JOINT MODELING OF TIME-TO-EVENT DATA WITH COMPETING RISKS

by

Bo Fu

B.S. Electrical Engineering, Xidian University, Xi'an, China, 2004

M.S. Electrical Engineering, Xi'an Electrical Engineering Institute, Xi'an, China, 2007

M.S. Statistics, West Virginia University, 2009

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF BIOSTATISTICS

This dissertation was presented

by

Bo Fu

It was defended on

April 16th, 2013

and approved by

Chung-Chou H. Chang, Ph.D., Associate Professor
Departments of Medicine and Biostatistics
School of Medicine and Graduate School of Public Health
University of Pittsburgh

Mary Ganguli, M.D., M.P.H, Professor
Departments of Psychiatry and Epidemiology, and Neurology
School of Medicine
University of Pittsburgh

Francis Pike, Ph.D., Assistant Professor
Departments of Medicine and Biostatistics
University of Pittsburgh

Ruosha Li, Ph.D., Assistant Professor
Department of Biostatistics
University of Pittsburgh

Lisa A. Weissfeld, Ph.D., Professor
Department of Biostatistics
University of Pittsburgh

Dissertation Director: Chung-Chou H. Chang, Ph.D., Associate Professor
Departments of Medicine and Biostatistics
School of Medicine and Graduate School of Public Health
University of Pittsburgh

JOINT MODELING OF TIME-TO-EVENT DATA WITH COMPETING RISKS

Bo Fu, PhD

University of Pittsburgh, 2013

When analyzing time-to-event data, informative dropout due to competing risks is one practical aspect that researchers should take into account. If we fail to account for the association between the event of interest and informative dropouts, unknown amplitude bias may be encountered when identifying the effects of potential risk factors related to time to the main cause of failure. A joint modeling approach of time to the main event and time to the competing events is proposed, to capture the dependence between the main event and the informative dropout due to competing risks via a set of random terms. Two fundamental likelihood functions with different structures of the random terms are provided, which may be combined in practice. We used three methods to optimize the corresponding likelihood functions in order to estimate the unknown covariate effects: Gaussian quadrature method, Bayesian Markov Chain Monte Carlo method, and hierarchical likelihood method. Four bias reduction correction methods for the h-likelihood estimation approach are explored. These methods were aimed to improve the accuracy of parameter estimation. The performances of the three methods were compared via simulations. We applied proposed methods to identify risk factors for dementia.

Time-to-event data have been widely investigated from clinical trials and from observational studies. The proposed joint modeling method is significantly meaningful to public health research because informative dropout commonly exists for the time-to-event data. Methods that have been currently used either fail to adjust for the association between the main event and the informative dropout due to competing events or the methods used to

adjust for the association are not easy to implement. In this dissertation, we showed that the proposed joint modeling approach provides less bias estimates on the effect of a risk factor and has fairly straightforward implementation, which will lead to benefits for medical research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Competing risks data include multiple events due to different causes of failures. In practice, unlike the random censoring, other causes of failure might be due to informative dropouts which are associated with the event of interest. For example, in an observational study of dementia among the elderly, the objective was to identify risk factors of the progression to dementia. The event of interest was progression to dementia, and the time of progression was recorded for each participant. There were non-negligible potential informative dropouts because patients who were too sick to stay in the study or who died before progressing to dementia. Those participants might have high counterfactual hazard of progression to dementia if they stayed in the study.

In analyzing competing risks data with informative dropouts, three commonly used techniques exists in the literature. The first method is the use of cause-specific hazards models, such as the Cox proportional hazards model (Cox, 1972) or the accelerated failure time model (Prentice and Kalbfleisch, 1980) for the event of interest. This model may conveniently be used to determine the hazard ratios. However, one of the fundamental assumptions of these models is independence among different events. The second method involves models for the subdistribution or cumulative incidence function. Models based on this quantity include that the Fine and Gray's proportional subdistribution hazards model (Fine and Gray, 1999), Klein and Andersen's pseudo observation model (Klein and Andersen, 2005), and the inverse-probability-weighted binomial model (Scheike et al., 2008). The subdistribution framework assumes that if a patient experiences a competing event before the main event, his or her time to the main event is infinity. This assumption implies a perfectly negative correlation between the main and the competing events. The third method is the use of copula in relating correlated different types of events. Archimedian copula models have been

suggested in the literature to account for the flexible dependence structures in the absence of prior knowledge about the association between the main and competing events (Escarela and Carriere, 2003). However, the form of copula is mathematically complex, and the estimation might be sensitive to the chosen copula.

To model the dependence between the main and competing events, and to avoid mathematical complexity, we propose to use the joint modeling technique to analyze data with competing risks. The joint modeling technique was raised originally to handle the not missing at random (NMAR) mechanism of longitudinal covariate measured with error when the time to dropouts were recorded (Wulfsohn and Tsiatis, 1997). In this setting, the informative dropouts are treated as NMAR and a functional structure of random effects is used to link the two components–the event of interested and the informative dropouts– and the full dependence between them is assumed to be captured by the random effects. By conditioning on the random effects, the dropout mechanism due to competing risks is missing at random (MAR) theoretically. Applying this idea to time-to-event data with competing risks, the NMAR of time to the main event occurs when individuals fail because of the competing risks. Therefore, we propose to jointly model the cause-specific hazards of time to the main event and that of the competing risks, and use an appropriate random effects mechanism to handle the dependence between different causes of failures. This assumes that the time to the main event is independent with the time to informative dropouts conditional on random effects. That is, the unobserved variation, which is assumed to contain all information of association between the main event and informative dropouts, is treated as random effects with specific distributions.

To apply the proposed model in identifying risk factors of dementia for the study described above, we jointly model the time to progression of Alzheimer's disease and the time to informative dropouts because of death or being too sick. The dependence of these two causes of failures is handled by functional structure of random terms. Two basic functional forms of random terms will be discussed in detail with pros and cons in this study. Joint modeling is advantageous in this scenario because we can account for informative dropout due to death or being too sick and allows flexible dependence structure between time to progression and the dropout process. In addition, the covariate structure for each failure

type is allowed to be different because the marginal models are independent of each other given the random effects. Under the joint modeling framework, the regression coefficients and association parameters are estimated simultaneously.

To optimize the likelihood function and estimate the effects of covariates, one of the computational challenges involves multidimensional integration with respect to the random effects. Likelihood functions of time-to-event models usually do not belong to exponential family and the integration may not be traceable, i.e., there may be no explicit mathematical form to integrate the random terms out. One direct way to optimize it is to numerically integrate out the random effects, using the gaussian quadrature method (Wulfsohn and Tsiatis, 1997). This method works even when the outcomes are non-gaussian type. Software package is available (e.g. NLMIXED procedure in SAS). Using this method, the efficiency may be low and convergence maybe an issue. Another approach is the Bayesian Markov Chain Monte Carlo (MCMC) method (Guo and Carlin, 2004), which can bypass the integration by using the MCMC sampling technique on the posterior function of each parameter. Priors must be specified for all unknown parameters.

Lee and Nelder (1996) proposed the hierarchical likelihood (h-likelihood) approach, which has been used for parameter estimation in frailty model (Ha et al., 2001; Ha and Lee, 2003), and joint models of longitudinal and time-to-event data (Ha et al., 2003), this approach has been reported to provide less bias results than the maximum likelihood (Nielsen et al., 1992) or the penalized likelihood method (Ripatti and Palmgren, 2000). By applying the h-likelihood approach on parameter estimation of our proposed model, the integrations of the random effects are bypassed. The Newton-Raphson method will be used in the optimization procedure, which converges quickly. Extra programming is needed because there is no standard package can be applied directly.

The aims of the dissertation are:

1. to propose a joint model for data with competing risks, in order to reduce the estimation bias caused by the dependence between the main event and the competing events, in the presence of additional independent censoring;

2. to apply the gaussian quadrature method, the Bayesian-MCMC method, and the h-likelihood method on the proposed joint model, and show the performance of them through simulation studies.

In the next chapter, we will give an overview of the three existing approaches to analyze data with competing risks; review the joint modeling method in relation to the three types of investigative problems. After which, the proposed joint model of time to the main event and time to the competing event submodels along with the corresponding likelihood functions for different functional form of random effects will be presented. The three likelihood function optimization methods (the Gaussian quadrature method, the Bayesian-MCMC method and the h-likelihood method) will be discussed. The performance of the 3 estimating methods will be compared via simulations. Finally, we apply the proposed model to identify risk factors of dementia among the elderly.

2.0 METHODS FOR COMPETING RISKS

In this chapter, we discuss the preliminary concepts associated with the models we consider in the dissertation. First, competing risks quantities (cause-specific hazard and subdistribution) are defined and modeling approaches are introduced. Fundamental concepts on copula are then presented for readers who are unfamiliar with the topic. Second, we introduce the notion of joint modeling in the analysis of longitudinal data with informative dropout.

2.1 CAUSE-SPECIFIC HAZARDS FUNCTION

The cause-specific hazards model is used in a competing risks data, when investigators are interested in the hazard rate, which is defined by

$$\lambda_k(t) = \lim_{\mathrm{d}t \to 0} \frac{P(t \leq T < t + \mathrm{d}t, \epsilon = k | T \geq t)}{\mathrm{d}t},$$

where $k = 1, \cdots, K$ denotes the kth event, $T = \min(T_1, \cdots, T_K)$ represents the survival time, where T_k is the time to event k, ϵ tells which event had happened, that is $\epsilon = k$ if $T = T_k$. $\lambda_k(t) \geq 0$, and $\lambda_k(t)\mathrm{d}t$ describes the chance that an individual experiencing the kth event within the time period $[t, t + \mathrm{d}t)$. Five widely used shapes of hazard functions (decreasing hazard, increasing hazard, constant hazard, hump-shaped hazard, and bathtub-shaped shaped hazard) are shown in Figure 1. By assuming mutually independence between different causes of failure, the joint survival function is $S(t_1, \cdots, t_K) = P(T_1 > t_1, \cdots, T_K > t_K) = \prod_{k=1}^{K} S_k(t_k)$, where the $S_k(t_k)$ is the kth marginal survival function, by taking $t_{k'} =$

$0, \forall k' \neq k$, then the cause-specific hazards rate is given by

$$\lambda_k(t) = -\frac{\partial S(t_1, \cdots, t_K)/\partial t_k|_{t_1=\cdots=t_K=t}}{S_T(t)}, \tag{2.1}$$

where $S_T(t) = S(T_1 = t, \cdots, T_K = t)$. Whereas the marginal hazard rate is calculated as

$$\bar{\lambda}_k(t) = -\frac{\partial S_k(t_k)/\partial t_k}{S_k(t_k)} \tag{2.2}$$

Notice that the calculation of the cause-specific hazard rate from the Equation 2.1 would not be the same with the marginal hazard rate without the independence assumption. If there is no independence assumption, $S_k(t)$ is bounded by the (Peterson, 1976) that $S_T(t) \leq S_k(t) \leq 1 - F_k(t)$, where $F_k(t) = P[T \leq t, \epsilon = k]$.

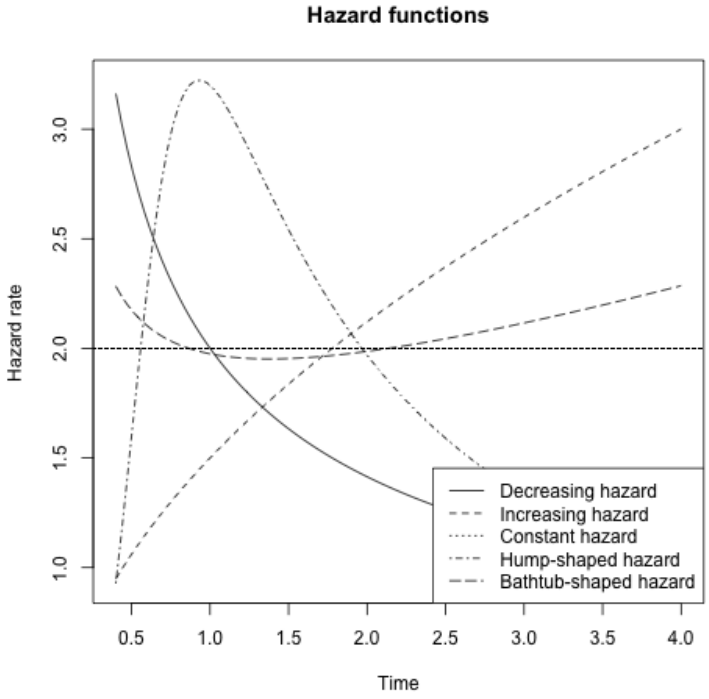

Figure 1: Five different shapes of hazard functions.

To sum up, the cause-specific model assumes independence among different causes of failure, that is, all events other than the interested one are treated as random censoring, bias might be induced due to the existence of informative dropouts.

2.2 SUBDISTRIBUTION HAZARDS MODEL

For competing risks data with covariates, the effect of a particular variate might be very difference on the cause-specific hazard function from that on the corresponding cumulative incidence function (Gray, 1988). So that it may be not possible to test for covariate effects on subdistribution, when modeling the cause-specific hazard function directly. In situations the covariate effects on the probability of different endpoints being observed is of interest, investigators model the cumulative incidence function for failure from event of interest ($k = 1$). The subdistribution hazards function for the event of interest by Gray (Gray, 1988) is defined by

$$\begin{aligned}\tilde{\lambda}_1(t) =& \lim_{\mathrm{d}t\to 0} P\{t \le T \le t+\mathrm{d}t, \epsilon = 1 | T \wedge C \ge t \cup (T \le t \cap \epsilon \ne 1 \cap C \ge t)\}/\mathrm{d}t \\ =& \frac{\partial F_1(t)/\partial t}{1-F_1(t)} \\ =& -\frac{\partial \log\{1-F_1(t)\}}{\partial t},\end{aligned} \tag{2.3}$$

where C is the random censoring time. From the expression of the subdistribution hazards function (2.3), the risk set also includes participants failed due to events other than the event of interest. That is, if an individual drops out because of competing risks, the event of interest would never occur, which means that, the relationship between the event of interest and the competing risks are perfect negative.

A frequently cited subdistribution regression model is the Fine and Gray proportional subdistribution hazards model (Fine and Gray, 1999). An increasing function $g(\cdot)$ is assumed, and $g\{F_1(t;Z)\} = h_0(t) + Z^T\beta$, where $h_0(\cdot)$ is a function that complete unspecified, invertible, and monotone increasing, Z is bounded time-independent covariate vector, and β is the parameter of Z. A suggested function of $g = \log\{-\log(1-u)\}$, which links the subdistribution cumulative incidence function to proportional hazards function. A weighted partial likelihood score equation is used for parameters' estimation

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_i - \frac{\sum_j \omega_j(t) \exp(\beta^T Z_j) Z_j}{\sum_j \omega_j(t) \exp(\beta^T Z_j)} \right\} \omega_i(t) \mathrm{d}N_i(t)$$

where $\omega_i(t) = I(C_i \ge T_i \wedge t) \frac{\hat{G}(t)}{\hat{G}(T_i \wedge t)} Y_i(t)$, $\hat{G}(t)$ is suggested to be the Kaplan-Meier estimator

of the random censoring survival distribution, $Y_i(t) = I(T_i < t, \epsilon_i \neq 1) + I(T_i \geq t)$, $\tau = \sup\{t : \Pr(\omega(t) \geq \epsilon > 0) > 0\}$, and $N_i(t) = I(T_i \leq t, \epsilon_i = 1)$ is the counting process. From the formula of getting the at risk indicator $Y_i(t)$, individuals failed not because of the event of interest are counted in.

Recently, more methods were developed based on subdistribution mechanism, such as modeling with pseudo-observations (Klein and Andersen, 2005), using of a parametric regression on cumulative incidence function (Jeong and Fine, 2007),and inverse-probability-weighted binomial model (Scheike et al., 2008).

2.3 COPULA METHOD

If the direction and amplitude of the dependence between the event of interest and the competing risks is unknown, a copula method can be considered.

A copula function is defined as a binary function $C : [0,1]^2 \to [0,1]$, which satisfies the following properties:

1. $\forall u, v \in [0,1], C(u,0) = 0 = C(0,v)$, $C(u,1) = u$, and $C(1,v) = v$,
2. $\forall 0 \leq u_1 \leq u_2 \leq 1$ and $\forall 0 \leq v_1 \leq v_2 \leq 1$, $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) = C([u_1,v_1] \times [u_2,v_2]) \geq 0$.

The Sklar's theorem (Sklar, 1959) proved that for a joint distribution $H(x,y)$ with two marginal distributions $F(x)$ and $G(y)$, there exists a copula C that $\forall x, y \in \bar{R}$, $H(x,y) = C(F(x), G(y))$. If $F(x)$ and $G(y)$ are continuous, then C is unique. The Sklar's theorem illustrates that copula plays a role of elucidating the relationship between a joint distribution and its margins. This theory ensures that the uniqueness of copula C, to describe the dependency of joint survival function and its margins, after a copula family is chosen.

Suppose there are one event of interest (T_1) and one competing risk (T_2), the bivariate survival function can be written in copula format (Escarela and Carriere, 2003)

$$S(t_1,t_2) = C(S_1(t_1), S_2(t_2)) = \psi_\theta[\psi_\theta^{-1}\{S_1(t_1)\} + \psi_\theta^{-1}\{S_2(t_2)\}],$$

where $S(t_1, t_2)$ is the joint survival distribution of t_1 and t_2, $S(t_1)$ and $S(t_2)$ are marginal survival functions of T_1 and T_2 respectively, $\psi_\theta : [0, \infty] \to [0, 1]$ is a generator function of copula in the Archimidean family. ψ_θ satisfies $\psi_\theta(0) = 1$, and twice differentiable with $\psi'_\theta, \psi''_\theta < 0$, and θ is she parameter of copula. By assuming the random censoring times are independent with all other causes of failure, the likelihood is written as

$$L = \prod_{i=1}^{n} \left(\prod_{k=1}^{K} [\lambda_k(X_i)]^{\delta_{ik}} \right) [S_T(X_i)]^{1-\delta_{i\cdot}},$$

where, $i = 1, \ldots, n$ are the n participants, $X_i = T_i \wedge C_i$ denotes the observed time, $\delta_{ik} = I(X_i = T_{ik})$ denotes the indicator of the ith person experienced the kth event, $\delta_{i\cdot} = I(X_i = T_i)$ denotes the indicator of the ith person experienced an event. Once a type of copula is assumed, the mathematical form of the likelihood function above could be derived. The association can be captured by the parameter θ, which is related to Kendall's τ or Spearman's ρ. The copula method is able to estimate coefficients effects under different association structure between the different causes of failures. However, the chosen copula type the Archimedean family is arbitrary, and the estimates of likelihood function might be sensitive to the specific copula selected.

2.4 JOINT MODELING APPROACH

Suppose n individuals participated in a study in which a longitudinal variable $Z_{ij}, i = 1, \cdots, n,\ j = 1, \cdots, n_i$ were measured n_i times for the ith person. Same as the previous section, assume T_i to be the time to dropouts and C_i be the time to random censoring, and $X_i = T_i \wedge C_i$ be the observed dropout time. Joint modeling method is mainly used to deal with the following three scenarios

(1) interest is on the distribution function of survival time $f(X_i|Z_{ij})$ (Wulfsohn and Tsiatis, 1997), when the longitudinal variable Z_{ij} measured with error,

(2) interest is on the distribution function of longitudinal measurements $f(Z_{ij}|X_i)$ (Little, 1993), when there are the presence of informative dropouts, or

(3) interest is on both distribution functions of $f(Z_{ij})$ and $f(X_i)$ (Henderson and Dobson, 2000), simultaneously.

The main challenge of the first scenario is that the Z_{ij} measured with error, so it would be biased when exploring the effect of the Z_{ij} on the survival time X_i. Joint modeling $f(X_i|Z_{i,j})f(Z_{ij}) = f(X_i, Z_{ij})$ is used for a purpose of removing the bias. By maximizing the likelihood function of the joint model, more accurate effect of Z_{ij} can be calculated, because the longitudinal covariate $f(Z_{ij})$ provided the true estimates of Z_{ij} at each survival time. Random effects v_i in the likelihood function of modeling Z_{ij} are used to capture the unobserved variation of Z_{ij}, which are assumed to be normally distributed here.

Considering the scenario (ii), the informative dropouts lead the data to NMAR, which violates the assumption of missing of mixed effects model. Jointly modeling $Z_{ij}|X_i$ with X_i as $f(Z_{ij}|X_i)f(X_i) = f(X_i, Z_{ij}|)$ adds the information of informative dropouts into the likelihood, and converts the NMAR to missing at random, by modeling the missing indicator and time to dropouts simultaneously with the longitudinal model.

For the last scenario, statistical inference on the marginal properties $f(Z_{ij})$ and $f(X_i)$ can be obtained from modeling likelihood function $\int_{v_i} f(Z_{ij}|v_i)f(X_i|v_i)f(v_i)\mathrm{d}v_i = f(Z_{ij}, X_i)$. The association of those two marginals are assumed to be captured by the random effects v, that is, Z_{ij}'s are independent with X_i, conditional on v_i.

The question addressed by our proposed method is similar to scenario (1). But, the time to the event of interest is focused on instead of the longitudinal measurements Z. The model we proposed models the time to the main event, adjusting for informative dropout due to competing risks.

3.0 JOINT MODELING OF TIME-TO-EVENT DATA WITH COMPETING RISKS

3.1 INTRODUCTION

For time-to-event analyses in clinical trials or observational studies, several methods have been proposed to estimate the effect of a covariate on the main event of interest in the presence of competing events. Among the most popular methods used are the cause-specific hazards model, the subdistribution hazards model, and the copula model. When analysts are interested in determining the hazards or hazard ratios, they generally use a cause-specific hazards model such as the Cox proportional hazards model (Cox, 1972) or an accelerated failure time model (Prentice and Kalbfleisch, 1980). Because competing events are treated as censoring, one of the key assumptions of this type of model is that the main event is independent of all other causes of failures. In situations in which covariate effects on the failure probabilities are of interest, analysts often use a subdistribution (or cumulative incidence function) hazards model such as the Fine and Gray model (Fine and Gray, 1999). This type of model assumes that if a patient experiences a competing event before the main event, his or her time to the main event is infinity, i.e., the patient will never experience the event of interest. Therefore, this model assumes that the main event and the competing events are perfectly negatively correlated. If dependence between the main event and the competing events cannot be presumed, analysts usually use a model with an assumed Archimedean copula (Escarela and Carriere, 2003). Although this model can allow for flexibility in the dependence of the main event and the competing events, the exact form of the copula is mathematically complex and the results might be sensitive to the chosen copula.

In this chapter, we propose a method to estimate covariate effects on the main event by treating the occurrence of the competing events as informative dropouts. This method will result in a flexible but mathematically simpler form of dependence between the main and competing events. The method entails jointly modeling the time to the main event and the time to the competing events and linking the two submodels via unobserved random terms to capture the association between different causes of failure. The joint model assumes that conditional on the random effects, all event times are independent. In other words, the model assumes that the unobserved random terms contain all the information about associations between the main event and the competing events.

When we estimate unknown parameters of the joint model, our main challenge is to optimize the likelihood function, which contains a multidimensional integration with respect to all the random terms. The integration is often untraceable, which means that there may be no explicit mathematical form to integrate the random terms out. To approximate the integration and resolve the issue, we propose to use three different methods: the Gaussian quadrature method, the Bayesian-Markov chain Monte Carlo (Bayesian-MCMC) method, and the hierarchical likelihood (h-likelihood) method. We then use simulations to compare the performance of these methods.

The proposed approach was motivated by the desire to identify risk factors of Alzheimer's disease and other forms of dementia among elderly individuals. The study population was recruited from a U.S. community and was followed for about 20 years. Because the individuals recruited were at least 65 years old, a nontrivial number of them died during the study follow-up or felt too ill to continue participating in the study. We believe that those who dropped out because of death or severe illness would have been at potentially higher risk of developing dementia if they had continued in the study. For this reason, we believe that using a cause-specific model or a subdistribution model may have led to a biased estimate of covariate effects. We therefore analyzed the data using our proposed approach and compared its results with those of a cause-specific model and a subdistribution model.

In subsequent sections of this article, we introduce notations and our proposed approach with two correlation structures of likelihood functions (Section 3.2); we introduce the Gaussian quadrature, Bayesian-MCMC, and h-likelihood methods for estimation (Section 3.3); we

show the performance of each of the estimation methods via simulation studies (Section 5); we present the results of our analyses of covariate effects on the risk of developing dementia (Section 6); and we present our conclusions (Section 7).

3.2 NOTATIONS AND MODEL

Suppose that there are n independent patients enrolled in a study. Duration time to each of the K types of events and censoring are recorded. Without lost of generality, let $K = 2$, that $k = 1$ indicates the event of interest, and $k = 2$ indicates the competing event.. For subject i $(i = 1, \cdots, n)$, let $T_{ki} = (T_{\epsilon i}|\epsilon = k), k = 1, 2$ be the failure time with respect to event k, and C be the censoring time. The observed time for individual i is $X_i = T_i \wedge C_i$, where $T_i = T_{1i} \wedge T_{2i}$. Therefore, for individual i, we observe $\mathbf{O}_i = \{X_i, \delta_{ki}, Z_{ki}\}$, where $\delta_{ki} = I(X_i = T_{ki})$ are the event indicator; Z_{1i} and Z_{2i} are p-dim and q-dim vectors of covariates related to the main event and competing event, respectively.

Let $\lambda_k, k = 1, 2$ denote the cause-specific proportional hazards for event k with the form

$$\begin{aligned} \lambda_k(t) &= \lim_{\mathrm{d}t \to 0} \{\Pr(t \leq T < t + \mathrm{d}t, \epsilon = k | t \leq T)/\mathrm{d}t\} \\ &= \lambda_{k0} \exp(\beta_k^T Z_k), \end{aligned} \tag{3.1}$$

where λ_{k0} represent the baseline hazards for event k; and β_k is a vector of unknown regression coefficients. We propose to jointly model these K cause-specific hazards, using random terms to account for the dependence among the k events. The following assumptions are made throughout this paper:

Assumption 1. Conditional on the random terms v_i, time to the event of interest is independent of the time to the competing event, i.e., $T_{1i} \perp T_{2i}|v_i$, where v_i is the random effects for the ith subject.

Assumption 2. Censoring time is independent of the main and competing events, i.e., $T_{ki} \perp C_i$.

Assumption 3. The probability density function of the censoring time is not identical to that of the main event or competing event.

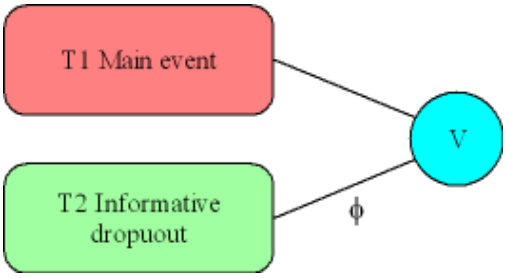

Figure 2: Structure of shared random effects model

Therefore, the likelihood function can be written as

$$L_i(\mathbf{O_i}; \mathbf{\Omega}, Z_i, v_i) = \prod_{i=1}^{n} \prod_{k=1}^{2} l_k(T_{ki}; \mathbf{\Omega}, Z_{ki}|v_{ki}) f(v_i), \tag{3.2}$$

where $\mathbf{\Omega} = \{\beta_k, \lambda_{k0}, \theta\}$; θ is a vector of parameters in the density function of the random terms; $f(v_i)$ is the probability density function of the random terms v_i, and l_k denotes the marginal likelihood function of event k with the following form

$$l_k(t|v_i) = \{\lambda_{ki}(t|v_i)\}^{\delta_{ki}} \exp\left\{-\int_0^t (\lambda_{ki}(t|v_i)\mathrm{d}t\right\}. \tag{3.3}$$

The marginal likelihood with respect to $\mathbf{O_i}$ conditional on $\mathbf{\Omega}$ and Z_i, can be obtained by integrating the unobserved random terms out from (3.2). That is

$$L(\mathbf{O_i}; \mathbf{\Omega}, Z_i) = \prod_{i=1}^{n} \int_{v_i} \prod_{k=1}^{2} l_k(T_{ki}|v_{ki}) f(v_i)\mathrm{d}v_i. \tag{3.4}$$

We propose two structures to link between the two time-to-event submodels: shared random effects structure and correlated random effects structure. They are described in detail in Sections (3.2.1), and (3.2.2).

3.2.1 Shared random effects structure

The model with shared random effects structure is depicted in Figure 3.2.1. The likelihood function can be written as the form

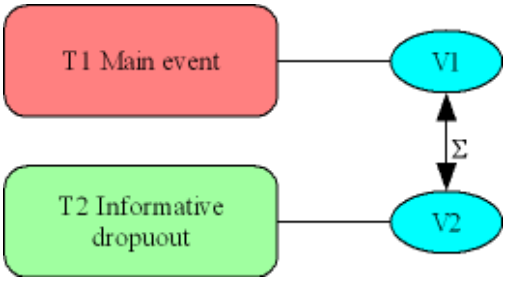

Figure 3: Structure of correlated random effects model

$$L = \prod_{i=1}^{n} \int_{v_i} \{\lambda_{1i}(t|v_i)\}^{\delta_{1i}} e^{\left\{-\int_0^{X_i} \lambda_{1i}(s|v_i)\mathrm{d}s\right\}} \{\lambda_{2i}(t|v_i)\}^{\delta_{2i}} e^{\left\{-\int_0^{X_i} \lambda_{2i}(s|v_i)\mathrm{d}s\right\}} f(v_i)\mathrm{d}v_i, \tag{3.5}$$

which contains three parts: the likelihood function of the main event, the likelihood function of the competing event, and the density function of the random terms. We assume that λ_{ki} follow the proportional hazards setting, and $v_i \sim N(0, \sigma^2)$, i.e.,

$$\lambda_{1i}(t) = \lambda_{10}(t) \exp(\beta_1^T Z_{1i} + v_i)$$
$$\lambda_{2i}(t) = \lambda_{20}(t) \exp(\beta_2^T Z_{2i} + \phi v_i).$$

Based on this setting, parameters need to be estimated are $\mathbf{\Omega} = \{\beta_k, \lambda_{0k}, \phi, \sigma\}, k = 1, 2$.

3.2.2 Correlated random effects structure

The model with correlated random effects structure is depicted in Figure 3.2.1. The likelihood function can be written as the form

$$\begin{aligned} L = \prod_{i=1}^{n} \int_{v_{1i}, v_{2i}} & \{\lambda_{1i}(t|v_{1i})\}^{\delta_{1i}} e^{\left\{-\int_0^{X_i} \lambda_{1i}(s|v_{1i})\mathrm{d}s\right\}} \\ & \{\lambda_{2i}(t|v_{2i})\}^{\delta_{2i}} e^{\left\{-\int_0^{X_i} \lambda_{2i}(s|v_{2i})\mathrm{d}s\right\}} f(v_{1i}, v_{2i})\mathrm{d}v_{1i}\mathrm{d}v_{2i}. \end{aligned} \tag{3.6}$$

Assume v_{1i}, v_{2i} follow a bivariate distribution with density function f. If f is a bivariate normal density, $(v_1, v_2)^T \sim N(0, \Sigma)$ and $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where $\rho \in [-1, 1]$ is

the correlation between v_{1i} and v_{2i}. The three parts of the likelihood function under the proportional hazards assumption become

$$\lambda_{ki}(t) = \lambda_{k0}(t) \exp\left(\beta_k^T Z_{1i} + v_{ki}\right), \quad k = 1, 2;$$
$$f(v_i) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left(-\frac{1}{2} v_i^T \Sigma^{-1} v_i\right).$$

Based on this setting, parameters need to be estimated are $\mathbf{\Omega} = \{\beta_k, \lambda_{0k}, \sigma_k, \rho\}$.

3.3 ESTIMATION METHODS

From the likelihood functions in (3.5) and (3.6), the main challenge of parameter estimation is the multidimensional integration towards the random terms. To approximate the integration, we propose to use three different optimization methods to solve the estimating equations. The three methods are the Gaussian quadrature method, the Bayesian-MCMC method, and the h-likelihood method.

The Gaussian quadrature method approximates an integral by a weighted sum over the predefined abscissas for the random effects (Liu and Huang, 2008). Based on the Bayes rule, the Bayesian-MCMC method estimates parameters by estimating the underlying posterior distribution, and generates sequence of random samples from an arbitrary probability density function (Iversen, 1984). The h-likelihood approach bypasses the integration of random terms by treating the random terms as unknown parameters (Lee and Nelder, 1996). A bias correction procedure was proposed to improve the estimation performance when a high censoring rate was observed (Jeon and Hsu, 2011).

3.3.1 Gaussian quadrature method

If we rewrite the likelihood function (3.4) as the form

$$L = \prod_{i=1}^{n} L_i = \prod_{i=1}^{n} \int_{v_i} \exp \sum_{k=1}^{2} \tilde{l}_k(T_{ki}|v_{ki}) f(v_i) \mathrm{d}v_i, \tag{3.7}$$

where $\tilde{l}_k(T_{ki}|v_{ki}) = \log(l_k(T_{ki}|v_{ki}))$; and if function $f(x)$ can be written into the form $f(x) = W(x)g(x)$, where $g(x)$ is approximated by a polynomial of order $(2Q-1)$, then the Gaussian quadrature method can be used to approximate the integrand $W(x)g(x)$

$$\int_a^b f(x)\mathrm{d}x = \int_a^b W(x)g(x)\mathrm{d}x \approx \sum_{i=0}^{Q} w_i g(x_i),$$

where x_i's are the nodes or augments, sampling from the domain of x; and w_i's are the weights. For a chosen kernel $W(x)$ and a fixed domain of x, x_i and w_i can be determined uniquely (Pinheiro and Bates, 1995). If the random terms $f(v_i)$ follows a normal distribution, likelihood function (3.7) can be approximated by

$$L_i \approx \sum_{q=1}^{Q} \exp\left(\sum_{k=1}^{2} \log(l_k(T_{ki}|u_q))\right) f(u_q)w_q,$$

where $q = 1, \cdots, Q$ are the Q predetermined quadrature points with respect to the random terms v_i. Liu and Huang (2008) suggested to set $u_q = \sqrt{2}z_q, w_q = \sqrt{2}\eta_q \exp(z_q^2)$ in a frailty model with normal distributed frailty, where z_q and η_q can be obtained from tables in Abramowitz and Stegun (1972).

3.3.2 Bayesian-MCMC method

Bayesian method makes inference of the parameters of interest through the expectation of posterior probability function. Based on the Bayes' rule, we need first obtain the likelihood of the observed data and assume prior distributions for all unknown parameters.

Based on the conditional independence assumption, the likelihood of observed information for subject i given unknown parameters can be written as

$$L_i(X_i, \delta_{ki}; Z_{ki}|\beta_1, \beta_2, \theta, \lambda_{10}, \lambda_{20}, v_i) = (\lambda_{ki}(t|v_i))^{\delta_{ki}} \exp\left(-\int_0^t (\lambda_{ki}(t|v_i))\mathrm{d}v_i\right).$$

Denote $\mathbf{\Omega}' = \{\beta_1^T, \beta_2^T, \theta^T, \lambda_{10}, \lambda_{20}, v_i\}$, θ is parameters of random term density; and denote $\mathbf{\Omega}'_s, s = 1, \cdots, S$ as the s^{th} component of $\mathbf{\Omega}'$, where S is the length of vector $\mathbf{\Omega}'$. Then $L_i(X_i, \delta_{ki}; Z_{ki}|\beta_k, \theta, \lambda_{k0}, v_i) = L_i(\mathbf{O}_i|\mathbf{\Omega}')$, where $\mathbf{O}_i = \{X_i, \delta_{ki}, Z_{ki}\}$ is defined in Section 3.2.

The next step of Bayesian procedure is to set up priors, $\pi(\mathbf{\Omega}')$, for each unknown pa-

rameters in $\mathbf{\Omega}'$. Prior reflects the researcher's acquaintance about parameters regardless the information from data. For the coefficients of covariates β_k in the proposed model, non-informative normal distributed priors with small variance (e.g. 0.01) are assigned, so that the priors will have minimal impact on the posterior distribution. For all other parameters, conjugate priors are used.

The posterior density $f_{\mathbf{\Omega}'|\mathbf{O}_i}$, which is the probability function of all parameters $\mathbf{\Omega}'$ conditional on the observed data $\mathbf{O}$, has the form

$$f_{\mathbf{\Omega}'_s|\mathbf{O}_i}(\mathbf{\Omega}'_s|\mathbf{O}_i) = \frac{L_i(\mathbf{O}_i|\mathbf{\Omega}'_s)\pi(\mathbf{\Omega}'_s)}{\int_{\mathbf{\Omega}'_s} L_i(\mathbf{O}_i|\mathbf{\Omega}'_s)\pi(\mathbf{\Omega}'_s)\mathrm{d}\mathbf{\Omega}'_s}. \tag{3.8}$$

Note that it is difficult to derive an explicit mathematical form for $f_{\mathbf{\Omega}'_s|\mathbf{O}_i}(\mathbf{\Omega}'_s|\mathbf{O}_i)$ through Equation (3.8). MCMC sampling technique is used to generate a Markov chain $\hat{\mathbf{\Omega}}_s'^{(b)}, b = 1, \cdots, B$ from Equation (3.8). An example on how to accomplish this sampling procedure using the Metropolis-Hastings (MH) algorithm is given in Appendix A. Other MCMC sampling techniques can also be used.

Parameter estimates will be obtained by calculating of the posterior mean as follows

$$\hat{E}(\mathbf{\Omega}'_s|\mathbf{O}_i) = \frac{1}{B - B'} \sum_{i=B'+1}^{B} \hat{\mathbf{\Omega}}_s'^{(i)},$$

where B' is the number of burn-in processes. The burn-in process ignores a certain number of iterations in the initial portion in order to minimize the effect of starting values on the estimation of posterior distributions.

Deviance information criterion (DIC) can be used for variable selection. Based on Spiegelhalter et al. (2002), define the deviance as $D(\mathbf{\Omega}') = -2\log(l(\mathbf{O}|\bar{\mathbf{\Omega}}')) + 2\log g(\mathbf{O})$, where $g(\mathbf{O})$ is some fully specified function of observed data only, which will be cancelled out for the model comparison purpose. Define the effective dimension as $p_D = \bar{D}(\mathbf{\Omega}') - D(\tilde{\mathbf{\Omega}}')$, where $\bar{D}(\mathbf{\Omega}') = -2E'_{\mathbf{\Omega}}\{\log f(\mathbf{O}|\mathbf{\Omega}')|\mathbf{O}\} + 2\log g(\mathbf{O})$ is the posterior mean deviance, $\tilde{\mathbf{\Omega}}'$ is usually chosen as the posterior mean $\bar{\mathbf{\Omega}}' = E(\mathbf{\Omega}'|\mathbf{O})$. Therefore, DIC is defined as

$$
\begin{aligned}
DIC =& \bar{D}(\boldsymbol{\Omega}') + p_D = 2\bar{D}(\boldsymbol{\Omega}') - D(\tilde{\boldsymbol{\Omega}}) \\
=& -4E'_{\boldsymbol{\Omega}}\{\log f(\mathbf{O}|\boldsymbol{\Omega}')|\mathbf{O}\} + 2\log f(\mathbf{O}|\tilde{\boldsymbol{\Omega}}').
\end{aligned}
$$

3.3.3 H-likelihood method

The h-likelihood approach bypasses the multidimension integration in the optimization procedure by treating the random effect terms as unknown parameters. The Newton-Raphson procedure is then to be used to optimize the likelihood function.

From Section 3.2, the full likelihood of the proposed joint model with multivariate normal shared structure has the form

$$
\begin{aligned}
L_{full} =& \prod_{i=1}^{n}\prod_{k=1}^{2} f_k(T_k|v_i)f(v_i) \\
=& \prod_{i=1}^{n}\left[\prod_{k=1}^{2}\{\lambda_{0k}\exp(Z_{ki}^T\beta_k + v_{ki})\}^{\delta_{ki}}\exp\{-\Lambda_{0k}(X_i)\exp(Z_{ki}^T\beta_k + \phi^{I(k=2)}v_i)\}\right] \\
& \frac{1}{\sqrt{2\pi|\Sigma|}}\exp\left(-v_i^T\Sigma^{-1}v_i\right)
\end{aligned}
$$

where $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(s)d(s)$ is the cumulative baseline hazards for event $k = 1, 2$. $I(k = 2)$ is an indicator function of the competing event, that is, $\phi^{I(k=2)} = \phi$ when $k = 2$. The h-likelihood function is defined as

$$
h = \log L = \sum_{k=1}^{2}\sum_{i=1}^{n} l_{ki} + \sum_{i=1}^{n} l_{3i},
$$

where

$$
\begin{aligned}
l_{ki} =& \log\left[\{\lambda_{0k}(X_i)\exp(Z_{ki}^T\beta_k + \phi^{I(k=2)}v_i)\}^{\delta_{ki}}e^{-\Lambda_{0k}(X_i)\exp(Z_{ki}^T\beta_k + \phi^{I(k=2)}v_i)}\right] \\
=& \delta_{ki}(\log\lambda_{0k}(X_i) + Z_{ki}^T\beta_k + \phi^{I(k=2)}v_i) - \Lambda_{0k}(X_i)\exp(Z_{ki}^T\beta_k + \phi^{I(k=2)}v_i); \quad k = 1, 2 \\
l_{3i} =& -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}v_i^T\Sigma^{-1}v_i.
\end{aligned}
$$

Using a Weibull baseline hazards $\lambda_{0k}(t) = \tau_k\gamma t^{\tau_k - 1}$ with a shape parameter τ_k and a scale parameter γ. For simplicity, we set this parameter value equal to 1 subsequently. Therefore,

the likelihood components can be rewritten as

$$\begin{aligned} l_{ki} =& \delta_{ki}(\log \lambda_{0k}(X_i) + Z_{ki}^T\beta_k + v_{ki}) - \Lambda_{0k}(X_i)\exp(Z_{ki}^T\beta_k + \phi^{I(k=2)}v_i) \\ =& \delta_{ki}\log\mu_{ki} - \mu_{ki} + \delta_{ki}(\log\tau_k - \log t_i), \end{aligned}$$

where μ_{ki} represent the cumulative hazards function with the form

$$\mu_{ki} = \int_0^{X_i} \lambda_{0k}(t)\exp(Z_{ki}^T\beta_k + v_{ki})\mathrm{d}t = \exp\left(\tau_k\log(X_i) + Z_{ki}^T\beta_k + \phi^{I(k=2)}v_i\right).$$

The parameters need to be estimated in the likelihood are $\Omega = (\alpha, v_{ki}, \theta)$, where $\alpha = \{\beta_1, \beta_2, \tau_1, \tau_2, \phi\}, \theta = \{\sigma^2\}$.

The h-likelihood estimation procedure is summarized as follows. Appendix B gives details of the derivatives that will be used in the estimating equations with respect to the h-likelihood function.

1. To estimate the parameters of interest, $\beta_k, k = 1, 2$, and the nuisance parameters τ_k, ϕ, v_i, and θ, $k = 1, 2$; all parameters are split into three parts: $\{v_{ki}\}$, $\{\alpha\}$, and $\{\theta\}$. We will then apply the Newton-Raphson procedures for the three estimating equations with respect to these three sets of parameters that suggested in Ha et al. (2010). Initial values, $\alpha^{(0)} = \{\hat\beta_1^{(0)}, \hat\beta_2^{(0)}, \hat\tau_1^{(0)}, \hat\tau_2^{(0)}, \hat\phi^{(0)}\}$ can be obtained from the estimates of the cause-specific models for the main event and competing event, separately. Initial values of θ can be set to $\hat\theta^{(0)} = \{\hat\sigma^{(0)} = 1\}$. Random terms, $\hat v_i^{(0)}$s will be generated from $N(0, \hat\sigma^{2(0)})$.

2. For the m^{th} iteration, the maximum h-likelihood estimators (MHLE) of $\hat v_i^{(m)}$ can be obtained from the following equation

$$\left(\hat v^{(m)}\right)_{(n\times 1)} = \left(\hat v^{(m-1)}\right)_{(n\times 1)} + \left[H^{-1}_{(n\times n)}\left(\frac{\partial h}{\partial v}\right)_{(n\times 1)}\right]\Bigg|_{(v)=(\hat{}v)^{(m-1)})}, \tag{3.9}$$

   where $\hat v^{(m-1)}$ denote the estimators of v at the $(m-1)^{th}$ iteration; $H = -\frac{\partial^2 h}{\partial v^T \partial v}$ is the hessian matrix, which is the negative second derivatives of h with respect to the random terms v.

3. The following adjusted profile h-likelihood (APHL) is used to estimate the MHLE of $\hat\alpha^{(m)} = \{\hat\beta_1^{(m)}, \hat\beta_2^{(m)}, \hat\tau_1^{(m)}, \hat\tau_2^{(m)}, \hat\phi^{(m)}\}$, because the maximal profile likelihood estimator

might have potential bias Lee and Nelder (1996).

$$h_v = h|_{(v=\hat{v}^{(m)})} - \log \left| \frac{H}{2\pi} \right| \Big|_{(v=\hat{v}^{(m)})}. \tag{3.10}$$

4. The following APHL is used to estimate the MHLE of $\hat{\theta}^{(m)} = \{\hat{\sigma}^{2(m)}\}$,

$$h_{\alpha,v} = h|_{(\alpha_k=\hat{\alpha}_k^{(m)};v=\hat{v}^{(m)})} - \log \left| \frac{H}{2\pi} \right| \Big|_{(\alpha_k=\hat{\alpha}_k^{(m)};v=\hat{v}^{(m)})}. \tag{3.11}$$

5. Iterate steps 2, 3, and 4 until it converges.

For the correlated random effects setting, the estimating procedure is similar with that stated above except in the step 1. The format of that variance structure becomes $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. And the three parts of the parameters become $\alpha = \{\beta_1, \beta_2, \tau_1, \tau_2\}, \theta = \{\sigma_1^2, \sigma_2^2, \rho\}$.

4.0 BIAS REDUCTION FOR H-LIKELIHOOD METHOD

We identify three possible issues when applying h-likelihood method in estimating unknown parameters in our proposed joint model. First, the h-likelihood method treats random terms as unknown nuisance parameters. When the sample size increases, the number of nuisance parameters increase as well. This might lead to non-negligible bias to the coefficient parameter estimation. The second potential issue of using h-likelihood estimator is that it could be substantially biased under high censoring situation (Jeon and Hsu, 2011). For time-to-event data with competing risks, censoring rate should include both noninformative and informative dropouts. Therefore, it is worth to explore appropriate bias reduction procedure when applying h-likelihood approach to our proposed model. The third issue is related to the binary outcome data. If we view time-to-event data as a series of binary outcomes over time, bias correction might be necessary under our setting. Breslow and Lin (1995) indicated that under bivariate binary outcomes, the estimation of the covariate coefficients could have a negligible asymptotic bias if variance of the random terms were incorrectly specified.

In this study, we propose four different bias reduction methods that aim to improve the estimation because of the three possible issues described above. The details of each of the methods are summarized in Sections 4.1 – 4.4.

4.1 BIAS REDUCTION 1 – CORRECTION ON ESTIMATES OF RANDOM EFFECTS

A bias correction procedure was introduced by Jeon and Hsu (2011) towards the h-likelihood approach on the frailty models. The procedure showed benefits in the situation of data with

high censoring rates. We expand the idea of this correction procedure to our proposed joint model. H-likelihood method has a working assumption that $\hat{v}_i|v_i \sim N(v_i, \gamma_i^2)$, where γ_i^2 is the variance of $\hat{v}_i$ obtained directly from H^{-1}, the inverse hessian matrix in Equation (3.9). Given $v_i \sim N(0, \sigma^2)$, $(v_i, \hat{v}_i)$ follows a bivariate normal distribution with the form

$$\begin{pmatrix} v_i \\ \hat{v}_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 + \gamma_i^2 \end{pmatrix}\right).$$

Therefore, the conditional distribution of $v_i|\hat{v}_i$ follows $N(\xi_i\hat{v}_i, \sigma^2(1-\xi_i))$, where $\xi_i = \sigma^2/(\sigma^2 + \gamma_i^2)$. We can use this conditional distribution to calculate the conditional mean of v_i, which is the mean function of v_i given the estimated $\hat{v}_i$.

For the shared random effects structure, the following two steps are implemented in the h-likelihood procedure in Section (3.3.3) to reduce the bias of $\hat{v}_i$. In the mth step,

1. Calculate $\xi_i^{(m)} = \hat{\sigma}^{2(m-1)}/(\hat{\sigma}^{2(m-1)} + \hat{\gamma}_i^{2(m)})$, where $\hat{\sigma}^{2(m-1)}$ is the variance estimate of $v_i^{(m-1)}$, and $\hat{\gamma}_i^{2(m)}$ is the variance of $\hat{v}_i^{(m)}$ given v_i.
2. Estimate parameters in Step 2 of Equation (4.3) and in Step 3 of Equation (3.11) of the h-likelihood procedure by replacing $\hat{v}_i$ with $E[v_i|\hat{v}_i] = \xi_i\hat{v}_i$ and $\exp(\hat{v}_i)$ with $E[\exp(v_i)|\hat{v}_i] = \exp(\xi_i\hat{v}_i + \sigma^2(1-\xi_i)/2)$.

The corresponding bias reduction procedure under the correlated random effects structure is provided in Appendix C.

4.2 BIAS REDUCTION 2 – CORRECTION FOR ZERO ESTIMATES OF VARIANCE COMPONENT

The method based on optimizing adjusted profile h-likelihood function (3.11) might lead to zero estimation for variance component σ^2. The zero estimation may cause a problem on constructing the confidence interval of v_i. It will also influence the accuracy of the estimators of other parameters, which may be a potential reason for nonconvergence in the estimation procedure. Morris (2006) identified this problem and proposed an adjustment on

the likelihood function under the linear mixed model context. Ha et al. (2013) expanded this adjustment procedure to the estimation of the h-likelihood dispersion parameters. Their idea is to add one extra term to the adjusted profile h-likelihood function for estimating σ^2. For our model, we let $h_{\alpha,v}$ and $h^*_{\alpha,v}$ be the profile h-likelihood and the adjusted profile h-likelihood, respectively. Therefore,

$$h^*_{\alpha,v} = h_{\beta,\tau,v} + \log\det(\sigma^2), \tag{4.1}$$

where $\sigma^2 = \Sigma$ for correlated random effects model, and Σ is a variance matrix for correlated random effects model. The term $\log\det(\sigma^2)$ is asymptotically negligible (Ha et al., 2013; Li and Lahiri, 2010). Therefore, the estimators from optimizing $h^*_{\beta,\tau,v}$ is asymptotically equivalent to the estimators from optimizing $h_{\beta,\tau,v}$ in 3.11. The adjusted profile h-likelihood has the following property:

$$\exp(h^*_{\alpha,v}) = \exp(h_{\alpha,v})\det(\Sigma) \geq 0.$$

Notice that $\exp(h^*_{\alpha,v}) = 0$ only if $\det(\Sigma) = 0$ (Appendix of Li and Lahiri (2010)). This insures that the zero estimate of the dispersion parameter could be avoided.

4.3 BIAS REDUCTION 3 – MODIFICATION ON ADJUSTED PROFILE H-LIKELIHOOD

Adjusted profile h-likelihood (APHL) in () and () has the same form with the adjusted profile likelihood proposed by Cox and Reid (1987). As defined before, α contains parameters of interest, and θ contains nuisance parameters. The modified profile likelihood $h_M(\alpha)$ of parameter α has the form

$$h_M(\alpha) = h_p(\alpha) + M(\alpha).$$

Based on the original version of $M(\alpha)$ proposed by Barndorff-Nielsen (1980, 1983), Barndorff-Nielsen (1980, 1983) proposed to use the following specification for the term $M(\alpha)$

$$M(\alpha) = -\frac{1}{2}\log|-l_{\theta\theta}(\alpha,\hat{\theta}_\alpha)| + D(\alpha),$$

where $\hat{\theta}_\alpha$ is the maximum likelihood estimate of θ given a fixed value of α, $D(\alpha) = \left|\frac{\partial\hat{\theta}_\alpha}{\partial\hat{\theta}}\right|$, $\hat{\theta}$ is the overall maximum likelihood estimate of θ, and $l_{\theta\theta} = \partial^2 h(\alpha,\theta)/\partial\theta\partial\theta^T$. In practice, the correction term $D(\alpha)$ is difficult to compute. When α and θ are orthogonal, $D(\alpha) = 0$. Under this orthogonal parametrization assumption, $M(\alpha)$ has the same form as the APHL. However, the orthogonal parametrization assumption between α and θ is often too strong. To balance the computation complexity and assumption, Severini (1998) proposed an approximation that has the form

$$M(\alpha) = -\frac{1}{2}\log|-l_{\theta\theta}(\alpha,\hat{\theta}_\alpha)| - \log|I_{\theta\theta}(\hat{\alpha},\hat{\theta};\alpha,\hat{\theta}_\alpha)|,$$

where $I_{\theta\theta}(\hat{\alpha},\hat{\theta};\alpha,\hat{\theta}_\alpha) = E_{\hat{\alpha},\hat{\theta}}\{l_\theta(\hat{\alpha},\hat{\theta})l_\theta\alpha,\hat{\theta}_\alpha\}$, where $l_\theta = \partial l(\alpha,\theta)/\partial\theta$. For the competing risks data, an empirical form of $I_{\theta\theta}(\hat{\alpha},\hat{\theta};\alpha,\hat{\theta}_\alpha)$ can be expressed as

$$\hat{I}_{\theta\theta}(\hat{\alpha},\hat{\theta};\alpha,\hat{\theta}_\alpha) = \left\{\frac{\partial\log h(T|Z,v;\alpha,\hat{\theta}_\alpha)}{\partial\theta}\right\}^T \frac{\partial\log h(T|Z,v;\alpha,\hat{\theta}_\alpha)}{\partial\theta}.$$

and hence,

$$M(\alpha) = -\frac{1}{2}\log|-l_{\theta\theta}(\alpha,\hat{\theta}_\alpha)| - \hat{I}_{\theta\theta}(\hat{\alpha},\hat{\theta};\alpha,\hat{\theta}_\alpha).$$

Therefore, the modified APHL for α, and θ have form

$$\begin{aligned} h_{M\alpha} &= h|_{(v=\hat{v})} + M(\alpha), \\ h_{M\alpha} &= h|_{(\alpha=\hat{\alpha},v=\hat{v})} + M(\theta), \end{aligned} \tag{4.2}$$

where h is the h-likelihood function defined in Section 3.3.3. Similar formula for modified profile likelihood in Equation (4.2) has been proposed by Bartolucci et al. (2012) for models with panel data with mutually independent subjects within the same panel.

4.4 BIAS REDUCTION 4 – CONSIDERATION OF DEPENDENCE BETWEEN RANDOM EFFECTS AND COEFFICIENT PARAMETERS

The derivatives of the random terms and other coefficients, $(\partial\hat{v}/\partial\alpha, \partial\hat{v}/\partial\theta)$, would be a potential reason of bias in parameter estimation for data with binary outcomes (Lee, 2004). As we mentioned earlier, time-to-event data with censoring (noninformative or informative due to competing risks) can be viewed as a series binary outcomes over time; therefore, we may have issue of biased estimation. It is worth noting that this type of bias will not attenuate with sample size increasing. That is because the number of random terms increase when sample size increases. To account for the dependence between the random terms v and other parameters, we cannot ignore the derivatives $\partial\hat{v}/\partial\alpha$ and $\partial\hat{v}/\partial\theta$ in the APHL (Ha and Lee (2003); Ha et al. (2011)).

Let α and θ be the coefficients and dispersion parameters defined in Section 3.3.3. Given $g(\alpha_s) = \partial h/\partial v|_{v=\hat{v}} = 0$, we can calculate $\partial\hat{v}/\partial\alpha_s$, where $s = 1, , 2, \cdots$, the length of α , as

$$\frac{\partial^2 g(\alpha_s)}{\partial\alpha_s} = \frac{\partial^2 h}{\partial v\partial\alpha_s}\bigg|_{v=\hat{v}} + \left(\frac{\partial^2 h}{\partial v^2}\bigg|_{v=\hat{v}}\right)\left(\frac{\partial\hat{v}}{\partial\alpha_s}\right) = 0,$$

$$\Rightarrow \frac{\partial\hat{v}}{\partial\alpha_s} = -\left(-\frac{\partial^2 h}{\partial v\partial v^T}\right)^{-1}\left(-\frac{\partial^2 h}{\partial v\partial\alpha_s}\right)\bigg|_{v=\hat{v}}. \tag{4.3}$$

Similarly, $\partial\hat{v}/\partial\theta$ can be derived in the same way as

$$\frac{\partial\hat{v}}{\partial\theta} = -\left(-\frac{\partial^2 h}{\partial v\partial v^T}\right)^{-1}\left(-\frac{\partial^2 h}{\partial v\partial\theta}\right)\bigg|_{\alpha=\hat{\alpha},v=\hat{v}}. \tag{4.4}$$

We will then substitute $\partial\hat{v}/\partial\alpha$ and $\partial\hat{v}/\partial\theta$ to the adjusted profile h-likelihood equation. The details can be found in Appendix B.

Of the four proposed bias reduction methods, the first one make correction on the estimate variances, weighting them by the dispersion parameters. The second method corrects the zero estimation of dispersion terms, by adding a term on the likelihood function; The third method corrects the dependence between the random terms and other parameters of coefficients; The fourth method does modification on the APHL.

5.0 SIMULATION

5.1 JOINT MODELING APPROACH FOR COMPETING RISKS DATA

5.1.1 Data generation

In this section, we compare the performance of the proposed model and the three optimization methods that we used to estimate the covariate effects. The data were generated based on assuming cause-specific hazards settings for each type of event. Two types of events were generated, the main event of interest and the dependent competing event.

We generated the time to the event of interest (T_1) and the time to a competing event (T_2), respectively. An identifiable dependence between the main event and competing event were represented by shared or correlated random effects struture. All three dependent scenarios (positive, negative, and independent) were considered. For each scenario, we generated a dataset consisting of $n = 400$ independent samples. We assumed a following standard exponential distributions for T_1 and T_2. For the shared random effects setting, the hazards function for T_1 and T_2 were $\lambda_{1i} = \exp(\beta_1^T Z_i + v_i)$ and $\lambda_{2i} = \exp(\beta_2^T Z_i + \phi v_i)$, respectively, where the random terms, v_i were generated from a standard normal distribution parameter $\phi = 1, -1$, or 0 represents a positive, a negative, or an independent association. For the correlated random effects setting, the hazards function for T_1 and T_2 $\lambda_{1i} = \exp(\beta_1^T Z_i + v_{1i})$ and $\lambda_{2i} = \exp(\beta_2^T Z_i + v_{2i})$, respectively, where $(v_{1i}, v_{2i})^T \sim N((0,0)^T, \Sigma)$, and $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Parameter $\rho = 0.3(-0.3, 0)$ represents a positive, a negative, or an independent association. The censoring times, C_i, were generated from an exponential distribution with parameters set to achieve an approximate 15% right

censoring rate. We then obtained the observed times X from the minimm of T_1, T_2 and C (i.e. $X = T_1 \wedge T_2 \wedge C$). We used 500 Monte Carlo replications for each scenario considered.

5.1.2 Results

Tables 1, 2, 3, and 4 summarize the results related to the parameter estimates, including bias, average of the estimated standard errors, empirical Monte Carlo standard errors, and empirical 95% coverage probabilities.

Table 1 shows the estimates of covariate effects of Z on $T_k, (k = 1, 2)$, β_k under the shared random effects setting. Specifically, Tables 1(a), 1(b), and 1(c) summarize the values of β_k that correspond to the positive association ($\phi = 1$), negative association ($\phi = -1$), and independence ($\phi = 0$) between T_1 and T_2, respectively. From the results, the coverage rates for the Gaussian quadrature method and the Bayesian-MCMC method were reasonable, and both methods yielded small biases for the estimates of β_k. However, the Gaussian quadrature method performed the best in the scenarios of the positive association and independence, whereas the Bayesian-MCMC method performed the best in the scenario of negative association. Although the h-likelihood method yielded a bias that was slightly larger than that of the other two methods, the application of a bias correction procedure reduced the bias of the h-likelihood method adequately and produced coverage rates closer to the nominal level.

Table 2 shows the estimates of all other nuisance parameters. The Gaussian quadrature method performed the best in terms of bias and coverage rates. The Bayesian-MCMC method did not perform as well as the Gaussian quadrature method but performed better than the h-likelihood method, which tended to underestimate the frailty standard deviations. When we applied a bias correction procedure for the h-likelihood method, the performance improved slightly but was still inferior to the other methods. Because the bias correction procedure that we used was designed to reduce the bias of the estimates of β_k by improving the estimate of σ, we did not expect it to yield good estimates for the nuisance parameters.

Table 3 shows the estimates of covariate effects of Z on $T_k, (k = 1, 2)$, β_k under the correlated random effects setting. Specifically, Tables 3(a), 3(b), and 3(c) summarize the

values of β_k that correspond to the positive association ($\rho = 0.3$), negative association ($\rho = -0.3$), and independence ($\rho = 0$) between T_1 and T_2, respectively. From the results, the coverage rates for the Gaussian quadrature method and the Bayesian-MCMC method were reasonable, and both methods yielded small biases for the estimates of β_k. However, the Gaussian quadrature method performed the best in the scenarios of the positive association and independence, whereas the Bayesian-MCMC method performed the best in the scenario of the negative association. The h-likelihood method yielded a bias that was initially a bit larger than that of the other two methods.

Table 4 shows the estimates of all other nuisance parameters. The Gaussian quadrature method performed the best in terms of bias and coverage rates, and its results were in the reasonable range. The Bayesian-MCMC method did not perform as well as the Gaussian quadrature method but performed better than the h-likelihood method, which tended to underestimate the standard deviations.

5.2 BIAS REDUCTION ON H-LIKELIHOOD

Two bias reduction methods introduced in Chapter 4 were implemented in the simulations. The bias reduction proposed by Jeon and Hsu (2011) described in Section 4.1, and the method to adjust zero estimate of dispersion parameters described in Section 4.2. Although we also conducted simulations for the bias reduction method of Ha et al. (2011) in Section 4.3, and the modification of profile likelihood in Section 4.4, convergence issue occurred. This indicates that theoretical considerations need to be further studied for applying these methods to our proposed models.

To show the performance of these bias reduction methods, the shared random effects model was used in this set of simulations. We used the same generated data as described in Section 5.1.1. Associations considered between the main event of interest T_1 and the competing risks T_2 are positive, negative, and independent.

Table 1: Estimates of the main parameters of interest from the simulated joint model with shared random effects structure. Total 500 replicates with a sample size of 400 in each replicate. The true parameter values are $\beta_1 = 0.5$ and $\beta_2 = 0.25$. Censoring rate = 15%. Estimation methods used were the Gaussian quadrature method (GQ), the Bayesian-MCMC method, the h-likelihood method (HL), the cause-specific Cox PH model with a normal frailty (Cox frailty), and the cause-specific Cox PH model (Cox).

(a) Simulation results for $\phi = 1$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	β_1	0.0044	0.0979	0.1022	0.942
	β_2	0.0020	0.0987	0.1027	0.946
Bayesian	β_1	-0.0054	0.0957	0.0931	0.952
	β_2	-0.0085	0.0985	0.1016	0.944
HL	β_1	0.0143	0.0741	0.0791	0.902
	β_2	0.0176	0.0750	0.0802	0.906
Cox frailty	β_1	0.0775	0.0857	0.0926	0.782
Cox	β_1	0.1096	0.0799	0.0841	0.754

(b) Simulation results for $\phi = -1$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	β_1	0.0024	0.0986	0.0978	0.940
	β_2	-0.0021	0.0988	0.1017	0.946
Bayesian	β_1	0.0008	0.0955	0.0981	0.952
	β_2	0.0006	0.0995	0.0950	0.956
HL	β_1	0.0298	0.0742	0.0774	0.910
	β_2	0.0171	0.0749	0.0873	0.892
Cox frailty	β_1	-0.0471	0.0827	0.0874	0.872
Cox	β_1	0.0746	0.0780	0.789	0.834

(c) Simulation results for $\phi = 0$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	β_1	-0.0015	0.0982	0.1017	0.950
	β_2	0.0023	0.0780	0.0748	0.962
Bayesian	β_1	-0.0039	0.0909	0.0966	0.936
	β_2	0.0021	0.0790	0.0843	0.942
HL	β_1	0.0182	0.0728	0.0807	0.902
	β_2	0.0184	0.0756	0.0792	0.906
Cox frailty	β_1	0.0085	0.083	0.089	0.926
Cox	β_1	0.027	0.0773	0.0789	0.908

Table 2: Estimates of the nuisance parameters from the simulated joint model with shared random effects structure. Total 500 replicates with a sample size of 400 in each replicate. The true parameter values are $\sigma_1 = 1$ and $\sigma_2 = 1$. Censoring rate = 15%. Estimation methods used were the Gaussian quadrature method (GQ), the Bayesian-MCMC method, the h-likelihood method (HL).

(a) Simulation results for $\phi = 1$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	σ	0.0052	0.1387	0.1298	0.967
	ϕ	0.0312	0.2145	0.2107	0.965
Bayesian	σ	-0.0644	0.1419	0.1527	0.914
	ϕ	-0.1649	0.2936	0.3322	0.967
HL	σ	-0.6931	0.2642	0.4209	0.548
	ϕ	0.3272	0.0533	0.2564	0.616

(b) Simulation results for $\phi = -1$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	σ	0.0294	0.1580	0.1566	0.958
	ϕ	-0.0033	0.2423	0.2467	0.934
Bayesian	σ	0.1336	0.1890	0.2039	0.896
	ϕ	0.3140	0.5643	0.6312	0.988
HL	σ	-0.7012	0.1807	0.4213	0.560
	ϕ	0.0308	0.1185	0.2416	0.624

(c) Simulation results for $\phi = 0$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	σ	-0.0064	0.1680	0.1763	0.952
	ϕ	-0.1767	0.1594	0.1230	0.954
Bayesian	σ	0.3127	0.2381	0.2895	0.836
	ϕ	0.1522	0.9761	0.4468	0.980
HL	σ	-0.6616	0.1571	0.5027	0.616
	ϕ	-0.2084	0.0664	0.1696	0.662

Table 3: Estimates of coefficients of covariates from simulation of correlated random effects model. Totally 500 runs with sample size 400 each, and the setting values are $\beta_1 = 0.5$, $\beta_2 = 0.25$, censoring rate is 15%. GQ: Gaussian quadrature method; the Bayesian-MCMC method, the h-likelihood method (HL), the cause-specific Cox PH model with a normal frailty (Cox frailty), and the cause-specific Cox PH model (Cox).

(a) Simulation results for $\rho = 0.3$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	β_1	0.0084	0.1001	0.1024	0.942
	β_2	0.0081	0.1008	0.1023	0.944
Bayesian	β_1	-0.0165	0.1116	0.1010	0.966
	β_2	0.0273	0.1114	0.1043	0.952
HL	β_1	0.0178	0.0746	0.0834	0.918
	β_2	0.0143	0.0734	0.0824	0.912
Cox frailty	β_1	-0.0886	0.0876	0.0912	0.852
SEP	β_1	0.0934	0.0857	0.0918	0.748

(b) Simulation results for $\rho = -0.3$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	β_1	0.0038	0.1017	0.1060	0.944
	β_2	0.0001	0.1018	0.1037	0.952
Bayesian	β_1	-0.0153	0.1118	0.1055	0.958
	β_2	0.0212	0.1118	0.1003	0.962
HL	β_1	0.0184	0.0758	0.0892	0.916
	β_2	0.0182	0.0739	0.0838	0.896
Cox frailty	β_1	-0.0676	0.0766	0.0819	0.830
SEP	β_1	0.0751	0.0864	0.0928	0.827

(c) Simulation results for $\rho = 0$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	β_1	0.0035	0.1010	0.0097	0.960
	β_2	0.0039	0.1012	0.1023	0.942
Bayesian	β_1	-0.0098	0.1115	0.0980	0.958
	β_2	0.0174	0.1117	0.1003	0.962
HL	β_1	0.0176	0.7547	0.8226	0.908
	β_2	0.0199	0.7304	0.7958	0.910
Cox frailty	β_1	0.0032	0.0968	0.1028	0.927
SEP	β_1	0.0034	0.1009	0.0968	0.974

Table 4: Estimates of parameters in variance matrix of random terms from simulation of correlated random effects model. Totally 500 runs with sample size 400 each, and the setting values are $\sigma_1 = 1, \sigma_2 = 1$, censoring rate is 15%. GQ: Gaussian quadrature method; Bayesian: Bayesian-MCMC method; HL: H-likelihood method.

(a) Simulation results for $\rho = 0.3$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	σ_1	0.0169	0.1737	0.1691	0.968
	σ_2	-0.0063	0.1759	0.1667	0.962
	ρ	-0.0068	0.3004	0.2748	0.912
Bayesian	σ_1	0.5082	0.0678	0.0432	0.000
	σ_2	0.5084	0.0678	0.0405	0.000
	ρ	0.2848	0.0468	0.0240	0.000
HL	σ_1	0.2977	0.0624	0.2209	0.290
	σ_2	-0.2650	0.0574	0.3046	0.172
	ρ	0.3079	0.0627	0.218	0.294

(b) Simulation results for $\rho = -0.3$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	σ_1	0.0190	0.1748	0.1744	0.972
	σ_2	0.0180	0.1745	0.1742	0.972
	ρ	0.0070	0.2525	0.2480	0.948
Bayesian	σ_1	0.4742	0.0722	0.0428	0.000
	σ_2	0.4795	0.0716	0.0407	0.000
	ρ	-0.3857	0.0503	0.0216	0.000
HL	σ_1	0.4098	0.0619	0.1851	0.2361
	σ_2	-0.4862	0.0622	0.1809	0.2602
	ρ	0.4194	0.0330	0.0821	0.1030

(c) Simulation results for $\rho = 0$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
GQ	σ_1	0.0055	0.1733	0.1769	0.966
	σ_2	0.0041	0.1749	0.1712	0.964
	ρ	0.0150	0.2711	0.2805	0.958
Bayesian	σ_1	0.4935	0.0697	0.0414	0.000
	σ_2	0.4958	0.0696	0.0386	0.000
	ρ	-0.0496	0.0482	0.0221	0.970
HL	σ_1	0.3887	0.0634	0.2024	0.294
	σ_2	-0.3293	0.0630	0.2103	0.296
	ρ	0.3773	0.0448	0.1711	0.106

Table 5 and Table 6 show the simulation results for evaluating two bias reduction methods of h-likelihood method. HL-BD1 referred to the method described in Section 4.1 and HL-DB2 referred to the method described in Section 4.2. Simulation results of the h-likelihood method were also included in these tables for the comparison purpose. Tables 5(a), 5(b), and 5(c) show the results corresponding to the positive association ($\phi = 1$), negative association ($\phi = -1$), and independence ($\phi = 0$) between T_1 and T_2, respectively. When the association between the event of interest and the competing event was positive, the bias were slightly improved by using the method of Jeon and Hsu (2011), and the corresponding coverage rates became much more reasonable. For the method in which one term with dispersion parameter is added (HL-BD2), the bias of one coefficient were smaller, and the coverage rates were better than that without this bias reduction method. Moreover, by using HL-BD2, there was hardly any convergence issue. For the situations of negative relation, and independence between the main event of interest and the competing event, similar results can be drawn as that of the positive association. Although the bias reduction methods were implemented, the estimate of dispersion parameters are still not satisfactory (Table 6).

Table 5: Estimates of the main parameters of interest from the simulated joint model with shared random effects structure. Total number of replicates is 500 with a sample size of 400 in each replicate. The true parameter values are $\beta_1 = 0.5$ and $\beta_2 = 0.25$. Censoring rate = 15%. Estimation methods used were the h-likelihood method (HL), the h-likelihood method with bias reduction procedures (HL-BR1, HL-BR2).

(a) Simulation results for $\phi = 1$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
HL	β_1	0.0143	0.0741	0.0791	0.902
	β_2	0.0176	0.0750	0.0802	0.906
HL-BR1	β_1	0.0110	0.0737	0.0736	0.952
	β_2	0.0160	0.0756	0.0758	0.934
HL-BR2	β_1	-0.0151	0.0749	0.0937	0.917
	β_2	-0.0017	0.0759	0.0960	0.928

(b) Simulation results for $\phi = -1$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
HL	β_1	0.0298	0.0742	0.0774	0.910
	β_2	0.0171	0.0749	0.0873	0.892
HL-BR1	β_1	-0.0104	0.0813	0.0817	0.940
	β_2	0.0174	0.1133	0.1145	0.966
HL-BR2	β_1	-0.0275	0.0698	0.0907	0.908
	β_2	0.0030	0.0727	0.0775	0.922

(c) Simulation results for $\phi = 0$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
HL	β_1	0.0182	0.0728	0.0807	0.902
	β_2	0.0184	0.0756	0.0792	0.906
HL-BR1	β_1	-0.0097	0.0797	0.0879	0.944
	β_2	0.0109	0.1175	0.1206	0.960
HL-BR2	β_1	-0.0134	0.0638	0.0826	0.922
	β_2	-0.0022	0.0660	0.0680	0.946

Table 6: Estimates of the nuisance parameters from the simulated joint model with shared random effects structure. Total 500 replicates with a sample size of 400 in each replicate. The true parameter values are $\sigma_1 = 1$ and $\sigma_2 = 1$. Censoring rate = 15%. Estimation methods used were the h-likelihood method (HL), the h-likelihood method with bias reduction procedures (HL-BR1, HL-BR2).

(a) Simulation results for $\phi = 1$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
HL	σ	-0.6931	0.2642	0.4209	0.548
	ϕ	0.3272	0.0533	0.2564	0.416
HL-BR1	σ	0.0010	0.0002	0.0004	0.630
	ϕ	0.2007	2.0614	0.0349	1.000
HL-BR2	σ	-0.0310	0.0885	0.2227	0.550
	ϕ	0.2171	0.1120	0.3930	0.427

(b) Simulation results for $\phi = -1$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
HL	σ	-0.7012	0.1807	0.4213	0.560
	ϕ	0.0308	0.1185	0.2416	0.624
HL-BR1	σ	0.0003	0.0004	0.0003	0.613
	ϕ	- 0.2194	2.8772	0.0337	1.000
HL-BR2	σ	-0.6510	0.2772	0.1737	0.000
	ϕ	0.3445	0.0520	0.0560	0.010

(c) Simulation results for $\phi = 0$

Estimation method	Parameter	Bias	SD-EST	SD-EMP	CR
HL	σ	-0.6616	0.1571	0.5027	0.616
	ϕ	-0.2084	0.0664	0.1696	0.662
HL-BR1	σ	-0.0002	0.0005	0.0016	0.643
	ϕ	0.2060	2.0071	0.0363	1.000
HL-BR2	σ	0.3049	0.1461	0.1729	0.396
	ϕ	-0.1984	0.0576	0.1615	0.408

6.0 APPLICATION

We analyzed data from a prospective cohort study of the Monongahela Valley Independent Elders Survey (MoVIES). One of the objectives of the MoVIES study was to identify risk factors associated with Alzheimers disease and other forms of dementia. The original cohort of the study was assembled between 1987 and 1989 and included 1,681 individuals aged 65 years or older. Until 2002, members of the cohort were screened once every 2 years for cognitive function and risk factors.

There were 1,204 individuals who had information on the risk factors of interest. Among these, 151 developed dementia, 506 dropped out because of death or severe illness, and the remaining 547 were lost to follow-up or were censored at the end of the study, as shown in Table 7. People who died or were too sick to continue participating in the study might be likely to develop dementia, therefore, the 506 dropouts might have had a greater chance of developing dementia if they had continued in the study. Hence, we considered them to be informative dropouts in our analysis.

Previous studies showed that several factors are potentially associated with the progression of dementia. Ganguli et al. (2005, 2013) showed that drinking alcohol lead to slower decline of several cognitive composite domain scores, and has protective effect on incidence mild cognitive impairment (MCI). Chang et al. (2012) identified that smoking is a risk factor for Alzheimer's disease (AD) by considering the age-specific mortality rates and adjusting for competing risks. Depression associates with cognitive impairment cross-sectionally (Ganguli et al., 2006). Dodge et al. (2011) indicated that the risk of AD could be reduced by 9.1% for non-apolipoprotein E (APOE) *4 carriers, by calculating the population-attributable risk percent.

Table 7: Descriptive statistics of potential risk factors for dementia stratified by participants who progressed to dementia, had informative dropouts because of death or too sick, and were censoring. For categorical variables, frequencies (%) are shown; and for continuous variables, mean (standard deviation [SD]).

	Progressed	Death/too sick	Censoring
Total	151	506	547
Age (mean/SD)	77.67/5.24	76.20/5.60	72.30/3.84
Female (n/%)	104/68.19	267/68.87	373/52.77
Medications ≥ 3 (n/%)	54/20.47	219/35.76	112/43.28
Smoking (n/%)	19/12.79	60/5.96	45/12.45
Depression (n/%)	9/8.29	63/12.58	70/11.95
Greater than high school (n/%)	81/53.64	274/54.15	377/68.92
APOE 4 (n/%)	30/27.52	44/20.47	102/19.92
LIPID use (n/%)	4/2.65	12/2.37	16/2.93
NASID use (n/%)	111/73.51	345/68.18	381/69.65
Poor health status (n/%)	1/0.66	17/3.36	6/1.10
Drinking (n/%)	68/45.03	242/47.92	344/62.89

For this study, variables considered were age (in years), gender, smoking status (yes / no), education level (greater than high school or not), alcohol consumption (yes/no), presence of depression measured by the modified Center for Epidemiological Studies-Depression (mCESD) score (< 5 or ≥ 5), APOE *4 carrier (yes/no), number of prescription medications (< 3 or ≥ 3), lipid use, non-steroidal anti-inflammatory drug (NSAID) use, and whether having a poor general health status. During this period, most of the missing data related to alcohol consumption and APOE*4. Wave 2 of the MoVIES is the baseline of the current study. Several risk factors were not measured until wave 2. The frequencies and percentages for categorical variables, and means and standard deviations for continuous variables are provided in Table 7.

Three models were applied to the data, and the estimation results are shown in Table 8. The three models are: (1) the proposed joint modeling method with shared random effects structure, and Weibull baseline hazards function, (2) the cause-specific Cox proportional hazards (PH) model, and (3) the Fine and Gray proportional subdistribution hazards function model. As shown in Table 8, age, APOE*4 gene, and depression were identified as risk factors for dementia by all 3 methods. In particular, the estimated hazards ratios of older age were 2.13, 2.10, and 1.9 from the joint model, Cox PH model, and Fine and Gray model, respectively. The joint model estimates that presence of APOE*4 gene and depression increase the risk of dementia by 1.67 and 1.89 times, respectively. The Cox PH model (APOE*4's HR=1.68; depression's HR=1.95) and Fine and Gray (APOE*4's HR=1.67; depression's HR=1.85) yielded comparable estimates.

In summary, the joint model captured same significant risk factors as did the other two models, which may because that the estimated correlation parameter ϕ ($p = 0.60$) and the standard deviation of the random effects σ ($p = 0.62$) were not significantly different from 0. Therefore, there was no significant association between the development of dementia and dropping out of the study because of death or being too sick.

Table 8: Estimated hazard ratio (HR) or subdistribution hazard ratio (SHR) of the risk factors for developing dementia. Three fitted models were the proposed joint model with shared random effects structure and Weibull baseline hazards function, a cause-specific Cox proportional hazards model, and the Fine and Gray proportional subdistribution hazards model.

	Joint Model			Cox			Fine & Gray		
Variable	HR	SE	P value	HR	SE	P value	HR	SE	P value
Model for the main event									
Smoke	0.559	0.466	0.212	0.586	0.467	0.253	0.549	0.458	0.190
Age	2.126	0.103	<0.001	2.100	0.103	<0.001	1.897	0.108	<0.0001
Female	1.067	0.228	0.776	1.039	0.233	0.870	1.073	0.232	0.760
>=high school	1.050	0.207	0.813	1.065	0.203	0.754	1.028	0.206	0.900
APOE 4	1.671	0.220	0.020	1.675	0.220	0.019	1.672	0.225	0.022
$\geq$ 3 Meds	0.945	0.225	0.800	0.955	0.222	0.836	0.879	0.233	0.580
LIPID use	1.959	0.559	0.230	1.791	0.560	0.298	1.878	0.519	0.220
NSAID use	1.361	0.237	0.193	1.394	0.238	0.163	1.334	0.248	0.250
Depression	1.894	0.297	0.0322	1.947	0.295	0.024	1.847	0.304	0.043
Poor general health	0.657	1.027	0.683	0.715	1.024	0.743	0.673	0.160	0.730
Drinking	0.731	0.202	0.122	0.763	0.144	0.060	0.776	0.144	0.077
Model for the informative dropouts									
Smoke	1.636	0.253	0.052						
Age	2.105	0.140	<0.001						
Female	0.496	0.194	<0.001						
>=high school	0.902	0.171	0.546						
APOE 4	1.130	0.203	0.547						
$\geq$ 3 Meds	1.882	0.194	0.001						
LIPID use	1.207	0.468	0.687						
NSAID use	1.297	0.185	0.159						
Depression	0.915	0.309	0.775						
Poor general healthy	1.511	0.798	0.605						
Drinking	0.780	0.171	0.146						
Other parameters									
τ_1	1.608	0.149	<0.001						
τ_2	6.856	0.631	<0.001						
ϕ	-4.344	9.896	0.596						
σ	0.163	0.390	0.616						

7.0 DISCUSSION

The presence of competing risks introduces challenges to the analysis of time-to-event data in medical research. The currently used methods: cause-specific hazards models and sub-distribution models do not allow for a flexible association between the main event and the competing events. On the other hand, copula models are mathematically complex and may be sensitive to the chosen form of the dependence structure. To deal with these challenges, we developed an approach that treats competing risks as informative dropouts, jointly models the time to the event of interest and the time to the competing events, and uses random terms to capture the dependence of the events. Our model assumes that conditional on the random terms, the main event and competing events are independent. The advantages of this conditional independence assumption are that it can handle a flexible association and lend ease of interpretation. In practice, we can fit separate cause-specific hazards models for each cause of failure, and then include the resulting covariates in submodels of the final joint model.

In our model, we provide two sturctures of fundamental likelihood function. The first is the shared random effects structure, and the second is the correlated random effects structure. The random effects in the first structure have a simpler distribution than those in the second structure. The first structure requires less unknown parameters and is easier to be extended if more competing risks are present. For example, in the shared random effects structure, if one more competing event is added, then the analyst needs to add only one more cause-specific frailty submodel with a random effect of $\phi_{K+1}v_i$ to the existing likelihood function. This is because the shared random effects structure assumes a proportional dependence between the main event and the competing event via parameter ϕ. Unfortunately, this assumption may not hold in practice. In contrast, in the correlated random effects structure, if one

more competing event is added, the dimensions of the variance-covariance matrix Σ would increase. This greatly enlarges the burden of computations but at the same time allows for a flexible dependence of unobserved variations. Even though both structures can indicate the direction of association between the main event and the competing events, the correlated random effects structure can better define the dependence. In practice, whether to use one structure or the other or to combine the structures will depend on the purpose of the study.

To estimate unknown parameters, we provide three likelihood optimization methods: the Gaussian quadrature method, the Bayesian-MCMC method, and the h-likelihood method. All three can approximate multidimensional integration of the random terms. The Gaussian quadrature method performs the best in terms of bias and coverage rates. If the SAS procedure NLMIXED is used for the Gaussian quadrature method, it is necessary to specify a parametric form of the likelihood function. The Bayesian-MCMC method produces reasonable estimates in our simulation studies and can be implemented with a standard software package (e.g., winBUGs), but it requires analysts to specify prior distributions for all parameters. The h-likelihood method is expected to be computationally efficient because it uses the Newton-Raphson approach. Simulations show that it performs well after a bias correction procedure is applied and that it can deal with both parametric and semiparametric likelihood functions. In practice, if nuisance parameters are of interest, we do not recommend this method because simulations show that the bias is large and the coverage rate is incorrect for estimating these parameters. Moreover, the method requires special programming, which we have developed and is available upon request from the study authors. A potential limitation of the proposed method is that it assumes the submodel of the informative dropouts is correct. In the other words, all covariates related to the informative dropouts are included, or captured by the random effects. In the future, we will study the impact of model misspecification on the estimation.

Although we prefer the Gaussian quadrature method, we recognize that convergence might be a problem with some initial values that are chosen. If convergence cannot be reached, analysts can first use the Bayesian-MCMC or h-likelihood method to find appropriate initial values and then incorporate these values into the Gaussian quadrature method. For the h-likelihood, we considered bias reduction methods for the parameter estimation, es-

pecially for the nuisance parameters. However, the bias reduction methods used still failed to meet expected improvement in estimation. Future development on bias reduction techniques are still needed.

Another future work with respect to the proposed model is to further explore and study the performance of the proposed model under varying sample sizes and independent censoring rates. In addition, we plan to expand the likelihood function of the proposed joint models to work under a semiparametric framework, that is, the use of an arbitrary baseline hazards function rather than a specific parametric distribution.

APPENDIX A

AN EXAMPLE OF THE MCMC SAMPLING ALGORITHM

The following steps describes the Metropolis-Hastings algorithm of sampling random numbers $\mathbf{\Omega}_s^{\prime(b)}, b = 1, \cdots, B$ from the posterior distribution (3.8).

(1) Give an initial value, $\mathbf{\Omega}_s^{\prime(0)}$ of $\mathbf{\Omega}_s^{\prime}$. one might want to try different initial values for a robust Bayesian inference.

(2) Generate a random number $\tilde{\Omega}$, from the assigned prior distribution with density function $\pi(\mathbf{\Omega}_s^{\prime})$. Generate a random number, ν from a standard uniform distribution $U(0, 1)$.

(3) Therefore, the bth sampling iteration of $\mathbf{\Omega}_s^{\prime}$ is obtained from

$$\mathbf{\Omega}_s^{\prime(b+1)} = \begin{cases} \tilde{\Omega} & \text{if } \nu \leq \min\left(1, \frac{L_i(\mathbf{O}_i|\tilde{\Omega})\pi(\tilde{\Omega})}{L_i(\mathbf{O}_i|\mathbf{\Omega}_s^{\prime(b)})\pi(\mathbf{\Omega}_s^{\prime(b)})}\right) \\ \mathbf{\Omega}_s^{\prime(b)} & \text{otherwise} \end{cases}.$$

(4) Iterate steps (2) and (3) B times to obtain a chain of random numbers $\{\mathbf{\Omega}_s^{\prime(b)}, b = 1, \cdots, B\}$.

APPENDIX B

DERIVATIVES OF THE H-LIKELIHOOD

In estimating v_i (step 2 in section 3.3.3), the gradient $\frac{\partial h}{\partial v}$ and the hessian matrix H are needed in the Newton-Raphson procedure, which can be calculated as following,

$$\begin{aligned}\frac{\partial h}{\partial v_i} =&(\delta_{1i} - \mu_{1i}) + \phi(\delta_{2i} - \mu_{2i}) + \frac{\partial l_{3i}}{\partial v_{ki}}\\ -\frac{\partial^2 h}{\partial v_{i'}\partial v_i} =&\mu_{1i} + \phi^2\mu_{2i} + \frac{\partial^2 l_{3i}}{\partial v_{ki'}\partial v_{ki''}},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial l_{3i}}{\partial v_{ki'}} =& -\frac{v_i}{\sigma^2}\\ \frac{\partial^2 l_{3i}}{\partial v_{i'}\partial v_{i''}} =& \begin{cases} \frac{1}{\sigma^2} & \text{if } i = i' = i'' \\ 0 & \text{otherwise} \end{cases}.\end{aligned}$$

Therefore, the hessian matrix H has the form

$$H = diag\left\{(\mu_1 + \phi^2\mu_2)\right\} + 1/\sigma^2 I_n = \mathcal{I}^T W \mathcal{I} + m,$$

where I_n denotes the $n \times n$ identity matrix, $\mathcal{I} = \begin{pmatrix} 1_{n\times 1} & 0 \\ 0 & \phi 1_{n\times 1} \end{pmatrix}$, $m = 1/\sigma^2 I_n$, and $W = diag\{(\mu_1^T, \mu_2^T)^T\}$.

After estimating random effects $\hat{v}$, Ha et al. (2007) proposed to use the restricted maximum likelihood (REML) estimators of $\alpha = \{\beta_1, \beta_2, \tau_1, \tau_2, \phi\}$ and $\theta = \{\phi\}$, respectively,

as in steps 3 and 4 in Section 3.3.3. In order to estimate α and θ, we will need the first and second derivatives of h_v with respect to α, and $h_{(\alpha,v)}$ with respect to θ. Let $\hat{h}_v = h|_{(v=\hat{v})}, \hat{H} = H|_{(v=\hat{v})}, \hat{h}_{\alpha,v} = h|_{(\alpha=\hat{\alpha},v=\hat{v})}, \hat{H}_{\alpha,v} = H|_{(\alpha=\hat{\alpha};v=\hat{v})}$, where $\hat{h}_v$ and $\hat{H}$ denote the profile likelihood and the observed information matrix evaluated at the current estimates of $\hat{v}$, respectively; $\hat{h}_{\alpha,v}$ and $\hat{H}_{\alpha,v}$ denote the profile likelihood and the observed information matrix evaluated at the current estimates of $(\hat{\alpha}, \hat{v})$, respectively. Separate the constant term 2π out of H and $\hat{H}_{\alpha,v}$, equations (4.3) and (3.11) can be rewritten as $h_v = \hat{h}_v - \log|\hat{H}| + n\log(2\pi), h_{(\alpha,v)} = \hat{h}_{\alpha,v} - \log|\hat{H}_{\alpha,v}| + (n+p)\log(2\pi)$, respectively. Therefore, the estimators of $\hat{\alpha_k}$ and $\hat{\theta}$ can be obtained by solving the following equations iteratively: $\frac{\partial h_v}{\partial \alpha} = 0$ and $\frac{\partial h_{(\alpha,v)}}{\partial \theta} = 0$.

To estimate β_k, we express the first and the second derivatives of h_v with respect to α as

$$\frac{\partial h_v}{\partial \alpha_s} = \frac{\partial \hat{h}_v}{\partial \alpha_s} - \frac{1}{2}\mathrm{tr}\left(\hat{H}^{-1}\frac{\partial \hat{H}}{\partial \alpha_s}\right) \tag{A.1}$$

$$-\frac{\partial^2 h_v}{\partial \alpha_s' \alpha_s} = -\frac{\partial^2 \hat{h}_v}{\partial \alpha_s' \alpha_s} + \mathrm{tr}\left(-\hat{H}^{-1}\frac{\partial \hat{H}}{\partial \alpha_{s'}}\hat{H}^{-1}\frac{\partial \hat{H}}{\partial \alpha_s} + \hat{H}^{-1}\frac{\partial^2 \hat{H}}{\partial \alpha_{s'}\partial \alpha_s}\right), \tag{A.2}$$

where $\alpha_s, \alpha_s'; s, s' = 1, 2, \cdots, S$ are the sth and s'th components of α, S is the length of α, and $\mathrm{tr}(A)$ is the trace of matrix A.

The first and the second derivatives of h_v and the hessian matrix $\hat{H}$ with respect to α_k in equations (A.1) and (A.2) are

$$\begin{aligned}
\frac{\partial \hat{h}_v}{\partial \beta_k} &= \left.\frac{\partial h_v}{\partial \beta_k}\right|_{v=\hat{v}} + \left.\frac{\partial h}{\partial v}\frac{\partial \hat{v}}{\partial \beta_k}\right|_{v=\hat{v}} \\
&= \sum_i Z_i^T(\delta_{ki} - \mu_{ki}) \\
\frac{\partial \hat{h}_v}{\partial \tau_k} &= \left.\frac{\partial h_v}{\partial \tau_k}\right|_{v=\hat{v}} + \left.\frac{\partial h}{\partial v}\frac{\partial \hat{v}}{\partial \tau_k}\right|_{v=\hat{v}} \\
&= \sum_i \log(X_i)(\delta_{ki} - \mu_{ki}) + \tau_k^{-1}\sum_i \delta_{ki} \\
\frac{\partial \hat{h}_v}{\partial \phi} &= \left.\frac{\partial h_v}{\partial \phi}\right|_{v=\hat{v}} + \left.\frac{\partial h}{\partial v}\frac{\partial \hat{v}}{\partial \phi}\right|_{v=\hat{v}} \\
&= \sum_i v_i(\delta_{2i} - \mu_{2i}).
\end{aligned}$$

Note that $\partial h/\partial v|_{v=\hat{v}} = 0$ in above equations, because $\hat{v}$ is obtained from solving $\partial h/\partial v|_{v=\hat{v}} = 0$. And,

$$-\frac{\partial^2 \hat{h}_v}{\partial \alpha_{s'} \partial \alpha_s} = -\frac{\partial^2 h_v}{\partial \alpha_s' \partial \alpha_s}\bigg|_{v=\hat{v}} + \frac{-\partial^2 h_v}{\partial \alpha_{s'} \partial v}\left(\frac{\partial \hat{v}}{\partial \alpha_s}\right)\bigg|_{v=\hat{v}},$$

where

$$\begin{aligned} -\frac{\partial^2 h_v}{\partial \alpha_s' \partial \alpha_s} &= -\begin{pmatrix} \frac{\partial^2 \hat{h}_v}{\partial \beta^T \partial \beta} & \frac{\partial^2 \hat{h}_v}{\partial \tau^T \partial \beta} & \frac{\partial^2 \hat{h}_v}{\partial \phi \partial \beta} \\ \frac{\partial^2 \hat{h}_v}{\partial \beta^T \partial \tau} & \frac{\partial^2 \hat{h}_v}{\partial \tau^T \partial \tau} & \frac{\partial^2 \hat{h}_v}{\partial \phi \partial \tau} \\ \frac{\partial^2 \hat{h}_v}{\partial \beta^T \partial \phi} & \frac{\partial^2 \hat{h}_v}{\partial \tau^T \partial \phi} & \frac{\partial^2 \hat{h}_v}{\partial \phi^2} \end{pmatrix} \\ &= \begin{pmatrix} Z^T W Z & Z^T W \mathcal{X} & Z^{*T} W V \\ \mathcal{X}^T W Z & \mathcal{X}^T W \mathcal{X} + M & \mathcal{X}^{*T} W V \\ V^T W Z^* & V^T W \mathcal{X}^* & V^T W V \end{pmatrix}, \end{aligned}$$

where Z, $\mathcal{X}$, Z^*, $\mathcal{X}^*$, V and M are block diagonal matrices, defined as

$$Z = \begin{pmatrix} Z_1 & 0 \\ 0 & Z_2 \end{pmatrix}, \mathcal{X} = \begin{pmatrix} \log X & 0 \\ 0 & \log X \end{pmatrix}, Z^* = \begin{pmatrix} 0_{\dim(Z_1)} & 0 \\ 0 & Z_2 \end{pmatrix},$$
$$\mathcal{X}^* = \begin{pmatrix} 0_{n\times 1} & 0 \\ 0 & \log X \end{pmatrix}, \text{and } M = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix},$$

where $M_k = (1/\sigma^2)\sum_i^n \delta_{ki}$.

Then, the calculation of $-\partial^2 h_v/\partial\alpha_{s'}\partial v$ is provided as,

$$\begin{aligned}
\frac{-\partial^2 h_v}{\partial\beta_1\partial v} &= Z_1\mu_1 \\
\frac{-\partial^2 h_v}{\partial\beta_2\partial v} &= \phi Z_2\mu_2 \\
\frac{-\partial^2 h_v}{\partial\tau_1\partial v} &= \log X\mu_1 \\
\frac{-\partial^2 h_v}{\partial\tau_2\partial v} &= \phi\log X\mu_2 \\
\frac{-\partial^2 h_v}{\partial\phi\partial v} &= \delta_2 - \mu_2.
\end{aligned} \tag{A.3}$$

Since $\hat{v}(\alpha)$ may be function of α (Section 4.3), it is necessary to derive the derivative of $\hat{v}$ on α. From the Equation 4.4, the derivative $\partial\hat{v}/\partial\alpha_s$ is

$$\frac{\partial\hat{v}}{\partial\alpha_s} = -\left(-\frac{\partial^2 h}{\partial v\partial v^T}\right)^{-1}\left(-\frac{\partial^2 h}{\partial v\partial\alpha_s}\right)\Bigg|_{v=\hat{v}} \tag{A.4}$$

where $\left(-\frac{\partial^2 h}{\partial v\partial v^T}\right) = H$, and $-\partial^2 h/\partial v\partial\alpha_s$ is derived in Equation (A.3).
The next term need to be derived in Equation (A.1) is $\partial\hat{H}/\partial\alpha_s$, which has form

$$\frac{\partial\hat{H}}{\partial\alpha_s} = \left(\frac{\partial H}{\partial\alpha_s} + \frac{\partial H}{\partial v}\frac{\partial\hat{v}}{\partial\alpha_s}\right)\Bigg|_{v=\hat{v}}, \tag{A.5}$$

Because $H = \mathcal{I}^T W\mathcal{I} + m$ and m is not a function of α or v, to calculate $\frac{\partial\hat{H}}{\partial\beta_s}$ and $\frac{\partial^2\hat{H}}{\partial\alpha_s'\partial\alpha_s}$, we only need to calculate $\frac{\partial W}{\partial\alpha_k}$ and $\frac{\partial^2 W}{\partial\alpha_k'\partial\alpha_k}$ via the forms:

$$\frac{\partial\hat{H}}{\partial\alpha_q} = \mathcal{I}^T\left(\frac{\partial W}{\partial\alpha_q} + \frac{\partial W}{\partial v}\frac{\partial\hat{v}}{\partial\alpha_q}\right)\mathcal{I}\Bigg|_{v=\hat{v}},$$

where

$$\begin{aligned}
\partial W/\partial \beta_1 =& diag[\{(Z_1\mu_1)^T, 0_{1\times n}\}^T] \\
\partial W/\partial \beta_2 =& diag[\{0_{1\times n}, (Z_2\mu_2)^T\}^T] \\
\partial W/\partial \tau_1 =& diag[\{(\log X\mu_1)^T, 0_{1\times n}\}^T] \\
\partial W/\partial \tau_2 =& diag[\{0_{1\times n}, (\log X\mu_2)^T\}^T] \\
\partial W/\partial \phi =& diag[\{0_{1\times n}, (v\mu_2)^T\}^T] \\
\partial W/\partial v =& W.
\end{aligned}$$

After obtaining all terms in Equation (A.1), we need to derive all derivatives in the Equation (A.2). The second derivative $\partial^2\hat{H}/\partial\alpha_{s'}\partial\alpha_s$ is derived as

$$\begin{aligned}
-\frac{\partial^2\hat{H}}{\partial\alpha_{s'}\partial\alpha_s} =& \mathcal{I}\left\{\frac{\partial^2 W}{\partial\alpha_{s'}\partial\alpha_s} + \frac{\partial^2 W}{\partial v\partial\alpha_s}\left(\frac{\partial\hat{v}}{\partial\alpha_{s'}}\right) + \frac{\partial^2 W}{\partial\alpha_{s'}\partial v}\left(\frac{\partial\hat{v}}{\partial\alpha_s}\right)\right. \\
&\left. + \frac{\partial W}{\partial v}\left(\frac{\partial^2\hat{v}}{\partial\alpha_{s'}\partial\alpha_s}\right) + \frac{\partial\hat{v}}{\partial\alpha_{s'}}\frac{\partial^2 W}{\partial v^T\partial v}\frac{\partial\hat{v}}{\partial\alpha_s}\right\}\mathcal{I},
\end{aligned} \tag{A.6}$$

where

$$\frac{\partial^2 W}{\partial \beta_{k'} \partial \beta_k} = \begin{cases} diag[\{(Z_1^2\mu_1)^T, 0_{1\times n}\}^T] & k = k' = 1 \\ diag[\{0_{1\times n}, (Z_2^2\mu_2)^T\}^T] & k = k' = 2 \\ 0_{2n\times 2n} & k \neq k' \end{cases}$$

$$\frac{\partial^2 W}{\partial \tau_{k'} \partial \tau_k} = \begin{cases} diag([\{(\log X)^2\mu_1\}^T, 0_{1\times n}]^T) & k = k' = 1 \\ diag([0_{1\times n}, \{(\log X)^2\mu_2\}^T]^T) & k = k' = 2 \\ 0_{2n\times 2n} & k \neq k' \end{cases}$$

$$\frac{\partial^2 W}{\partial \phi^2} = diag([0_{1\times n}, (\{\log X)v\mu_2\}^T]^T)$$

$$\frac{\partial^2 W}{\partial \tau_{k'} \partial \beta_k} = \begin{cases} diag[\{(Z_1\mu_1 \log X)^T, 0_{1\times n}\}^T] & k = k' = 1 \\ diag[\{0_{1\times n}, (Z_2\mu_2 \log X)^T\}^T] & k = k' = 2 \\ 0_{2n\times 2n} & k \neq k' \end{cases}$$

$$\frac{\partial^2 W}{\partial \phi \partial \beta_k} = \begin{cases} diag[\{0_{1\times n}, (Z_2\mu_2 v)^T\}^T] & k = 2 \\ 0_{2n\times 2n} & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 W}{\partial \phi \partial \tau_k} = \begin{cases} diag[\{0_{1\times n}, (\log X\mu_2 v)^T\}^T] & k = 2 \\ 0_{2n\times 2n} & \text{otherwise} \end{cases}$$

And $\partial^2 v / \partial \alpha_{s'} \partial \alpha_s$ is derived as

$$\frac{\partial^2 v}{\partial \alpha_{s'} \partial \alpha_s} = -(\mathcal{I}^T W \mathcal{I} + m)^{-1} \mathcal{I}^T \left(\frac{\partial \hat{W}}{\partial \alpha_{s'}} \right) \left(\mathcal{A} + \mathcal{I} \frac{\partial \hat{v}}{\partial \alpha_{s'}} \right),$$

where $\mathcal{A} = \begin{cases} Z_k & \text{if } \alpha_s = \beta_k \\ logX & \text{if } \alpha_s = \tau_k \\ v & \text{if } \alpha_s = \phi \end{cases}$. Once all the explicit equations described above are provided, the estimates of α can be obtained iteratively.

The first and second derivatives of the adjusted profile h-likelihood $h_{(\beta,v)}$ with respect to $\theta = \sigma^2$ can be expressed as

$$\frac{\partial h_{(\alpha,v)}}{\partial \theta} = \frac{\partial \hat{h}_{\alpha,v}}{\partial \theta} - \frac{1}{2} \text{tr} \left(\hat{H}_{\alpha,v}^{-1} \frac{\partial \hat{H}_{\alpha,v}}{\partial \theta} \right) \tag{A.7}$$

$$-\frac{\partial^2 h_{(\alpha,v)}}{\partial\theta^2} = -\frac{\partial^2 \hat{h}_{\alpha,v}}{\partial\theta^2} + \frac{1}{2}\text{tr}\left(-\hat{H}_{\alpha,v}^{-1}\frac{\partial \hat{H}_{\alpha,v}}{\partial\theta}\hat{H}_{\alpha,v}^{-1}\frac{\partial \hat{H}_{\alpha,v}}{\partial\theta} + \frac{\partial^2 \hat{H}_{\alpha,v}}{\partial\theta^2}\right) \tag{A.8}$$

where all terms in Equation (A.7) are derived as

$$\begin{aligned}
\frac{\partial \hat{h}_{\alpha,v}}{\partial\theta} &= \frac{\partial h}{\partial\theta}\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} + \frac{\partial h}{\partial v}\frac{\partial \hat{v}}{\partial\theta}\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} \\
&= -\frac{n}{2\sigma^2} + \frac{v^T v}{2(\sigma^2)^2}. \\
\frac{\partial \hat{v}}{\partial\theta} &= -\left(-\frac{\partial^2 h}{\partial v \partial v^T}\right)^{-1}\left(-\frac{\partial^2 h}{\partial v \partial\theta}\right)\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} \\
&= (\mathcal{I}^T W \mathcal{I} + m)^{(-1)}\left(\frac{v}{(\sigma^2)^2}\right)\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} \\
\frac{\partial \hat{H}_{\alpha,v}}{\partial\theta} &= \mathcal{I}\left\{\frac{\partial W}{\partial\theta} + \frac{\partial W}{\partial v}\left(\frac{\partial \hat{v}}{\partial\theta}\right)\right\}\mathcal{I}\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} \\
&= \mathcal{I}\left\{\frac{\partial W}{\partial v}\left(\frac{\partial \hat{v}}{\partial\theta}\right)\right\}\mathcal{I}\bigg|_{\alpha=\hat{\alpha},v=\hat{v}},
\end{aligned}$$

where $\partial W \partial v = W$.

Underived derivatives in Equation (A.8) other than those above in Equation (A.7) are

$$\begin{aligned}
-\frac{\partial^2 \hat{h}_{\alpha,v}}{\partial\theta^2} &= -\frac{\partial^2 h}{\partial\theta^2}\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} - \frac{\partial^2 h}{\partial v\partial\theta}\left(\frac{\partial \hat{v}}{\partial\theta}\right)\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} \\
&= -\frac{n}{2(\sigma^2)^2} + \frac{v^T v}{4(\sigma^2)^3}\bigg|_{v=\hat{v}} - \frac{1}{(\sigma^2)^4}v^T(\mathcal{I}^T W\mathcal{I} + m)^{-1}v\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} \\
\frac{\partial^2 \hat{H}_{\alpha,v}}{\partial\theta^2} &= \mathcal{I}\left\{\frac{\partial^2 W}{\partial\theta\partial v}\left(\frac{\partial \hat{v}}{\partial\theta}\right) + \frac{\partial^2 W}{\partial v^T\partial v}\left(\frac{\partial \hat{v}}{\partial\theta}\right)^2 + \frac{\partial W}{\partial v}\left(\frac{\partial^2 \hat{v}}{\partial\theta^2}\right)\right\}\mathcal{I}\bigg|_{\alpha=\hat{\alpha},v=\hat{v}} \\
&= \mathcal{I}\left\{\frac{\partial^2 W}{\partial v^T\partial v}\left(\frac{\partial \hat{v}}{\partial\theta}\right)^2 + \frac{\partial W}{\partial v}\left(\frac{\partial^2 \hat{v}}{\partial\theta^2}\right)\right\}\mathcal{I}\bigg|_{\alpha=\hat{\alpha},v=\hat{v}},
\end{aligned}$$

where

$$\frac{\partial^2 \hat{v}}{\partial\theta^2} = \quad -(\mathcal{I}^T W\mathcal{I} + m)^{-1}\left\{\frac{2v}{(\sigma^2)^3} - \frac{2}{(\sigma^2)^4}(\mathcal{I}^T W\mathcal{I} + m)^{-1}v\right\}\bigg|_{\alpha=\hat{\alpha},v=\hat{v}}$$

APPENDIX C

BIAS REDUCTION METHOD 1 OF H-LIKELIHOOD APPROACH FOR THE CORRECTED RANDOM EFFECTS MODEL

For the correlated random effects structure, the mth iteration of the bias reduction procedure can follow the following steps:

1. Estimate $\hat{\alpha}_k^{(m)}$ and $\hat{v}^{(m)}$ from equations (3.9) and (4.3), and update the hessian matrix from $\hat{H}^{(m-1)}$ and $\hat{D}^{(m-1)}$ to $\hat{H}^{(m)}$ and $\hat{D}^{(m)}$.
2. Calculate the covariance matrix of $\hat{v_i}^{(m)}$ from
$$\hat{\Xi}_i^{(m)} = \begin{pmatrix} (\hat{H}^{(m)})^{-1}(2i-1,2i-1) & (\hat{H}^{(m)})^{-1}(2i-1,2i) \\ (\hat{H}^{(m)})^{-1}(2i,2i-1) & (\hat{H}^{(m)})^{-1}(2i,2i) \end{pmatrix},$$
3. Calculate $E[v_i|\hat{v}_i^{(t)}] = \Sigma(\Sigma+\Xi_i)^{-1}\hat{v}_i$, and $E[e^{v_i}|\hat{v}_i^{(t)}] = \exp\{\Sigma(\Sigma+\Xi_i)^{-1}\hat{v}_i + \frac{1}{2}\text{diag}(\Sigma - \Sigma(\Sigma+\Xi_i)^{-1}\Sigma)\}$, where $\hat{(\Sigma)}^{(m-1)}$ is the variance estimate of v_i at the $(m-1)^{th}$ iteration.
4. Estimate $\hat{\Sigma}^{(m)}$ from equation (3.11) in Step 4 of the h-likelihood approach.
5. Update h_v and $h_{\alpha,v}$ in the step 2 and step 3 of h-likelihood method procedure, by substituting v_i by $E[v_i|\hat{v}_i^{(t)}]$ and $e^{(v_i)}$ by $E[e^{v_i}|\hat{v}_i^{(t)}]$.
6. Iterate Steps 1 to 5 until converge.

BIBLIOGRAPHY

M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1972.

O. E. Barndorff-Nielsen. Conditionality resolutions. *Biometrika*, 67:293 – 310, 1980.

O. E. Barndorff-Nielsen. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70:343 – 365, 1983.

F. Bartolucci, R. Bellio, A. Salvan, and N. Sartori. Modified profile likelihood for panel data models. page Available at SSRN: http://ssrn.com/abstract=2000666 or http//dx.doi.org/10.2139/ssrn.2000666, 2012.

N. E. Breslow and X. Lin. Bias correction in generalized linear models with a single component of dispersion. *Biometrika*, 82:81–92, 1995.

C. Chang, H. Yang, G. Tang, and M. Ganguli. Minimizing attrition bias: a longitudinal study of depressive symptoms in an elderly cohort. *International Psychogeriatrics*, 21(5): 869–878, 2009.

C. Chang, Y. Zhao, C. Lee, and M. Ganguli. Smoking, death, and alzheimer disease: a case of competing risks. *Alzheimer Disease and Associated Disorders*, 26(4):300–306, 2012.

D. R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187 – 220, 1972.

D. R. Cox and N. Reid. Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society*, 49:1–39, 1987.

H. Dodge, C. Chang, I. Kamboh, and M. Ganguli. Risk of alzheimer's disease incidence attributable to vascular disease in the population. *Alzheimer's and Dementia*, 7(3):356–360, 2011.

G. Escarela and J. F. Carriere. Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research*, 12:333–349, 2003.

J. P. Fine and R. J. Gray. A proportional hazards model for the sub distribution of a competing risk. *Journal of the American Statistical Association*, 94:496–509, 1999.

M. Ganguli, J. Bilt, J. Saxton, C. Shen, and H. Dodge. Alcohol consumption and cognitive function in late life: A longitudinal community study. *Neurology*, 65:1210–1217, 2005.

M. Ganguli, Y. Du, H. Dodge, G. Ratcliff, and C. Chang. Depressive symptoms and cognitive decline in late life: a prospective epidemiological study. *Archives of General Psychiatry*, 63:153–160, 2006.

M. Ganguli, B. Fu, T. Hughes, and C. Chang. Mild cognitive impairment: incidence and vascular risk factors in a population-based cohort. *Neurology*, [PMC in process], 2013.

R. Gray. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16:1141–1154, 1988.

X. Guo and B. P. Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58:1–9, 2004.

I. D. Ha and Y. Lee. Estimating frailty models via poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*, 12:663–681, 2003.

I. D. Ha, Y. Lee, and J. K. Song. Hierarchical likelihood approach for frailty models. *Biometrika*, 88:233–243, 2001.

I. D. Ha, T. Park, and Y. Lee. Joint modeling of repeated measures and survival time data. *Biometrical*, 45:647–658, 2003.

I. D. Ha, Y. Lee, and G. MacKenzie. Model selection for multi-component frailty models. *Statistics in Medicine*, 26:4790–4807, 2007.

I. D. Ha, M. Noh, and Y. Lee. Bias reduction of likelihood estimators in semi parametric frailty models. *Scandinavian Journal of Statistics*, 37:307–320, 2010.

I. D. Ha, R. Sylvester, C. Legrand, and G. MacKenzie. Frailty modeling for survival data from multi-center clinical trials. *Statistical in Medicine*, 30:2144–2159, 2011.

I. D. Ha, F. Vaida, and Y. Lee. Interval estimation of random effects in proportional hazards models with frailties. *Statistical Methods in Medical Research*, 0:1–18, 2013.

R. Henderson and P. Diggle A. Dobson. Joint modeling of longitudinal measurements and event time data. *Biostatistics*, 1:465–480, 2000.

G. Iversen. *Bayesian Statistical Inference*. Sage Publications, Thousand Oaks, CA, 1984.

J. Jeon and L. Hsu. Bias correction in the hierarchical likelihood approach to the analysis of multivariate survival data. *Biostatistics*, 0:1–14, 2011.

J. Jeong and J. Fine. Parametric regression on cumulative incidence function. *Biostatistics*, 8:184–196, 2007.

J. Klein and P. Andersen. Regression modeling of competing risks data based on pseudo values of the cumulative incidence function. *Biometrics*, 61:223–229, 2005.

Y. Lee and J. A. Nelder. hierarchical generalized linear model. *Journal of the Royal Statistical Society*, 58:619–678, 1996.

Youngjo Lee. Estimating intraclass correlation for binary data using extended quasi-likelihood. *Statistical Modeling*, 4:113–126, 2004.

H. Li and P. Lahiri. An adjusted maximum likelihood method for solving small area estimation problems. *Journal of Multivariate Analysis*, 101:882–892, 2010.

R. Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88:125–134, 1993.

L. Liu and X. Huang. The use of gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine*, 27:2665–2683, 2008.

C. Morris. Mixed model prediction and small area estimation. *Test*, 15:72–76, 2006.

G. G. Nielsen, R. D. Gill, and P. K. Andersen. A counting process approach to maximum likelihood estimation in frailty models. *Scand J Statist*, 19:25–43, 1992.

A. V. Peterson. Jr. bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. *Proceedings of the National Academy of Scienes*, 73:11 – 13, 1976.

J. Pinheiro and D. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of the Computational and Graphical Statistics*, 4:12–35, 1995.

Ross L. Prentice and John D. Kalbfleisch. *The Statistical Analysis of Failure Time Data*. Wiley, New York, 1980.

S. Ripatti and J. Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56:1016–1022, 2000.

T. Scheike, M. Zhang, and T. Gerds. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95:205–220, 2008.

T. Severini. An approximation to the modified profile likelihood function. *Biometrika*, 85: 403–411, 1998.

A. Sklar. Fonctions de rpartition n dimensions et leurs marges. *Publ. Inst. Statist. Univ*, 8: 229 – 231, 1959.

D. Spiegelhalter, N. Best, B. Carlin, and A. Van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, 64:583–640, 2002.

K. Stowell, C. Chang, J. Bilt, G. Stoehr, and M. Ganguli. Sustained benzodiazepine use in a community sample of older adults. *Journal of the American Geriatrics Society*, 56(12): 2285–2291, 2008.

M. S. Wulfsohn and A. A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:330–339, 1997.