clinical endpoints if occurring prior to them, and enrolled study participants are usually required to be free of the disease and of course alive at the entry to follow-up. A real example of such studies is the Memory and Aging Project (MAP) (Bennett

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

et al., 2012). Since 1997, the MAP has recruited more than 1400 older individuals from about 40 retirement communities and senior housing facilities in the Chicago metropolitan area to study how dementia evolves in the elderly. The participants were all dementia-free when they entered the study and had yearly evaluation for dementia during the follow-up, which leads to left-truncated and interval censored age-to-dementia data. Dementia could be categorized into two major types, AD and other dementia, which are competing risks because time to the first incidence of dementia is of scientific interest. Besides, a considerable proportion of these elder participants passed away before they were diagnosed to be demented, making death another competing risk.

Competing risks data under interval censoring and left truncation was first studied by Hudgens et al. (2001), who developed the nonparametric maximum likelihood estimator (NPMLE) for the cumulative incidence function of a failure type. Since then, a deal of attention has been drawn to the inference of the cumulative incidence function with interval censored competing risks data. Relevant works include Jewell et al. (2003), Groeneboom et al. (2008a,b), Li and Fine (2013), Hudgens et al. (2014) and Li (2016) among others. But the important problem of inference of the cause-specific hazard with interval censored competing risks data has not been studied extensively. Along this line, Li and Fine (2013) proposed four cause-specific hazard estimators for current status competing risks data based on smoothing the nonparametric cumulative incidence estimators, and derived their asymptotic distributions. The estimation methods of Li and Fine (2013) can be naturally extended to the setting of mixed case interval censoring. With a slightly different target, Frydman and Liu (2013) developed the NPMLE of the cumulative cause-specific hazard function in an interval censored competing risks model. However, works on inference for cause-specific hazards with interval censored competing risks data, especially regression modeling, are still much needed.

We are aware that several works on inference for multistate models from interval censored data have been published in the past 17 years. Two major papers are Joly and Commenges (1999), which studied a progressive three-state model, and Joly et al. (2002), which studied an illness-death model. The censoring mechanisms for the event history data considered therein are different from the one we consider, though. Specifically, Joly and Commenges (1999) assumes that the second transition is only subject to right censoring; Joly et al. (2002) assumes that the death time is exactly known if a subject dies and that a subject could be ill at death even if s/he is healthy at the last seen time; we focus on the situation where there are several interval censored competing events but any two of them cannot occur within the same interval. Ideally, the penalized likelihood approach in Joly and Commenges (1999) and Joly et al. (2002) can be adopted to analyze competing risks data under interval censoring and left truncation, yet some gaps in practice and theory remain to be filled. In terms of practice, how to minimize the cross-validation criterion to select smoothing parameters was not discussed by either paper, neither was how to determine the number of knots as well as the boundary endpoints for spline smoothing. In terms of theory, there was no theoretical justification for using cubic splines to approximate the intensity functions in the particular penalized likelihoods proposed by the two papers, and neither paper investigated the finite sample performance of the proposed variance estimator and confidence interval for transition intensities. In this article, we present a slightly different penalized likelihood

approach to analyze competing risks data subject to interval censoring and left truncation and address the aforementioned issues. Additionally, we provide some heuristic arguments about the asymptotic theory associated with the methods.

The rest of the paper is organized as follows. Section 2 presents the estimation and inferential procedures for cause-specific hazards with competing risks data under interval censoring and possible left truncation when covariates are available. The associated asymptotic theory is also discussed therein. In Section 3, we conduct extensive simulations to investigate the finite sample performance of the proposed methods. An analysis of the MAP data using the proposed methods is given in Section 4, followed by some concluding discussions in Section 5 that point out several future research directions. The computational details are collected in the Appendix.

2. Cause-Specific Hazard Regression

2.1. Observations

We describe the competing risks data under interval censoring and possible left truncation as follows. Let $(V_1^{(i)}, V_2^{(i)}, \dots, V_{M_i}^{(i)})$ be the sequence of inspection times for subject i ($i = 1, \dots, n$) and $V_0^{(i)}$ be the left truncation time, if any, and 0 otherwise. Define

$\vec{V}_i = (V_0^{(i)}, V_1^{(i)}, \dots, V_{M_i}^{(i)})$. T_j denotes the failure time, K_j denotes the failure cause, and J denotes the number of failure causes. Define

$\Delta_{jk}^{(i)} = I(V_{j-1}^{(i)} < T_i \leq V_j^{(i)}, K_i = k)$, $\Delta_k^{(i)} = (\Delta_{1k}^{(i)}, \dots, \Delta_{M_i k}^{(i)})$ and

$\vec{\Delta}_i = (\Delta_1^{(i)}, \dots, \Delta_J^{(i)})$ ($i = 1, \dots, n$, $j = 1, \dots, M_i$, $k = 1, \dots, J$). \mathbf{Z}_i denotes a d -dimensional vector of time-independent covariates whose effects on the distribution of (T_i, K_i) are of interest. The observable competing risks data under interval censoring and possible left truncation consist of n i.i.d. vectors of $(M_i, \vec{V}_i, \vec{\Delta}_i, \mathbf{Z}_i)$.

2.2. Model and Likelihood

We are interested in estimating the conditional cause-specific hazard functions given the covariates, $\lambda_k(t|\mathbf{Z})$ ($k = 1, \dots, J$). For this purpose, we assume a proportional cause-specific hazards model:

$$\lambda_k(t|\mathbf{Z}) = \lambda_{0k}(t) \exp(\mathbf{Z}^T \boldsymbol{\beta}_k), \quad k = 1, \dots, J, \quad (1)$$

where $\lambda_{0k}(t)$'s are baseline cause-specific hazards and $\boldsymbol{\beta}_k$'s are cause-specific regression parameters.

The development of the likelihood considers two kinds of interval censoring schemes. The first one is a generalization of mixed case interval censoring (Schick and Yu, 2000) to the setting with covariates, which assumes that the inspection process is independent of the failure time and cause given the covariates, that is, for any subject i ,

$$(M_i, \vec{V}_i) \perp (T_i, K_i) | \mathbf{Z}_i. \quad (2)$$

The second censoring scheme is a generalization of the independent inspection process (IIP) model (Lawless, 2003, Section 2.3.1), for which the inspection process stops if any type of failure is detected, to the setting with covariates. Under this censoring scheme, it is assumed that the next inspection time is independent of the failure time and cause given the history of inspection times and failure information as well as the covariates, that is, for any subject i ,

$$V_j^{(i)} \perp (T_i, K_i) | (H_j^{(i)}, \mathbf{Z}_i), \quad (3)$$

where $H_j^{(i)} = (V_0^{(i)}, V_1^{(i)}, \dots, V_{j-1}^{(i)}, \Delta_{11}^{(i)}, \dots, \Delta_{j-1,1}^{(i)}, \dots, \Delta_{1j}^{(i)}, \dots, \Delta_{j-1,j}^{(i)})$ ($j \geq 1$) denotes the history of inspection times and failure information. Hudgens et al. (2014) discussed when the two censoring schemes will be reasonable to assume respectively and gave some illustrating examples. Under either censoring scheme, our work only considers the situation where $P(V_j - V_{j-1} > \varepsilon) = 1$ for some $\varepsilon > 0$, i.e., any two successive observation times are strictly separated.

No matter whether the censoring mechanism is mixed case interval censoring or the independent inspection process, the likelihood of $(M_i, \vec{V}_i \rightarrow_i \mathbf{Z}_i)$ ($i = 1, \dots, n$) under Model (1) equals, up to a multiplicative constant that doesn't depend on $\lambda_0(\cdot) = (\lambda_{01}(\cdot), \dots, \lambda_{0J}(\cdot))$ and $\beta = (\beta_1^T, \dots, \beta_J^T)^T$,

$$L(\beta, \lambda_0(\cdot)) = \prod_{i=1}^n \frac{\left\{ 1 - \sum_{k=1}^J F_k(V_{M_i}^{(i)} | \mathbf{Z}_i) \right\}^{\Delta_{M_i+1}^{(i)}} \prod_{l=1}^{M_i} \prod_{k=1}^J \left\{ F_k(V_l^{(i)} | \mathbf{Z}_i) - F_k(V_{l-1}^{(i)} | \mathbf{Z}_i) \right\}^{\Delta_{lk}^{(i)}}}{1 - \sum_{k=1}^J F_k(V_0^{(i)} | \mathbf{Z}_i)} \quad (4)$$

(Hudgens et al., 2014), where $\Delta_{M_i+1}^{(i)} = 1 - \sum_{l=1}^{M_i} \sum_{k=1}^J \Delta_{lk}^{(i)}$ and

$F_k(t | \mathbf{Z}) = \int_0^t \lambda_{0k}(s) \exp(\mathbf{Z}^T \beta_k) \exp\{-\int_0^s \sum_{j=1}^J \lambda_{0j}(v) \exp(\mathbf{Z}^T \beta_j) dv\} ds$ is the cumulative incidence function for Cause k ($k = 1, \dots, J$). Note that this likelihood assumes all J causes are interval censored and any two competing events do not occur within the same interval. Let L_i and R_i be the last inspection time before the failure and the first inspection time after the failure respectively for the i th subject ($R_i = \infty$ if no failure occurs by $V_{M_i}^{(i)}$). Let $K_i^* = K_i$ if a failure occurs to Subject i by $V_{M_i}^{(i)}$ and $J+1$ otherwise. The likelihood (4) can be reformulated as

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)) = \prod_{i=1}^n \prod_{k=1}^J \left[\int_{L_i}^{R_i} \lambda_{0k}(s) \exp(\mathbf{Z}_i^T \boldsymbol{\beta}_k) \exp \left\{ - \int_{V_0^{(i)}}^s \sum_{j=1}^J \lambda_{0j}(v) \exp(\mathbf{Z}_i^T \boldsymbol{\beta}_j) dv \right\} ds \right]^{I(K_i^* = k)} \left[\exp \left\{ - \int_{V_0^{(i)}}^{L_i} \sum_{j=1}^J \lambda_{0j}(v) \exp(\mathbf{Z}_i^T \boldsymbol{\beta}_j) dv \right\} \right]^{I(K_i^* = J+1)}$$

(5)

2.3. Penalized Likelihood Estimation

Unlike competing risks data subject to right censoring, there is no counterpart of partial likelihood for interval censored competing risks data which profiles out baseline cause-specific hazards for estimating regression parameters. However, when the cause-specific hazards are expected to be smooth, we can estimate $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}_0(\cdot)$ simultaneously by maximizing a penalized log likelihood. Specifically, we penalize the log likelihood by subtracting a term which has large values for rough baseline cause-specific hazard functions and then search for the minimizer $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}_0(\cdot))$ of the minus penalized log likelihood in the parameter space of $(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot))$. The estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\lambda}}_0(\cdot)$ obtained this way are the so-called maximum penalized likelihood estimators (MPLE). In the sequel, we assume that $\lambda_{0k}(\cdot)$'s are strictly positive over the interval $[\tau_1, \tau_2]$ with τ_1 being the lower bound of V_0 's support and τ_2 being the upper bound of V_M 's support, and $\lambda_{0k}(\cdot)$'s have first derivatives with finite L^2 -norms over $[\tau_1, \tau_2]$. The roughness penalty function in the penalized likelihood is then chosen to be the sum of the squared L^2 -norms of the first derivatives of the baseline hazards. Thus, the minus penalized log likelihood is defined as

$$-pl(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)) = -n^{-1}l(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)) + \sum_{k=1}^J h_k \int_{\tau_1}^{\tau_2} \{\lambda_{0k}^{(1)}(t)\}^2 dt, \quad (6)$$

where $l(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)) = \log L(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot))$ and h_k 's are smoothing parameters. The first derivatives of the baseline hazards are chosen for the penalization so that (6) can equal

$$-pl(\boldsymbol{\beta}, \boldsymbol{\Lambda}_0^{(1)}(\cdot|\tau_1)) = -n^{-1}l(\boldsymbol{\beta}, \boldsymbol{\Lambda}_0^{(1)}(\cdot|\tau_1)) + \sum_{k=1}^J h_k \int_{\tau_1}^{\tau_2} \{\Lambda_{0k}^{(2)}(t|\tau_1)\}^2 dt, \quad (7)$$

where $\Lambda_{0k}(t|\tau_1) = \int_{\tau_1}^t \lambda_{0k}(s) ds$.

In the absence of competing risk i.e. $J=1$, $l(\boldsymbol{\beta}, \boldsymbol{\Lambda}_0^{(1)}(\cdot|\tau_1))$ depends on $\Lambda_0(\cdot|\tau_1)$ through its values at $\{V_0^{(i)}, L_i; i=1, \dots, n\} \cup \{R_i; R_i < \infty, i=1, \dots, n\}$, and thus Theorem 2.3 in Green and Silverman (1993) implies that the MPLE $\hat{\Lambda}_0(\cdot|\tau_1)$ is a natural monotone cubic spline

over $[\tau_1, \tau_2]$ with interior knots being $T \equiv$ the unique values of

$\{V_0^{(i)}, L_i: i=1, \dots, n\} \cup \{R_i: R_i < \infty, i=1, \dots, n\}$. As a result, the MPLE $\hat{\lambda}_0(\cdot)$ is a nonnegative quadratic spline over $[\tilde{\tau}_1, \tilde{\tau}_2]$, where $\tilde{\tau}_1 = \min T$ and $\tilde{\tau}_2 = \max T$, with interior knots being $T \setminus \{\tilde{\tau}_1, \tilde{\tau}_2\}$, and it is constant over $[\tau_1, \tilde{\tau}_1]$ and $[\tilde{\tau}_2, \tau_2]$ respectively. In practice, τ_1 and τ_2 in the penalty term are often unknown and replaced by $\tilde{\tau}_1$ and $\tilde{\tau}_2$. Consequently, $\lambda_0(\cdot)$ is estimated only on $[\tilde{\tau}_1, \tilde{\tau}_2]$. To reduce computational burden, one can approximate $\hat{\lambda}_0(\cdot)$ over $[\tilde{\tau}_1, \tilde{\tau}_2]$ by quadratic spline with a subset of the knots needed for the full MPLE. According to Section 3.5 of Gu (2013), the approximation has the same asymptotic convergence rate as the full MPLE provided that its number of knots is sufficiently large, which will be discussed in Section 2.5.

In the presence of competing risks, there is no closed form for the MPLE $\hat{\lambda}_0(\cdot)$, but this can be rescued by approximating $\hat{\lambda}_0(\cdot)$ over $[\tilde{\tau}_1, \tilde{\tau}_2]$ by quadratic splines with interior knots being a subset of $T \setminus \{\tilde{\tau}_1, \tilde{\tau}_2\}$ as in the situation without competing risk. M-splines have been proposed to approximate $\hat{\lambda}_0(\cdot)$ for penalized likelihood hazard estimation with interval censored data (see, e.g., Joly et al., 1998; Joly and Commenges, 1999; Joly et al., 2002), because the corresponding $\hat{\Lambda}_0(\cdot | \tilde{\tau}_1)$ in the likelihood are then approximated by I-splines with the same spline coefficients as in the M-splines. We thus adopt quadratic M-splines to approximate $\hat{\lambda}_0(\cdot)$. Specifically, $\hat{\lambda}_0(\cdot)$ is approximated by J linear combinations of q_n quadratic M-spline basis functions over $[\tilde{\tau}_1, \tilde{\tau}_2]$,

$$\tilde{\lambda}_{0k}(\cdot) = \tilde{\theta}_k^T \mathbf{M}(\cdot), \quad k=1, \dots, J, \quad (8)$$

where $\tilde{\theta}_k = (\tilde{\theta}_{k1}, \dots, \tilde{\theta}_{kq_n})^T$ ($k=1, \dots, J$) is a vector of spline coefficients and $\mathbf{M}(\cdot) = (M_1(\cdot), \dots, M_{q_n}(\cdot))$ is a base of quadratic M-spline functions on $[\tilde{\tau}_1, \tilde{\tau}_2]$. Here we use a common set of knots $\tilde{\tau}_1 = d_0 < d_1 < \dots < d_{B_n} < d_{B_n+1} = \tilde{\tau}_2$ for approximating $\hat{\lambda}_{0k}(\cdot)$'s of different failure causes by M-splines in order to simplify the computation. $\{d_i\}_{i=1}^{B_n}$ are put at every p_n values of the ordered $T \setminus \{\tilde{\tau}_1, \tilde{\tau}_2\}$, where $p_n = \lceil (|T| - 1)/(B_n + 1) \rceil$. Note that $q_n = B_n + 3$. B_n is chosen to be $O(n^\nu)$ for some constant $\nu \in (0, 1]$. A small ν is computationally preferred, but too small a ν could result in model bias. How small it can be depends on the smoothness of $\lambda_0(\cdot)$, which will be discussed in Section 2.5. According to Ramsay (1988), we constrain $\tilde{\theta}_k$'s to be nonnegative so that $\tilde{\lambda}_{0k}(\cdot)$'s are nonnegative functions. The estimators of β and $\lambda_0(\cdot)$, denoted by $\tilde{\beta}$ and $\tilde{\lambda}_0(\cdot)$, are then obtained by minimizing (6) with respect to β and $\theta = (\theta_1^T, \dots, \theta_J^T)^T$ subject to the constraint $\theta \geq 0$. The details on the optimization such as the choice of initial values and the integral evaluations in the likelihood are provided in the Appendix.

2.4. Smoothing Parameter Selection

We select the smoothing parameters in (6) by minimizing the leave-one-out likelihood cross-validation criterion:

$$V(\mathbf{h}) = -n^{-1} \sum_{i=1}^n l_i(\tilde{\zeta}_{-i}(\mathbf{h})), \quad (9)$$

where $\mathbf{h} = (h_1, \dots, h_j)^T$, $\zeta = (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T$, $\tilde{\zeta}_{-i}(\mathbf{h})$ is the MPLE of ζ given smoothing parameter \mathbf{h} based on the sample in which the i th subject is removed, and l_i is the log likelihood contribution of this subject. Commenges et al. (2007) showed that the leave-one-out likelihood cross-validation criterion is a possible estimator up to a constant of the expected Kullback-Leibler divergence.

Direct calculation of (9) is computationally burdensome. Following O'Sullivan (1988), an approximation formula for (9) can be obtained:

$$V(\mathbf{h}) \approx -n^{-1} [l(\tilde{\zeta}(\mathbf{h})) - \text{trace}\{H_{pl}(\tilde{\zeta}(\mathbf{h}), \mathbf{h})^{-1} H_l(\tilde{\zeta}(\mathbf{h}))\}] \quad (10)$$

where $H_{pl}(\zeta, \mathbf{h}) = \frac{\partial p_l(\zeta)}{\partial \zeta \partial \zeta^T}$, $H_l(\zeta) = n^{-1} \frac{\partial l(\zeta)}{\partial \zeta \partial \zeta^T}$ and $\tilde{\zeta}(\mathbf{h}) = (\tilde{\boldsymbol{\beta}}(\mathbf{h})^T, \tilde{\boldsymbol{\theta}}(\mathbf{h})^T)^T$ is the MPLE of ζ given smoothing parameter \mathbf{h} based on the whole sample. The \mathbf{h} in the parentheses will be omitted in the sequel for a compact display. The computational details on minimizing (10) are given in the Appendix.

2.5. Inference

It is hard to rigorously derive the large sample theory associated with the above penalized likelihood approach. Much of the following argument is heuristic based on the relevant literature with the focus on the utility of the asymptotics to the inference for $\boldsymbol{\beta}$ and $\lambda_0(\cdot)$ with finite samples.

In the sequel, $\lambda_{0k}(\cdot)$'s are assumed to have at most c th-order ($c \geq 1$) derivatives that are square integrable over $[\tau_1, \tau_2]$. To sketch the large sample theory, we first discuss the determination of the number of interior knots B_n in light of the equivalence between (6) and (7). According to Section 3.5.4 of Gu (2013), the smoothing parameter selected by cross validation, $\tilde{\mathbf{h}}$, is of order $n^{-4/(4p+1)}$ for some $p \in [1, 2]$ depending on c if $q_n^{-1} = o(n^{-2/(4p+1)})$, and the corresponding $\tilde{\lambda}_0(\cdot)$ achieves the optimal convergence rate in L^2 -norm over $[\tau_1, \tau_2]$. Combining with Stone (1982), one would expect that this optimal convergence rate is $n^{-(2p-1)/(4p+1)}$ and that $p = (c+1)/2$ for $1 \leq c < 3$ and $p = 2$ for $c = 3$. This indicates that it suffices to have $\nu = 2/9$ in practice for $\tilde{\lambda}_{0k}(\cdot)$'s to achieve the optimal convergence rate if $\lambda_{0k}(\cdot)$'s have square integrable third-order derivative over $[\tau_1, \tau_2]$. Considering $n \rightarrow \infty$, we set $B_n = \max\{\lceil (3n)^{2/9} \rceil, 5\}$ in the subsequent simulation study and real data analysis. With the above $\tilde{\mathbf{h}}$ and q_n , when n is large, $\tilde{\zeta} - \zeta^{(n)}$ has an approximate multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $-n^{-1} H_{p_l}(\tilde{\zeta}, \tilde{\mathbf{h}})^{-1} H_l(\tilde{\zeta}) H_{p_l}(\tilde{\zeta}, \tilde{\mathbf{h}})^{-1}$ based on Gray (1992), where $\zeta^{(n)} = (\boldsymbol{\beta}^{(n)}, \boldsymbol{\theta}^{(n)})^T$ is the value of ζ that maximizes the expectation of $p_l(\zeta)$. From a Bayesian viewpoint, similar arguments to Section 5 of O'Sullivan (1988) can show that when n is large, ζ has an approximate Gaussian posterior distribution with mean $\tilde{\zeta}$ and covariance matrix $-n^{-1} H_{p_l}(\tilde{\zeta}, \tilde{\mathbf{h}})^{-1} - n^{-1} H_{p_l}(\tilde{\zeta}, \tilde{\mathbf{h}})^{-1} H_l(\tilde{\zeta}) H_{p_l}(\tilde{\zeta}, \tilde{\mathbf{h}})^{-1}$ and $-n^{-1} H_{p_l}(\tilde{\zeta}, \tilde{\mathbf{h}})^{-1}$

are asymptotically equivalent, but Rondeau et al. (2003) showed that the latter performs better with finite samples in the sense that it is closer to the empirical variance of the MPLE for finite dimensional parameters and its associated Wald confidence intervals for finite dimensional parameters have empirical coverage probability closer to the nominal level. Therefore, we used $-n^{-1}H_{p\lambda}(\tilde{\boldsymbol{\zeta}}, \tilde{\mathbf{h}})^{-1}$ in the simulation and the real data application.

With a smoothing parameter selected by cross validation, the regression parameter estimator from the penalized likelihood approach for semiparametric models usually has a bias of larger order than $\alpha(n^{-1/2})$ (Rice, 1986). One exception is that the covariates of the regression parameters are uncorrelated with the covariate of the infinite dimensional parameter (Rice, 1986). In our situation, that is equivalent to assuming

$$(M_i, \vec{V}_i) \perp \mathbf{Z}_i, \quad i=1, \dots, n. \quad (11)$$

In other words, the observation times are determined regardless of the explanatory variable values, which is true in most longitudinal studies. Under this assumption, the bias $\boldsymbol{\beta}^{(n)} - \boldsymbol{\beta}$ is of order $\alpha(n^{-1/2})$, so $\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normal with mean $\mathbf{0}$ and a covariance matrix that can be consistently estimated by the upper left $Jd \times Jd$ block of $-H_{p\lambda}(\tilde{\boldsymbol{\zeta}}, \tilde{\mathbf{h}})^{-1}$. One is able to construct Wald tests and confidence intervals for $\boldsymbol{\beta}$ based on this asymptotic distribution. Since $-nH_{p\lambda}(\tilde{\boldsymbol{\zeta}}, \tilde{\mathbf{h}})$ is the observed information matrix of the penalized likelihood, $\tilde{\boldsymbol{\zeta}}$ is semiparametrically efficient under (11).

To assess the estimation precision of $\tilde{\boldsymbol{\lambda}}_0(\cdot)$, one could use the posterior distribution of $\boldsymbol{\theta}$ derived above to construct approximate Bayesian confidence intervals for $\boldsymbol{\lambda}_0(\cdot)$. Specifically, the $100(1-\alpha)\%$ Bayesian confidence interval for $\boldsymbol{\lambda}_k(t)$ is

$$\tilde{\boldsymbol{\theta}}_k^T \mathbf{M}(t) \pm z_{\alpha/2} \sqrt{\mathbf{M}(t)^T \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}}_k} \mathbf{M}(t)}, \quad k=1, \dots, J, \quad (12)$$

where $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}}_k}$ is the block in $-n^{-1}H_{p\lambda}(\tilde{\boldsymbol{\zeta}}, \tilde{\mathbf{h}})^{-1}$ corresponding to $\tilde{\boldsymbol{\theta}}_k$. This Bayesian technique for generating confidence intervals was proposed by O'Sullivan (1988) for log hazard estimation with right censored data. Based on Nychka (1988), we expect that when the sample size is large, the average coverage probability of these pointwise confidence intervals across the knots should be close to the nominal level. In other words, for n large,

$$E[\text{ACP}_k(\alpha)] \approx 1-\alpha, \quad k=1, \dots, J, \quad (13)$$

where

$$ACP_k(\alpha) = \frac{\#\{i: |\tilde{\lambda}_{0k}(d_i) - \lambda_{0k}(d_i)| \leq z_{\alpha/2} \sqrt{\mathbf{M}(d_i)^T \Sigma_{\hat{\theta}_k} \mathbf{M}(d_i)}, i=0, \dots, B_n+1\}}{B_n+2}.$$

In practice, the lower bound might be negative in which case we replace it by 0.

3. Simulations

We investigate the finite sample performance of $\tilde{\beta}$ and $\tilde{\lambda}_0(\cdot)$ as well as their variance estimators and the corresponding confidence intervals through simulations. The failure time and cause were generated from the following two cause-specific hazard functions using the method of Beyersmann et al. (2009),

$$\lambda_1(t|Z_1, Z_2) = \lambda_{01}(t) \exp(0.3Z_1 - 0.15Z_2)$$

and

$$\lambda_2(t|Z_1, Z_2) = \lambda_{02}(t) \exp(0.1Z_1 - 0.4Z_2),$$

where Z_1 is a Bernoulli random variable with probability 0.6 of being 1, Z_2 has a uniform distribution over $[0, 1]$ denoted by $U(0, 1)$, Z_1 and Z_2 are independent, and the two baseline cause-specific hazards are

$$\lambda_{01}(t) = \frac{0.168(0.012t)^{19} e^{-(0.012t)^{20}} + 0.2392(0.0092t)^{19} e^{-(0.0092t)^{20}}}{0.35e^{-(0.012t)^{20}} + 0.65e^{-(0.0092t)^{20}}}$$

and

$$\lambda_{02}(t) = \frac{0.252(0.012t)^{19} e^{-(0.012t)^{20}} + 0.3588(0.0092t)^{19} e^{-(0.0092t)^{20}}}{0.35e^{-(0.012t)^{20}} + 0.65e^{-(0.0092t)^{20}}},$$

which are obtained by multiplying the hazard of a mixture of Weibull distributions, $0.35 * \text{Weibull}(0.012, 20) + 0.65 * \text{Weibull}(0.0092, 20)$, by 2 and 3 respectively. The left truncation time was generated by $V_0 = 80 + U(0, 10)$. Only subjects with $T > V_0$ entered the sample for analysis. Six potential inspection times were generated for each subject of the sample with the first one being $V_0 + U(1.8, 2.2)$, the second one being $V_0 + U(3.8, 4.2)$, and so on. A subject could miss an inspection with probability 0.05. This simulation scenario is to mimic the design features of the MAP study. From this set-up, it is easy to see that $\tau_1 = 80$ and $\tau_2 = 102.2$. The set-up also leads to that about 31.5% of the subjects in a sample will have Type 1 failure observed, about 37.5% will have Type 2 failure observed, and the rest will have no failure by the end of follow-up. We did two simulations with sample size 500 and 750

respectively. In each simulation, we generated 1000 samples. The smoothing parameters were selected by cross validation for the first sample of every simulation and the same value was used for the rest samples in that simulation. The theoretical standard errors for $\tilde{\beta}$ and $\tilde{\lambda}_{0k}(\cdot)$ were computed based on $-n^{-1}H_p(\tilde{\zeta}, \tilde{\mathbf{h}})^{-1}$ as described in Section 2.5.

All the numerical experiments were run in R-3.1.1. The computation is demanding, especially for the cross validation. We thus use parallel computing for evaluating the log likelihood and its gradient, allocating subjects to eight CPU cores that compute simultaneously. On a PC with a 3.40 GHz CPU and 8GB RAM, the smoothing parameter selection takes about 7.4 and 16.2 hours for $n = 500$ and 750 respectively, and the model fitting including covariance matrix estimation takes about 9.6 and 14.7 minutes. However, we expect that these times will be shortened remarkably if a better programmer codes the methods or better optimization engines are used.

The cross-validation selected $\tilde{\mathbf{h}}$ is (0.442, 0.330) and (0.153, 0.094) for $n = 500$ and 750 respectively. The decrease of $\tilde{\mathbf{h}}$ with n is consistent with the asymptotic theory on $\tilde{\mathbf{h}}$. The baseline hazard of the second cause received less penalty than that of the first, perhaps because the two baseline hazards have similar shape whereas there were more failures from Cause 2.

Table 1 presents the results for estimating β . Except that when $n = 500$, there are bias in estimating β_{12} and β_{22} and a bit of undercoverage by the confidence interval for β_{21} , the penalized likelihood approach performed very well in estimating β . Additionally, the standard error ratios between the two sample sizes stay around $(750/500)^{1/2} \approx 1.225$, as implied by the asymptotic distribution of $\tilde{\beta}$. We thus conclude that $n = 750$ is large enough for using the asymptotic distribution of $\tilde{\beta}$ to perform inference on β in the simulation scenario.

Figure 1 shows the averaged $\tilde{\lambda}_{0k}(\cdot)$'s and Bayesian confidence intervals (12) for $\lambda_{0k}(\cdot)$'s across the 1000 Monte Carlo samples as well as the true $\lambda_{0k}(\cdot)$'s. The time range is taken to be $[\max(\tilde{\tau}_1), \min(\tilde{\tau}_2)]$ over the Monte Carlo samples. One can see that the bias decreases with sample size and when $n = 750$, $\tilde{\lambda}_{0k}(\cdot)$ ($k = 1, 2$) has very little bias except near the boundaries where the number of observed failures is small. The widths of the confidence intervals appear to be of proper magnitude except near the boundaries where information from the data being vanishing widens the intervals. Figure 2 shows that the theoretical standard errors of the baseline cause-specific hazard estimates are very close to the empirical ones except again near the boundaries especially the left endpoint. Note that the theoretical standard errors are slightly larger than the empirical ones at most time points. This is because as Bayesian posterior standard deviations, the theoretical standard errors take the bias of $\tilde{\lambda}_{0k}(\cdot)$ ($k = 1, 2$) into account. Figure 3 shows the pointwise empirical coverage and the average coverage probabilities (13) of the 95% Bayesian confidence intervals for $\lambda_{0k}(\cdot)$'s. The empirical coverage approaches 95% as the sample size increases. All average coverage probabilities agree well with the nominal level. Figures 1, 2 and 3 together indicate that in the simulation scenario, $n = 750$ is large enough for the Bayesian inference on $\lambda_{0k}(\cdot)$'s to have good frequentist properties.

We also performed similar simulations with $B_n = \max\{\lceil n^{2/9} \rceil, 5\}$. The simulation results, not provided here due to space limit, showed bigger bias of $\tilde{\beta}$ and $\tilde{\lambda}_0(\cdot)$ for $n = 500$ compared to the results with $B_n = \max\{\lceil (3n)^{2/9} \rceil, 5\}$; when $n = 750$, the results under these two different numbers of knots are quite similar. Thus we recommend using $B_n = \max\{\lceil (3n)^{2/9} \rceil, 5\}$ in practice.

4. Application to the MAP Data

4.1. Data

We used the MAP data to investigate the effects of years of education (categorized into two levels: ≤ 12 years and > 12 years), gender, and the presence of the apolipoprotein E $\epsilon 4$ allele (ApoE4) on ages to incidences of AD, other dementia and death without dementia. The data set we used from the MAP study was frozen in 2010 with a sample size of 1168. We first excluded from that data set the 87 participants who had only baseline visit and were not known to have died, then removed one subject with $L = R$ due to measurement error in age at visit, and finally deleted the 104 subjects with missing ApoE4 status. The resulting sample for the analysis has 976 subjects. A summary of characteristics of the sample is presented in Table 2.

4.2. Statistical Analysis

We aim to estimate the cause-specific hazards of AD, other dementia and death without dementia across age adjusting for education, gender and ApoE4 status. Model (1) is assumed for all cause-specific hazard functions with t representing age. The fact that death terminates the follow-up makes neither mixed case interval censoring nor the IIP model applicable to the inspection process of the MAP study. Nevertheless, according to the study design, those deceased subjects in our data set would have dementia examinations approximately one year apart until 2010 had they not died ahead of that year (based on the data, study subjects miss a scheduled visit rarely). Therefore, it is reasonable to treat the subjects who deceased before developing dementia as being interval censored between the last visit before death and the first visit after death, which did not happen but was scheduled. This scheduled visit is m years apart from the last visit before death where m is a positive integer and most often equal to 1. The resulting data are age-to-event data of the competing risks, AD, other dementia and death, under interval censoring and left truncation, for which the likelihood is of the form (5) with $J = 3$. It loses statistical efficiency to treat death as being interval censored rather than use its exact time. This interval censoring treatment was performed in order to estimate the AD, other dementia and death hazards from the MAP data using the proposed methods. Since the time intervals bracketing death are mostly 1 year in length in the analytic sample (138 out of 181 death bracketing intervals), the efficiency loss is expected to be small. Annually scheduled dementia examinations also decrease the chance that dementia incidence was not caught in subjects who died after developing dementia.

Based on the analytic sample, the visit times range from 54.34 to 107.50 years in age scale. We estimated the AD, other dementia and death hazards in this age window with the cross-validation selected smoothing parameters being 41.3, 492.4 and 36.1 respectively. The baseline hazard of other dementia received a much larger roughness penalty because of few

such events. Table 3 shows the estimated cause-specific hazard ratios for the three covariates. The confidence intervals and two-sided p-values are based on the asymptotic normal distribution of the regression parameter estimators. The table indicates that controlling for the other covariates, carrying ApoE4 significantly increases the AD incidence rate, men are more likely to develop AD than women, and non-demented men have a significantly higher risk of death than non-demented women. Specifically, across the age period from 54.34 to 107.50 years, ApoE4 carriers have about two-fold the risk of developing AD that ApoE4 non-carriers have adjusting for gender and college attendance, the AD incidence rate is 44% higher in men than in women adjusting for college attendance and ApoE4 status, and the death risk for non-demented men is roughly 1.8 times that for non-demented women adjusting for college attendance and ApoE4 status. Due to the small number of people developing other dementia first, the three regression parameter estimators for other dementia have very large standard errors. Figure 4 shows the estimated cause-specific hazards for AD, other dementia and death without dementia from age 54.34 to 107.50 for a woman who didn't attend college and doesn't carry ApoE4, which are baseline hazards in Model (1). The incidence rates of AD and death appear to increase dramatically after ages 80 and 84 respectively.

5. Discussion

We studied the penalized likelihood estimation for the Cox-type cause-specific hazard regression with left truncated and interval censored competing risks data. The inferential procedures for regression parameters and baseline hazards were proposed based on the heuristically derived asymptotics, and demonstrated satisfactory finite sample performances in the simulations when the sample size reaches 750. Besides sample size, the number of events for each competing risk as well as the failure bracketing interval length also affect the precision of hazard estimation. Low number of events for a competing risk leads to large variance for the corresponding baseline hazard and regression parameter estimators, so do wide failure bracketing intervals, which are often caused by a high rate of missing scheduled visits. Power calculation for treatment effect evaluation based on interval censored competing risks data should take these factors into account.

All proposed methods apply directly to interval censored data without competing risks or left truncation or either. By modifying the likelihood and carefully choosing the spline order, the latter needing further investigation, our methods can also deal with right censored data with or without competing risks and even the competing risks data where failures from certain causes could be observed exactly but other types of failure are interval censored. To obtain the likelihood of the partly interval censored competing risks data, one just needs to

replace $\prod_{l=1}^{M_i} \{F_k(V_l^{(i)} | \mathbf{Z}_i) - F_k(V_{l-1}^{(i)} | \mathbf{Z}_i)\}^{\Delta_{lk}^{(i)}}$ in (4) by $\{f_k(T_i)\}^{I(K_i^* = k)}$, where $f_k(t) = dF_k(t)/dt$, if failure from cause k is observed exactly.

Several other future research directions are worth pursuing as well. The first is to rigorously prove those asymptotic results heuristically derived in Section 2.5. Secondly, it may lead to better finite sample performance if one chooses different knots for estimating baseline hazards of different causes via splines. For example, when sample size is small, $\tilde{\lambda}_{0k}(t)$ is

likely to be zero for some k 's and l 's, and $-n^{-1}H_p(\tilde{\zeta}, \tilde{\mathbf{h}})^{-1}$ may not be positive definite on the boundary of the parameter space. The problem of non-positive definite covariance matrix will occur less often if the interior knots for the spline approximating $\hat{\lambda}_{0k}(\cdot)$ are equally spaced with respect to the quantiles of the unique values of

$$\{L_i, R_i: \text{a failure from Cause } k \text{ occurs in } (L_i, R_i], i=1, \dots, n\},$$

because the corresponding spline coefficient estimates are less likely to be zero. Last but not least, cross-validation based smoothing parameter selection is computationally intensive as discussed in Section 3 and the Appendix. One may exploit the connection between the penalized likelihood and the mixed-effects model to automatically select the smoothing parameters through maximizing a marginal likelihood as in Cai and Betensky (2003).

Acknowledgments

The data for the example presented in this work were from the Rush Memory and Aging Project supported by the National Institute of Aging (R01AG17917). We are grateful to the principal investigator, David A. Bennett, MD, for the authorization to use the data. This work is supported in part by a grant from National Institute of Dental and Craniofacial Research (1R03DE023889).

References

- Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and findings from the Rush Memory and Aging Project. *Current Alzheimer Research*. 2012; 9:646–663. [PubMed: 22471867]
- Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Statistics in Medicine*. 2009; 28:956–971. [PubMed: 19125387]
- Cai T, Betensky RA. Hazard regression for interval-censored data with penalized spline. *Biometrics*. 2003; 59:570–579. [PubMed: 14601758]
- Commenges D, Joly P, Gégout-Petit A, Liqueur B. Choice between semi-parametric estimators of markov and non-markov multi-state models from coarsened observations. *Scandinavian Journal of Statistics*. 2007; 34:33–52.
- Frydman H, Liu J. Nonparametric estimation of the cumulative intensities in an interval censored competing risks model. *Lifetime Data Analysis*. 2013; 19:79–99. [PubMed: 23054241]
- Gray RJ. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*. 1992; 87:942–951.
- Green, PJ.; Silverman, BW. *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press; 1993.
- Groeneboom P, Maathuis MH, Wellner JA. Current status data with competing risks: consistency and rates of convergence of the MLE. *Annals of Statistics*. 2008a; 36:1031–1063.
- Groeneboom P, Maathuis MH, Wellner JA. Current status data with competing risks: Limiting distribution of the MLE. *Annals of Statistics*. 2008b; 36:1064–1089. [PubMed: 19888358]
- Gu, C. *Statistical Models and Methods for Lifetime Data*. 2. Springer-Verlag; New York: 2013.
- Hudgens MG, Li C, Fine JP. Parametric likelihood inference for interval censored competing risks data. *Biometrics*. 2014; 70:1–9. [PubMed: 24400873]
- Hudgens MG, Satten GA, Longini IM JR. Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics*. 2001; 57:74–80. [PubMed: 11252621]
- Jewell NP, Van Der Laan MJ, Henneman T. Nonparametric estimation from current status data with competing risks. *Biometrika*. 2003; 90:183–197.

- Joly P, Commenges D. A penalized likelihood approach for a progressive three-state model with censored and truncated data: Application to aids. *Biometrics*. 1999; 55:887–890. [PubMed: 11315022]
- 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; 3:433–443. [PubMed: 12933608]
- 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; 54:185–194. [PubMed: 9574965]
- Lawless, JF. *Statistical Models and Methods for Lifetime Data*. 2. John Wiley & Sons; Hoboken, NJ, USA: 2003.
- Li C. The Fine–Gray model under interval censored competing risks data. *Journal of Multivariate Analysis*. 2016; 143:327–344. [PubMed: 26543275]
- Li C, Fine JP. Smoothed nonparametric estimation for current status competing risks data. *Biometrika*. 2013; 100:173–187.
- Nychka D. Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*. 1988; 83(404):1134–1143.
- O’Sullivan F. Fast computation of fully automated log-density and log-hazard estimators. *SIAM J Sci Stat Comput*. 1988; 9:363–379.
- Ramsay JO. Monotone regression splines in action. *Statistical Science*. 1988; 3:425–441.
- Rice J. Convergence rates for partially splined models. *Statistics & Probability Letters*. 1986; 4:203–208.
- Rondeau V, Commenges D, Joly P. Maximum penalized likelihood estimation in a gamma-frailty model. *Lifetime Data Analysis*. 2003; 9:139–153. [PubMed: 12735493]
- Schick A, Yu Q. Consistency of the gmle with mixed case interval-censored data. *Scandinavian Journal of Statistics*. 2000; 27:45–55.
- Stone CJ. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*. 1982; 10:1040–1053.

Appendix: Computational Details

Substituting $\tilde{\lambda}_{0k}(\cdot)$'s for $\lambda_{0k}(\cdot)$'s in the minus penalized log likelihood and reparameterizing h_k 's by $h_k = h/\mu_k$ with μ_k 's satisfying the constraint that $\sum_{k=1}^J \mu_k \text{tr}(\mathbf{Q}_k) = J$, we rewrite the penalty term in the quadratic form of $\tilde{\boldsymbol{\theta}}_k$'s as follows,

$$\sum_{k=1}^J h_k \int_{\tau_1}^{\tau_2} \{\lambda_{0k}^{(1)}(t)\}^2 dt = h \sum_{k=1}^J \mu_k^{-1} \tilde{\boldsymbol{\theta}}_k^T \mathbf{Q}_k \tilde{\boldsymbol{\theta}}_k,$$

where \mathbf{Q}_k is a $q_n \times q_n$ matrix with the jl -th element being $\int_{\tau_1}^{\tau_2} M_j^{(1)}(t) M_l^{(1)}(t) dt$. Note that \mathbf{Q}_k does not depend on k if a common set of M-spline basis functions are used to approximate $\lambda_{0k}(\cdot)$'s of different failure causes, which is the case in this paper.

For fixed h and μ_k 's, we minimize the minus penalized log likelihood with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ subject to the constraint $\boldsymbol{\theta} \geq 0$ using the R built-in function “nlminb”. The initial values for the optimization are 0 for every regression parameter and 0.1 for every spline coefficient. The integrals over (L_j, R_j) in the likelihood do not have analytical expressions and are computed by Gauss-Jacobi quadrature with the number of quadrature points proportional to $R_j - L_j$ and 5 quadrature points for the shortest interval.

To select h and μ_k 's based on the approximate cross-validation criterion, we use the following algorithm that adapts from Algorithm 3.3 in Section 3.4 of Gu (2013).

Algorithm for selecting smoothing parameters:

1. Set $\mu_k^* = 1/\text{tr}(\mathbf{Q}_k)$ and minimize the approximate likelihood cross-validation criterion $V(h, \mu_1^*, \dots, \mu_J^*)$ with respect to h . Denote the resulting minimizer by h^* .
2. Minimize the minus penalized log likelihood with h^* and μ_k^* 's to get estimates $\boldsymbol{\beta}^*$ and $\lambda_{0k}^*(\cdot) = \mathbf{M}(\cdot)^T \boldsymbol{\theta}_k^*$ ($k=1, \dots, J$).
3. Select $\mu_k^0 \propto \int_{\tau_1}^{\tau_2} \{\lambda_{0k}^{*(1)}(t)\}^2 dt = \boldsymbol{\theta}_k^{*T} \mathbf{Q}_k \boldsymbol{\theta}_k^*$ and h^0 that minimizes $V(h, \mu_1^0, \dots, \mu_J^0)$.

Ideally, one needs to adapt Algorithm 3.2 in Section 3.4 of Gu (2013) with h^0 and μ_k^0 's as the starting values to minimize $V(h, \mu_1, \dots, \mu_J)$ with respect to (h, μ_1, \dots, μ_J) . However, as pointed out in Section 3.5.3 of Gu (2013), their Algorithm 3.2 can be slow if the number of μ_k 's is large, and the starting values h^0 and μ_k^0 's often deliver “80% or more” of the achievable performance. We therefore select $h_k^0 = h^0 / \mu_k^0$ ($k=1, \dots, J$) as the smoothing parameters for model fitting. To minimize $V(h, \mu_1, \dots, \mu_J)$ with (μ_1, \dots, μ_J) fixed in Steps 1 and 3, we set the initial value of h to be

$$h^{(0)} \equiv \frac{\sum_{k=1}^J \mu_k^2 \cdot \sum_{t \in T} \sum_{i=1}^{q_n} \left\{ \int_{\tilde{\tau}_1}^t M_i(s) ds \right\}^2}{n \sum_{k=1}^J \mu_k \text{tr}(\mathbf{Q}_k)},$$

which adapts from the “sspreg1” function in the R package “gss”, and then apply the “nlm0” function from the package over short searching intervals on which the curve may be bowl-shaped.

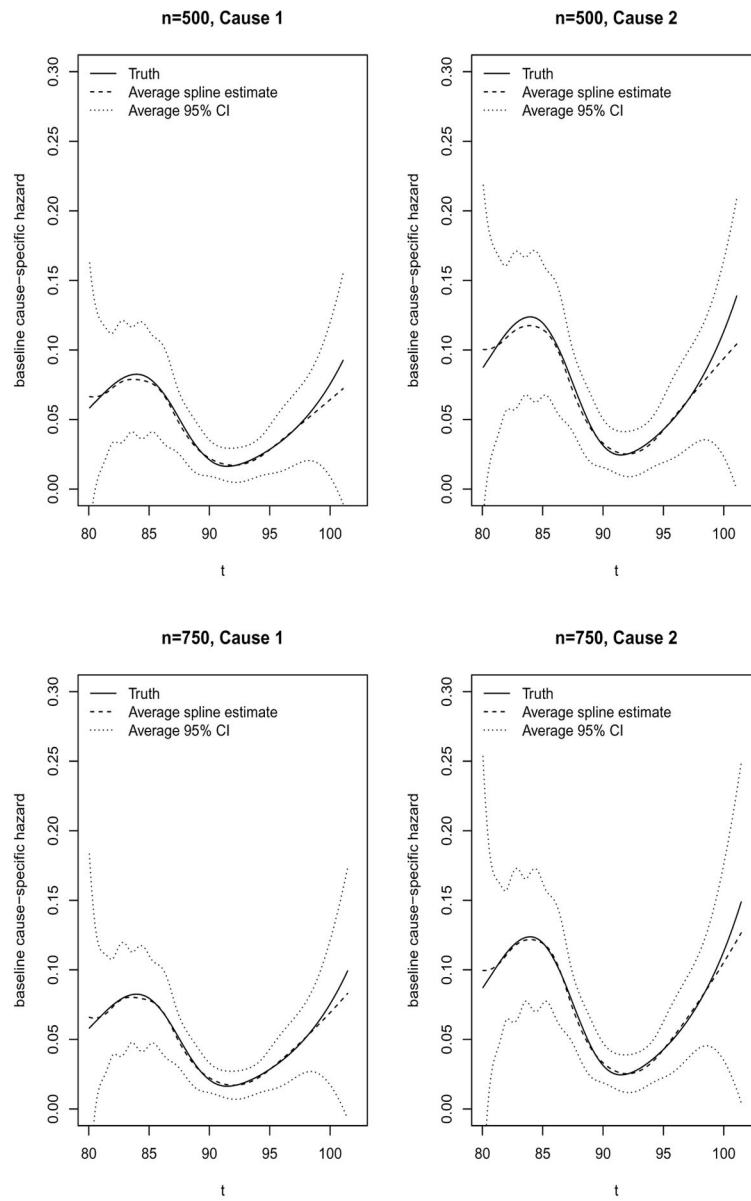


Figure 1. Averaged estimates and 95% pointwise Bayesian confidence intervals of the baseline cause-specific hazard functions across the 1000 Monte Carlo samples.

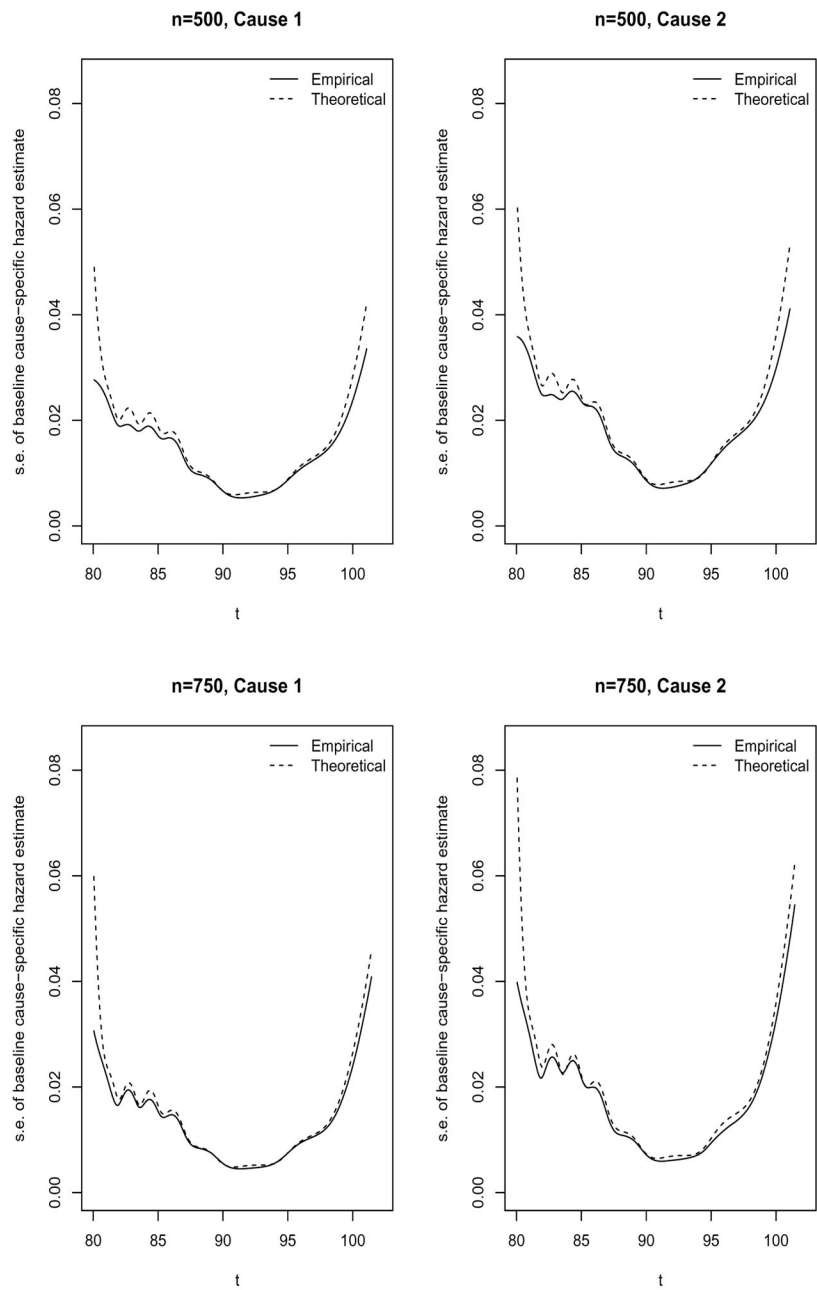


Figure 2. Pointwise empirical standard errors and empirical means of the theoretical standard error estimates of the estimated baseline cause-specific hazard functions.

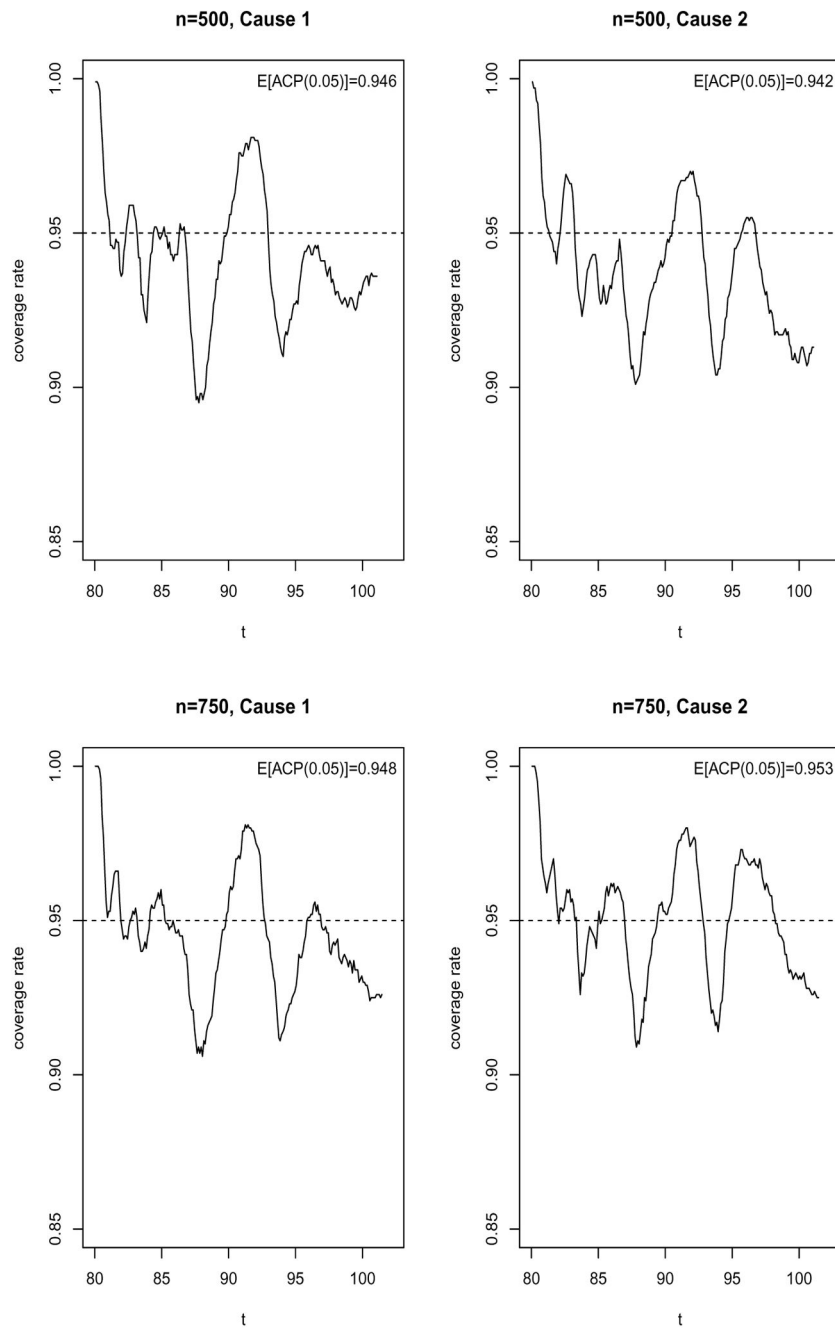


Figure 3. Pointwise empirical coverage and average coverage probabilities of the 95% Bayesian confidence intervals for the baseline cause-specific hazard functions. The solid curves represent the empirical coverage.

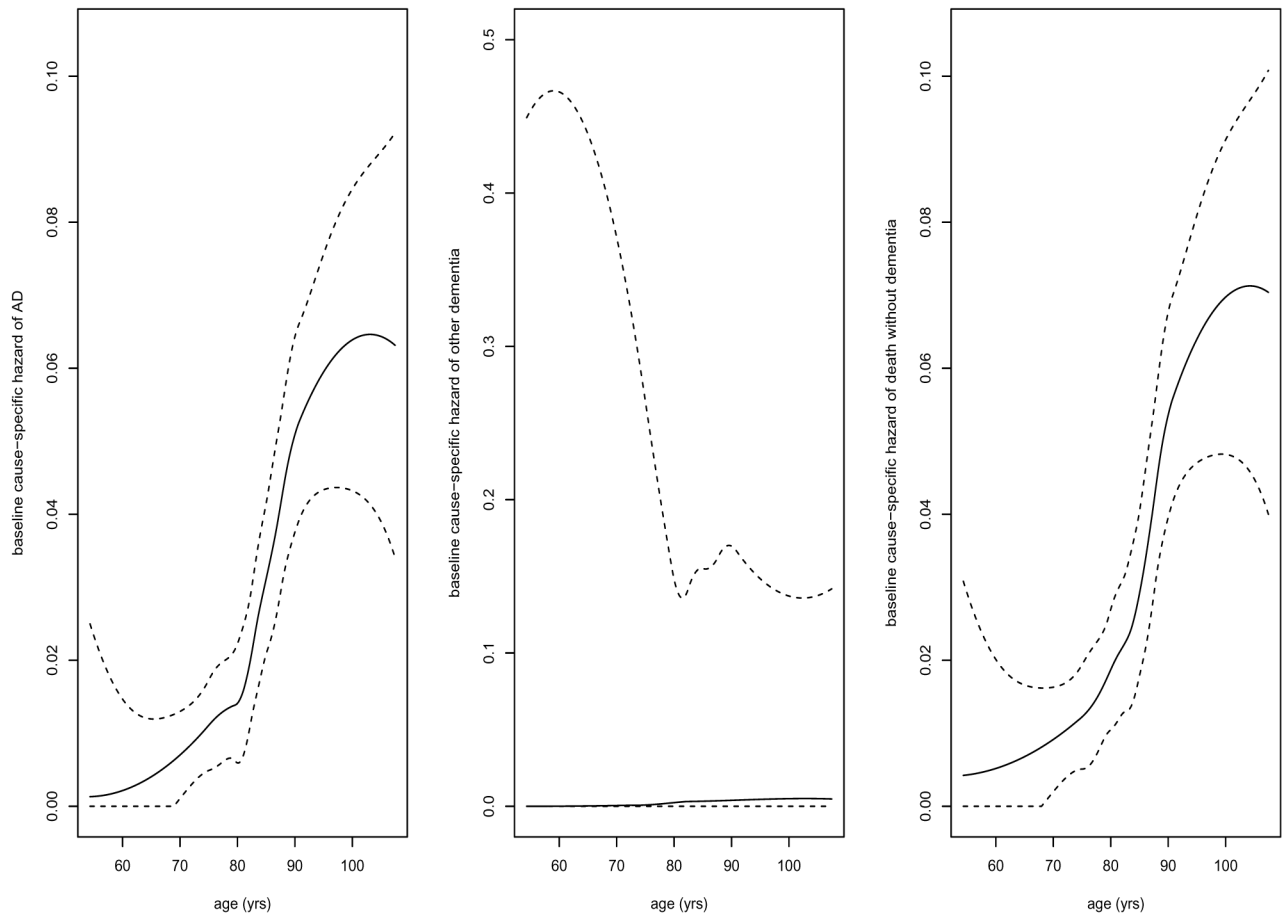


Figure 4.

The baseline cause-specific hazard functions of AD, other dementia and death without dementia. The solid curves represent the point estimates and the dashed curves represent the 95% Bayesian confidence intervals. The plot of other dementia hazard has a different y-axis scale because of the much wider confidence intervals.

Table 1

Simulation results of the estimates of regression parameters

n	Parameter	$E(\hat{\beta})$	$SE(\hat{\beta})$	$E[\widehat{se}(\hat{\beta})]$	$\Pr(\beta \in CI)$
500	$\beta_{11} = 0.30$	0.319	0.158	0.163	0.954
	$\beta_{12} = -0.15$	-0.109	0.274	0.277	0.956
	$\beta_{21} = 0.10$	0.119	0.154	0.148	0.935
	$\beta_{22} = -0.40$	-0.354	0.249	0.254	0.952
750	$\beta_{11} = 0.30$	0.316	0.130	0.133	0.952
	$\beta_{12} = -0.15$	-0.133	0.227	0.226	0.949
	$\beta_{21} = 0.10$	0.107	0.124	0.121	0.943
	$\beta_{22} = -0.40$	-0.384	0.206	0.207	0.954

$E(\hat{\beta})$: the empirical mean of $\hat{\beta}$, $SE(\hat{\beta})$: the empirical standard error of $\hat{\beta}$, $E[\widehat{se}(\hat{\beta})]$: the empirical mean of the theoretical standard error estimates of $\hat{\beta}$, CI : 95% Wald confidence interval.

Table 2

Characteristics of the MAP sample for the analysis

Characteristic	Analysis
Gender—no. (%)	
Male	258 (26.4%)
Female	718 (73.6%)
Years of education—no. (%)	
> 12 (attended college)	667 (68.3%)
12 (not attended college)	309 (31.7%)
ApoE4 status—no. (%)	
carrier	220 (22.5%)
non-carrier	756 (77.5%)
Age at entrance—mean (sd)	80.2 (7.2)
No. of study visits—mean (sd)	5.7 (2.4)
Years of follow-up—mean (sd)	4.8 (2.5)
Observed outcome event—no. (%)	
AD	172 (17.6%)
Other dementia	11 (1.1%)
Death	181 (18.5%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Cause-specific hazard ratios for education, gender and ApoE4

Cause	Covariate	β	$\hat{se}(\hat{\beta})$	$\exp(\hat{\beta})$	95% C.I. for $\exp(\hat{\beta})$	p-value
AD	College education (yes vs no)	0.002	0.167	1.00	(0.722, 1.391)	0.988
	Gender (male vs female)	0.364	0.170	1.44	(1.031, 2.007)	0.032
	ApoE4 (carrier vs non-carrier)	0.728	0.169	2.07	(1.488, 2.884)	<0.001
Other dementia	College education (yes vs no)	-0.803	24.7	0.448	$(4.4 \times 10^{-22}, 4.6 \times 10^{20})$	0.974
	Gender (male vs female)	-1.04	4.69	0.352	$(3.6 \times 10^{-5}, 3.5 \times 10^3)$	0.824
	ApoE4 (carrier vs non-carrier)	1.43	22.8	4.18	$(1.5 \times 10^{-19}, 1.1 \times 10^{20})$	0.950
Death	College education (yes vs no)	0.125	0.167	1.13	(0.817, 1.572)	0.452
	Gender (male vs female)	0.606	0.159	1.83	(1.342, 2.506)	<0.001
	ApoE4 (carrier vs non-carrier)	0.108	0.188	1.11	(0.771, 1.610)	0.565