Journal of Statistical Research *ISSN 0256 - 422 X*
2019, Vol. 53, No. 1, pp. 1-25

JOINT MODELING OF BIVARIATE LONGITUDINAL AND SURVIVAL DATA IN SPOUSE PAIRS

Jia-Yuh Chen

The Emmes Corporation, 401 North Washington Street, Suite 700, Rockville, MD 20850
Email: lchen@emmes.com

Richard Schulz

Department of Psychiatry, School of Medicine, University of Pittsburgh
3811 O'Hara Street, Pittsburgh, PA 15213
Email: schulz@pitt.edu

and

Stewart J. Anderson*

Department of Biostatistics, University of Pittsburgh Graduate School of Public Health
7119 Public Health, 130 Desoto St, Pittsburgh, PA 15261
Email: sja@pitt.edu

SUMMARY

We investigated the association between longitudinally measured depression scores and survival times simultaneously for paired spouse data from the Cardiovascular Health Study (CHS). We propose a joint model incorporating within-pair correlations, both in the longitudinal and survival processes. We use bivariate linear mixed-effects models for the longitudinal processes, where the random effects are used to model the temporal correlation within each subject and the correlation across outcomes between subjects. For the survival processes, we incorporate gamma frailties into Weibull proportional hazards models to account for the correlation between survival times within pairs. The two sub-models are then linked through shared random effects, where the longitudinal and survival processes are conditionally independent given the random effects. Parameter estimates are obtained via the EM algorithm by maximizing the joint likelihood for the bivariate longitudinal and bivariate survival data. We use our method to model data where the use of bivariate longitudinal and survival sub–models are apropos but where there are no competing risks, that is, the censoring of one spouse's time–to–mortality is not necessarily guaranteed by the death of the other spouse.

Keywords and phrases: Bivariate linear mixed-effects model; Bivariate survival data; Gamma frailty; Joint models; Spouse pairs; Weibull proportional hazards model.

* Corresponding author

1 Introduction

Multivariate longitudinal models arising when two or more related processes are being measured repeatedly were proposed as early as Potthoff and Roy (1964) who used a general multivariate linear model to characterize both univariate and multivariate longitudinal data. Reinsel (1982, 1984) introduced a multivariate linear random-effects model but that particular model could only be used to analyze complete and balanced multivariate longitudinal data in which all outcomes are measured at the same time point. In practice, however, the data can be highly unbalanced, where outcomes may be measured at different time points. Reinsel's work was extended by Shah et al. (1997) to accommodate the case of arbitrary measurement times. Their approach employed the EM algorithm for the parameter estimates. Schafer (1997) and Schafer and Yucel (2002) developed a similar model which allowed for multiple imputation in cases when there was missing data. Other similar work was introduced by Morrell et al. (2003) used the multivariate linear mixed-effects model in a Bayesian framework to predict hypertension based on body mass index (BMI), systolic blood pressure and triglyceride levels from the Baltimore longitudinal study of aging. Lin et al. (2002), Thiébaut et al. (2005), and Chi and Ibrahim (2006) also employed this approach. Dang et al. (2005) used a Kalman filtering approach to model bivariate unequally spaced longitudinal data. Nonlinear methods were also employed using both splines and parametric methods (Ruppert et al. (2003), Song et al. (2002)). In addition, Brown et al. (2005) developed a data-driven Bayesian approach B-spline model to describe how two biomarkers (CD4 counts and HIV RNA levels) change over time and to estimate the impact of a set of covariates on the two biomarkers in an AIDS clinical trial. Rizopoulos and Ghosh (2011) considered natural cubic splines.

Modeling bivariate survival data has also had a long history in the statistical literature. After Gumbel (1960) first proposed two bivariate distributions whose marginal distributions are exponential, others (Freund (1961), Marshall and Olkin (1967)) considered multivariate reliability models for two or more component life testing systems that relaxed the requirement for exponential failure times. Others (Klein et al. (1989); Ghosh and Gelfand (1998)) generalized the exponential models using more flexible Weibull distributions and where proportional hazards or accelerated failure time models are allowed. Others (Vaupel et al. (1979); Lancaster (1979)) introduced the notion of frailty, a random effect, to represent the unobserved population heterogeneity. In related work, Clayton (1978) proposed a continuous bivariate survival model where the conditional hazard for subject 1 at time t_1 given that another subject died at time t_2 and the conditional hazard for subject 1 at time t_1 given that the other subject survived at least to t_2 are proportional, that is,

$$\frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 \geq t_2)} = 1 + \Phi, \tag{1.1}$$

where the hazard ratio $(1 + \Phi)$ is constant over time. This model can be interpreted in terms of a proportional hazards model with a one-parameter gamma distributed frailty, μ, and $\Phi = \theta$. Oakes (1982) considered the bivariate case without covariates and showed that

θ (or Φ) is closely related to a measure of dependence, Kendall's τ, where $\tau = \theta/(\theta + 2)$ is calculated from an uncensored sample or from a right-censored sample using only those pairs that can be classified as either concordant or discordant. Clayton and Cuzick (1985) later extended the model to allow for covariates and adapted to study the problem of intra-class association (e.g., litter-matched and matched-pair failure-time data). Guo and Rodriguez (1992) generalized Clayton's model to the multivariate case and fitted the model using an accelerated EM algorithm. Wienke et al. (2003) suggested a correlated gamma frailty by extending Clayon's shared gamma frailty model to explain the correlation within clusters in a breast cancer incidence data for Swedish female monozygotic and dizygotic twin pairs. Hanagal also has published on the gamma shared frailty model (Hanagal, 2006, 2007; Hanagal and Dabade, 2013; Hanagal and Pandey, 2014).

Joint modeling, linking longitudinal data with survival data, has become a valuable tool for analyzing clinical trials data. The motivating idea behind this approach is to couple a survival model, which is of primary interest, with a suitable model for the repeated measurements of an endogenous outcome that will account for its special features. Of particular interest is how the temporal features of the endogenous outcome affect the survival outcome.

A well–known example where this methodology is used is in HIV clinical trials where longitudinally measured immunologic and physiological status such as CD4 count and RNA copy number are considered as the predictors for time to progression to acquired immunodeficiency syndrome (AIDS) or death (Thiébaut et al., 2005). Most joint models developed so far have focused on relating single or multiple observations measured longitudinally to a time-to-event endpoint (Choi et al., 2014; Henderson et al., 2000; Lin et al., 2002; Rizopoulos and Ghosh, 2011; Song et al., 2002). Newer work focuses on how longitudinal measurements relate to multiple time-to-event endpoints. For example, Chi and Ibrahim (2006) developed a model that related multiple quality of life (QOL) endpoints to disease-free survival (DFS) and overall survival (OS). Their model conditioned the multiple QOL longitudinal components on a single latent QOL process and assumed that longitudinal covariates were related to the QOL components only through the latent process. They also assume that the DFS and OS processes are *independent given a common frailty*. Furthermore, their broader model development induced a proportional hazards structure for the population hazard, both conditionally and marginally, and their model is capable of dealing with survival functions with different cure rate structures. Other related work focused on how longitudinal measurements relate to multiple competing risks time-to-event endpoints (Elashoff et al., 2007), or recurrent events (Liu et al., 2008).

Due to the nature of our application, we formulate a different model than those formulated previously. Our subjects are paired and hence, our "experimental unit" is a pair of different individuals that are likely to be correlated in both the longitudinal and time-to-event processes but do not involve competing risk events. We assume a bivariate Weibull structure for the time-to-event processes. This situation is fundamentally different than that involving a single individual with multiple longitudinal and time-to-event processes

especially when the time-to-event processes in each individual involve competing risks. A prototypical example of our situation involves measurements on twins, who have the same genes and also share a similar childhood environment and have correlated but often different trajectories of health measurements and also typically, have different survival times. Another example is married couples. The individuals in a married couple presumably do not have common genetic traits like twins but they usually have other common traits. For example, a non-smoker might prefer a non-smoker, leading to smoking concordance within pairs. Shared traits or risks may also be due to coexistent life styles; for example, even though a non-smoker chooses a smoker, they will both have elevated mortality risks; one as an active smoker, one as passive smoker. Married couples usually have similar diets and living environments. Moreover, effects observed in *older* married couples tend to be more strongly associated because of shared *lifetime* exposures. In either example of twins or spouse pairs, there is no inherent censoring of observing a death in one member of a pair due to the death of the other member of the pair. Accordingly, somewhat more straight forward models can account for the correlation within pairs when jointly modeling such characteristics in a population of pairs of different subjects.

Our proposed model was motivated by the Cardiovascular Health Study (CHS) (Fried et al., 1991). The CHS is a prospective, observational study designated to identify the risk factors for and consequences of cardiovascular disease in older adults. A total of 5888 men and women aged 65 or older were enrolled from four U.S. communities and underwent annual clinical examinations and completed an extensive array of demographic and health assessments. The CHS sample included 1330 married couples; we are specifically interested in this subsample. Depression is the most prevalent mental health problem in adulthood and a significant public health concern (Fisher et al., 1993). Epidemiological studies have found that 10-20% of community-dwelling elderly persons report clinically significant depressive symptomatology (Blazer et al., 1987; Kennedy et al., 1989; Murrell et al., 1983). However, findings of relationships between depression and mortality in older population have been inconsistent across studies with some investigators concluding that depression is associated with an increased risk of mortality and others failing to find this association (Schulz et al., 2000, 2002; Wulsin et al., 1999). Limitations due to lack of adequate information or less than optimal methodology may have contributed to inconsistent findings about the relationship between depression and mortality (Zhang et al., 2009). First, many studies only have a one-time assessment of depression. In such studies, the dynamic nature of depression cannot be captured. Second, some studies had relatively short follow-up periods to ascertain mortality, which only captured a few deaths and hence, produced potentially biased and underpowered results. In the present study, we examine the association between depression and mortality in a large community sample that utilizes longitudinal depression measures over a long period of time while controlling for potential confounding factors. The traditional time-dependent Cox model is only appropriate for exogenous time-dependent covariates and thus cannot easily handle longitudinal depression measures that are taken on the subjects and thus typically require the survival of the subject for their existence. In order to better

quantify the association of mortality and longitudinal course of depression, it is necessary to use a modeling approach to characterize both longitudinal and survival processes jointly.

We propose a joint modeling methodology for paired data where we use a bivariate Weibull model for the time-to-event processes. Our model incorporates the flexibility associated with Weibull models for the joint time-to-event hazards, while still allowing relatively easy computation via a maximum likelihood approach. We also do not assume an underlying cure rate structure. Our model, like the current joint models, accounts for within-pair correlation, both in the longitudinal and in the time-to-event processes. The longitudinal processes can have both serial and cross correlations. Specifically, we propose a joint model to investigate the association between time to mortality and longitudinal depression scores among married couples adjusted for the covariates related to mortality and longitudinal depression separately.

The rest of our paper is structured as follows. In sections **2** and **3**, we introduce our model and discuss how the parameters are estimated using likelihood techniques and the EM algorithm. Simulations are performed in section **4** to assess how well model parameters are estimated under different correlation structures and in section **5**, we give the results and interpretation of an analysis of the CHS spouse mortality data using our method. Finally, in section **6**, we make a few conclusions and discuss areas of further research.

2 The Model

2.1 The Bivariate Longitudinal Submodel

We propose a bivariate linear mixed-effects model to explicitly model the two sources of correlation: the correlation over time for each response and that between the two responses. Let $y_{ik}(t)$ be the response of subject k in pair i at time t ($i = 1, \ldots, n$; $k = 1, 2$). Each subject's response is described by

$$\begin{aligned} y_{ik}(t) &= \mathbf{x}_{ik}(t)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(t)\boldsymbol{b}_{ik} + \varepsilon_{ik}(t) \\ &= m_{ik}(t) + \varepsilon_{ik}(t),\ \boldsymbol{b}_i \sim N(\mathbf{0}, \mathbf{D}),\ \boldsymbol{\varepsilon}_i(t) \sim N(\mathbf{0}, \mathbf{R}) \ , \end{aligned}$$

where $\mathbf{x}_{ik}(t)$ and $\mathbf{z}_{ik}(t)$ are the $1 \times p$ and $1 \times r$ ($0 < r \leq p$) design matrices of fixed and random effects, respectively. The $p \times 1$ and $r \times 1$ vectors of corresponding fixed and random effects parameters are $\boldsymbol{\beta}_k$ and $\boldsymbol{b}_{ik}$; $m_{ik}(t)$ denotes the *true, unobserved* value of longitudinal response; and $\varepsilon_{ik}(t)$ is the measurement error. For our particular application, the subject–specific random effects are of the form $\mathbf{z}_{ik}(t)\boldsymbol{b}_{ik} = b_{0ik} + b_{1ik}t$, that is, $r = 2$, and follow a joint distribution given by

$$\begin{bmatrix} \boldsymbol{b}_{i1} \\ \boldsymbol{b}_{i2} \end{bmatrix} = \begin{bmatrix} b_{0i1} \\ b_{1i1} \\ b_{0i2} \\ b_{1i2} \end{bmatrix} \sim N(\mathbf{0}, \mathbf{D}) \ ,$$

where $\mathbf{D}$, the covariance (symmetric) matrix of the random effects, has the following structure to reflect the bivariate nature of the data:

$$\mathbf{D} = \begin{bmatrix} \sigma^2_{b_{01}} & \sigma_{b_{01}b_{11}} & \sigma_{b_{01}b_{02}} & \sigma_{b_{01}b_{12}} \\ & \sigma^2_{b_{11}} & \sigma_{b_{11}b_{02}} & \sigma_{b_{11}b_{12}} \\ & & \sigma^2_{b_{02}} & \sigma_{b_{02}b_{12}} \\ & & & \sigma^2_{b_{12}} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_2 \end{bmatrix} .$$

$\mathbf{D}$ can be partitioned in four sub-matrices: (1) $\mathbf{D}_1 = \begin{bmatrix} \sigma^2_{b_{01}} & \sigma_{b_{01}b_{11}} \\ & \sigma^2_{b_{11}} \end{bmatrix}$, the variances and covariances of random effects for the response of subject 1; (2) $\mathbf{D}_2 = \begin{bmatrix} \sigma^2_{b_{02}} & \sigma_{b_{02}b_{12}} \\ & \sigma^2_{b_{12}} \end{bmatrix}$, the variances and covariances of random effects for the response of subject 2; (3) $\mathbf{D}_{12} = \mathbf{D}_{21} = \begin{bmatrix} \sigma_{b_{01}b_{02}} & \sigma_{b_{01}b_{12}} \\ \sigma_{b_{11}b_{02}} & \sigma_{b_{11}b_{12}} \end{bmatrix}$, the covariances between the random effects of the different responses. If $\mathbf{D}_{12} = \mathbf{D}_{21} = \mathbf{0}$, the responses are independent at a given time. The two measurement errors are assumed to follow a joint distribution given by

$$\begin{bmatrix} \varepsilon_{i1}(t) \\ \varepsilon_{i2}(t) \end{bmatrix} \sim N(\mathbf{0}, \mathbf{R}) ,$$

where $\mathbf{R}$, the covariance matrix of the measurement errors. In our case, we assume $\mathbf{R}$ to be a diagonal matrix with the form $\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$, and σ_1^2 and σ_2^2 represent the variance of measurement errors of each response. We also assume the measurement errors are independent of the random effects, which implies that conditional on the random effects, both response trajectories are independent.

2.2 The Bivariate Survival Submodel

For the survival sub–model, we propose a bivariate Weibull model with a gamma frailty to jointly characterize the spouses' times to death. We assume that the times until mortality are conditionally independent given the pair-specific random effect (the frailty), μ_i. The shape parameter of the Weibull baseline hazard is assumed to be the same for the two spouse members. The conditional hazard function at time t for subject k ($k = 1, 2$) in pair i ($i = 1, \ldots, n$) is as follows:

$$\begin{aligned} h_{ik}(t|\mu_i) &= \rho\lambda_k t^{\rho-1}\mu_i \exp\{\mathbf{w}^*_{ik}\boldsymbol{\gamma}^*_k + \alpha_k m_{ik}(t)\}, \rho > 0, \lambda_k > 0 \\ &= \rho t^{\rho-1}\mu_i \exp\{\log\lambda_k + \mathbf{w}^*_{ik}\boldsymbol{\gamma}^*_k + \alpha_k m_{ik}(t)\} \\ &= \rho t^{\rho-1}\mu_i \exp\{\mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(t)\} , \\ \mu_i &\sim GAM(1/\theta, \theta) , \end{aligned} \tag{2.1}$$

where ρ is the Weibull shape parameter; λ_k is the Weibull scale parameter for subject k; μ_i is the frailty for pair i; $\mathbf{w}^*_{ik}$ is the baseline covariate vector associated with survival time and

$\boldsymbol{\gamma}_k^*$ is the corresponding effect for subject k; α_k represents the effect of the true longitudinal response, $m_{ik}(t)$, on the survival process for subject k; the intercept term in the baseline covariate vector ($\mathbf{w}_{ik}$) corresponds to $\log \lambda_k$. Equation (2.1) formulates the variability of the event times, coming from two sources. The first source is *natural* variability that is explained by the hazard function and the second is variability common to individuals in the same pair that is explained by the frailty, μ_i. We assume that the frailty, μ_i, has a one-parameter gamma distribution with mean 1 and variance θ and acts multiplicatively on the hazard. The density of μ_i is

$$g(\mu_i) = \frac{\mu_i^{1/\theta - 1} \exp\left(-\mu_i/\theta\right)}{\Gamma(1/\theta)\theta^{1/\theta}}.$$

Larger values of θ reflect greater heterogeneity between pairs and stronger association among individuals within a pair. The dependence of event time between the paired individuals can be measured by Kendall's τ using $\tau = \theta/(\theta + 2)$ (Oakes, 1982). When $\theta = 0$, both event times are assumed to be independent. Let (t_{i1}, t_{i2}) be a bivariate failure time in pair i. The joint survival function of (t_{i1}, t_{i2}) can be obtained by integrating out μ_i from a conditional survival function

$$\begin{aligned} S(t_{i1}, t_{i2}) &= \int_0^\infty S(t_{i1}, t_{i2}|\mu_i) g(\mu_i) \mathrm{d}\mu_i \\ &= \int_0^\infty \exp\left[-\sum_{k=1}^{2} \int_0^t \rho s^{\rho-1} \mu_i \exp\left\{\mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(s)\right\} \mathrm{d}s \right] g(\mu_i) \mathrm{d}\mu_i \ . \end{aligned}$$

2.3 Joint Likelihood

Let the observed data for each pair be $\left\{ \boldsymbol{T}_i = \begin{bmatrix} T_{i1} \\ T_{i2} \end{bmatrix},\ \boldsymbol{\Delta}_i = \begin{bmatrix} \Delta_{i1} \\ \Delta_{i2} \end{bmatrix}, \boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{y}_{i1} \\ \boldsymbol{y}_{i2} \end{bmatrix} \right\}$ $(i = 1, \ldots, n)$, where $\boldsymbol{y}_i$ is the longitudinal response vector for pair i, $\boldsymbol{T}_i$ is the event time vector, and $\boldsymbol{\Delta}_i$ is the vector of event indicators ($\Delta_{ik} = 1$ if the event occurs and $\Delta_{ik} = 0$, otherwise). Assume censoring is independent of the frailty, μ_i. We assume that given the shared random effects $\boldsymbol{b}_i$, the bivariate longitudinal response and the bivariate event time are independent. This means that these random effects account for both the association between bivariate longitudinal and bivariate survival outcomes, and the correlations between the repeated measurements in each response and between the two responses in the bivariate longitudinal process. Under these assumptions, we have that

$$p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i|\boldsymbol{b}_i; \boldsymbol{\phi}) = p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i|\boldsymbol{b}_i, \boldsymbol{\phi}) p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\phi}), \quad \text{and}$$

$$p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\phi}) = \prod_{k=1}^{2} \prod_{j=1}^{n_{ik}} p\{y_{ik}(t_{ikj})|\boldsymbol{b}_i, \boldsymbol{\phi}\} \ ,$$

where $p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i|\boldsymbol{b}_i, \boldsymbol{\phi})$ and $p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\phi})$ are the joint densities of bivariate survival and longitudinal processes, respectively; $\boldsymbol{\phi} = (\boldsymbol{\phi}_y^T, \boldsymbol{\phi}_t^T, \boldsymbol{\phi}_b^T)^T$ denotes the full parameter vector, with $\boldsymbol{\phi}_y$ denoting the parameters for the bivariate longitudinal process, $\boldsymbol{\phi}_t$ the parameters for the

bivariate survival process, and $\boldsymbol{\phi}_b$ the parameters for the random effects covariance matrix; n_{ik} is the number of repeated measurements of subject k in pair i.

We do not observe the random effects, $\boldsymbol{b}_i$. Hence, the log-likelihood of observed bivariate longitudinal response and bivariate event time can be formulated as

$$
\begin{aligned}
\ell(\boldsymbol{\phi}) &= \sum_{i=1}^{n} \log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) = \sum_{i=1}^{n} \log \int p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \log \int p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) p(\boldsymbol{b}_i | \boldsymbol{\phi}_b) \mathrm{d}\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \int \big\{ \log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) + \log p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) + \log p(\boldsymbol{b}_i | \boldsymbol{\phi}_b) \big\} \mathrm{d}\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \int \Bigg(D_i \log \theta + \log \frac{\Gamma(D_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} + \sum_{k=1}^{2} \Delta_{ik} \big\{ \log (\rho T_{ik}^{\rho-1}) + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(T_{ik}) \big\} \\
&\quad - (D_i + \frac{1}{\theta}) \log \Big[1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\{\alpha_k m_{ik}(s)\} \mathrm{d}s \Big] \\
&\quad + \sum_{k=1}^{2} \Big\{ -\frac{n_{ik}}{2} \big(\log 2\pi + \log |\mathbf{R}| \big) \Big\} - \frac{\big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\boldsymbol{b}_i\big)^T \mathbf{R}^{-1} \big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\boldsymbol{b}_i\big)}{2} \\
&\quad - \frac{q_b}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{D}| - \frac{\boldsymbol{b}_i^T \mathbf{D}^{-1} \boldsymbol{b}_i}{2} \Bigg) \mathrm{d}\boldsymbol{b}_i \ ,
\end{aligned}
$$

where $D_i = \sum_{k=1}^{2} \Delta_{ik}$; q_b denotes the dimensionality of the random-effects vector, and other quantities are as defined earlier.

3 Parameter Estimation using the EM Algorithm

In the joint modeling literature, the EM algorithm has been traditionally preferred, mainly due to the fact that in the M-step, some of the parameters have closed-form updates. In this study, the EM algorithm was used to estimate the parameters, $\boldsymbol{\phi} = (\boldsymbol{\phi}_y^T, \boldsymbol{\phi}_t^T, \boldsymbol{\phi}_b^T)^T$, by maximizing the joint likelihood of the observed data. Below we describe how we implemented our estimation procedure using the EM algorithm.

E-step: The aim of using the EM algorithm is to find the parameter values $\boldsymbol{\phi}$ that maximize the observed data log-likelihood $\ell(\boldsymbol{\phi})$, but by maximizing instead the expected value of the complete data log-likelihood with respect to the posterior distribution of random effects

as below.

$$
\begin{aligned}
\varrho(\boldsymbol{\phi} \mid \boldsymbol{\phi}^{(it)}) &= \sum_{i=1}^{n} \int \log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\phi}) p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \int \big\{ \log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) + \log p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) + \log p(\boldsymbol{b}_i | \boldsymbol{\phi}_b) \big\} p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \int \Bigg(D_i \log \theta + \log \frac{\Gamma(D_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} + \sum_{k=1}^{2} \Delta_{ik} \big[\log (\rho T_{ik}^{\rho-1}) + \mathbf{w}_{ik} \boldsymbol{\gamma}_k + \alpha_k \{ \mathbf{x}_{ik}(T_{ik}) \boldsymbol{\beta}_k + \mathbf{z}_{ik}(T_{ik}) \boldsymbol{b}_{ik} \} \big] \\
&\quad - (D_i + \frac{1}{\theta}) \log \Bigg(1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik} \boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp \big[\alpha_k \{ \mathbf{x}_{ik}(s) \boldsymbol{\beta}_k + \mathbf{z}_{ik}(s) \boldsymbol{b}_{ik} \} \big] \mathrm{d}s \Bigg) \\
&\quad + \sum_{k=1}^{2} \Bigg\{ - \frac{n_{ik}}{2} \big(\log 2\pi + \log |\mathbf{R}| \big) \Bigg\} - \frac{\big(\boldsymbol{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k - \mathbf{Z}_i \boldsymbol{b}_i \big)^T \mathbf{R}^{-1} \big(\boldsymbol{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k - \mathbf{Z}_i \boldsymbol{b}_i \big)}{2} \\
&\quad - \frac{q_b}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{D}| - \frac{\boldsymbol{b}_i^T \mathbf{D}^{-1} \boldsymbol{b}_i}{2} \Bigg) p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i \ .
\end{aligned}
\tag{3.1}
$$

The integrals of the survival density and random effects cannot be solved in closed form so numerical approaches are required to approximate these integrals. We employ the Gauss–Kronrod quadrature rule (Press et al., 2007) to approximate the one-dimensional integral with respect to time in the survival function. For the random effects, we extend the pseudo-adaptive Gauss–Hermite rule proposed by Rizopoulos (2012) to approximate the integral with respect to random effects. Rizopoulos' approach approximates the integral of subject-specific random effects across individuals. However, in this study, the random effects of each individual within a pair are tied together through a joint distribution. Thus, we have extended this approach to approximate the integral of subject-specific random effects across individuals and pairs. The details are given in Section A of the Supplementary Material.

Using these approaches, the log-likelihood (3.1) is approximated by

$$\begin{aligned}\varrho(\boldsymbol{\phi} \mid \boldsymbol{\phi}^{(it)}) \approx \sum_{i=1}^{n} \Bigg\{ & 2^{q_b/2} |\widetilde{\boldsymbol{B}}_i|^{-1} \sum_{t_1 \cdots t_{q_b}} \Bigg(D_i \log\theta + \log \frac{\Gamma(D_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} \\ & + \sum_{k=1}^{2} \Delta_{ik} \big[\log(\rho T_{ik}^{\rho-1}) + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k \{ \mathbf{x}_{ik}(T_{ik})\boldsymbol{\beta}_k + \mathbf{z}_{ik}(T_{ik})\widetilde{\boldsymbol{r}}_t \} \big] \\ & - (D_i + \frac{1}{\theta}) \log \Bigg[1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \frac{T_{ik}}{2} \Big(\sum_{g=1}^{m} \pi_g \rho\, t_g^{\rho-1} \exp\big[\alpha_k \{ \mathbf{x}_{ik}(t_g)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(t_g)\widetilde{\boldsymbol{r}}_t \} \big] \Big) \Bigg] \\ & + \sum_{k=1}^{2} \Big\{ - \frac{n_{ik}}{2} \big(\log 2\pi + \log|\mathbf{R}| \big) \Big\} - \frac{\big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\widetilde{\boldsymbol{r}}_t\big)^T \mathbf{R}^{-1} \big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\widetilde{\boldsymbol{r}}_t\big)}{2} \\ & - \frac{q_b}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{D}| - \frac{\widetilde{\boldsymbol{r}}_t^T \mathbf{D}^{-1} \widetilde{\boldsymbol{r}}_t}{2} \Bigg) p(\widetilde{\boldsymbol{r}}_t | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \pi_t \exp\left(\|\boldsymbol{b}_t\|^2\right) \Bigg\} ,\end{aligned}$$

where m is the number of Gauss–Kronrod quadrature points; $t_g = \frac{T_{ik}}{2} t_g^* + \frac{T_{ik}}{2}$ with Gauss-Kronrod quadrature points t_g^* and weights π_g; $\sum_{t_1 \cdots t_{q_b}}$ is shorthand for $\sum_{t_1=1}^{K} \cdots \sum_{t_{q_b}=1}^{K}$ with K denoting the number of Gauss-Hermite quadrature points, and $\boldsymbol{b}_t^T = (b_{t_1}, \ldots, b_{t_{q_b}})$ are the Gauss-Hermite quadrature points with corresponding weights $\boldsymbol{\pi}_t$; $\widetilde{\boldsymbol{r}}_t = \widetilde{\boldsymbol{b}}_i + \sqrt{2}\widetilde{\boldsymbol{B}}_i^{-1}\boldsymbol{b}_t$ with $\widetilde{\boldsymbol{b}}_i = \arg\max_b \{\log p(\boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\phi}_y)\}$ and $\widetilde{\boldsymbol{B}}_i$ denoting the Choleski factor of $\widetilde{\boldsymbol{H}}_i$ with $\widetilde{\boldsymbol{H}}_i = -\partial^2 \log p(\boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\phi}_y)/\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^T|_{\boldsymbol{b}_i = \widetilde{\boldsymbol{b}}_i}$.

M-step: In the M-step, we update the parameters by

$$\boldsymbol{\phi}^{(it+1)} = \arg\max_{\boldsymbol{\phi}} \varrho(\boldsymbol{\phi} \mid \boldsymbol{\phi}^{(it)}) \ .$$

Because the complete data log-likelihood consists of three parts, i.e., $\log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\phi}) = \log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) + \log p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) + \log p(\boldsymbol{b}_i | \boldsymbol{\phi}_b)$, maximization of $\varrho(\boldsymbol{\phi} \mid \boldsymbol{\phi}^{(it)})$ with respect to $\boldsymbol{\phi}$ involves only the parts where the respective parameters appear. The covariance matrix of the measurement errors in the bivariate longitudinal model and the covariance matrix of the random effects have a closed-form update

$$\widehat{\mathbf{R}} = \begin{bmatrix} \hat{\sigma}_1^2 & 0 \\ 0 & \hat{\sigma}_2^2 \end{bmatrix} \quad \text{with}$$

$$\hat{\sigma}_k^2 = N_k^{-1} \sum_{i=1}^{n} \big(\boldsymbol{y}_{ik} - \mathbf{X}_{ik}\boldsymbol{\beta}_k\big)^T \big(\boldsymbol{y}_{ik} - \mathbf{X}_{ik}\boldsymbol{\beta}_k - 2\mathbf{Z}_{ik}\widetilde{\boldsymbol{b}}_{ik}\big) + \operatorname{tr}(\mathbf{Z}_{ik}^T \mathbf{Z}_{ik} \widetilde{v\boldsymbol{b}}_{ik}) + \widetilde{\boldsymbol{b}}_{ik}^T \mathbf{Z}_{ik}^T \mathbf{Z}_{ik} \widetilde{\boldsymbol{b}}_{ik} \ ,$$

$$\widehat{\mathbf{D}} = n^{-1} \sum_{i=1}^{n} \{ \widetilde{v\boldsymbol{b}}_i - \widetilde{\boldsymbol{b}}_i^T \widetilde{\boldsymbol{b}}_i \} \ ,$$

where $N_k = \sum_i n_{ik}$, $\widetilde{\boldsymbol{b}}_i = E(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) = \int \boldsymbol{b}_i \, p(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i$, and $\widetilde{v\boldsymbol{b}}_i = \mathrm{var}(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) = \int (\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i)^T (\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i) \, p(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i$.

The estimates of the fixed effects, $\boldsymbol{\beta}$, and the parameters, $\boldsymbol{\phi}_t$, of the survival sub-model do not have a closed-form updates and hence, we use one–step Newton–Raphson approach to update these parameters as follows:

$$\hat{\boldsymbol{\beta}}^{(it+1)} = \hat{\boldsymbol{\beta}}^{(it)} - \left\{ \frac{\partial S(\hat{\boldsymbol{\beta}}^{(it)})}{\partial \boldsymbol{\beta}} \right\}^{-1} S(\hat{\boldsymbol{\beta}}^{(it)}),$$

$$\hat{\boldsymbol{\phi}}_t^{(it+1)} = \hat{\boldsymbol{\phi}}_t^{(it)} - \left\{ \frac{\partial S(\hat{\boldsymbol{\phi}}_t^{(it)})}{\partial \boldsymbol{\phi}_t} \right\}^{-1} S(\hat{\boldsymbol{\phi}}_t^{(it)}),$$

where $\hat{\boldsymbol{\beta}}^{(it)}$ and $\hat{\boldsymbol{\phi}}_t^{(it)}$ denote the values of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}_t$ at the current iteration, respectively; $S(\hat{\boldsymbol{\beta}}^{(it)})$ and $S(\hat{\boldsymbol{\phi}}^{(it)})$ denote the score vector of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}_t$, evaluated at $\hat{\boldsymbol{\beta}}^{(it)}$ and $\hat{\boldsymbol{\phi}}_t^{(it)}$, respectively; $\partial S(\hat{\boldsymbol{\beta}}^{(it)})/\partial \boldsymbol{\beta}$ and $\partial S(\hat{\boldsymbol{\phi}}_t^{(it)})/\partial \boldsymbol{\phi}_t$ denote the corresponding blocks of the Hessian matrix, evaluated at $\hat{\boldsymbol{\beta}}^{(it)}$ and $\hat{\boldsymbol{\phi}}_t^{(it)}$, respectively. The components of the score vector of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}_t$ are given in Section B of the Supplementary Material.

To compute the standard errors for the parameter estimates, we first calculate the score vector

$$S(\hat{\boldsymbol{\phi}}) = \sum_{i=1}^{n} \int \left[\frac{\partial}{\partial \boldsymbol{\phi}} \log \left\{ p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i|\boldsymbol{b}_i, \boldsymbol{\phi}) p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\phi}) p(\boldsymbol{b}_i|\boldsymbol{\phi}) \right\} \right] p(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \, \mathrm{d}\boldsymbol{b}_i \, |_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}$$

and the standard errors are calculated as

$$\widehat{\mathrm{var}}(\hat{\boldsymbol{\phi}}) = -\{H(\hat{\boldsymbol{\phi}})\}^{-1}, \quad \text{with} \quad H(\hat{\boldsymbol{\phi}}) = \left. \frac{\partial S(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} \right|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}$$

We implemented our method in `R`, version 3.1, by using the `lme()` function for the longitudinal part; creating a function to create the bivariate Weibull model; and then modifying Rizopoulos' `JM` procedure to accommodate our bivariate joint model. Details of our algorithm are given in the Supplementary Material.

4 Simulation Studies

We conducted simulations to evaluate the performance of our proposed model. We considered three situations representing different levels of dependence (i.e., low, moderate, high) on the bivariate longitudinal measurement as well as the bivariate survival time. In each simulation, we constructed several sets of association parameters to assess how well our model estimates the effect of longitudinal measurements on the risk of events. We generated data sets with 600 "female–male" pairs as described below.

Simulation 1. Low dependence on both bivariate longitudinal and bivariate survival outcomes: For the bivariate longitudinal outcome, we assumed seven repeated measurements were taken at fixed times 0, 0.5, 1, 1.5, 2, 2.5, and 3 years. The measurement y_{ikj} of subject k (k=1, female; 2, male) in pair i ($i = 1, \ldots, 600$) at time t_{ikj} ($j = 1, \ldots, 7$) was generated from a bivariate linear mixed-effects model with random intercept

$$y_{ikj} = \beta_{0k} + \beta_{1k} t_{ikj} + b_{0ik} + \varepsilon_{ikj},$$
$$\boldsymbol{b_{0i}} = \begin{pmatrix} b_{0i1} \\ b_{0i2} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{D}),\ \boldsymbol{\varepsilon_i} = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{R}),$$

where $\begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$; $\begin{pmatrix} \beta_{11} \\ \beta_{12} \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.1 \end{pmatrix}$; $\mathbf{D} = \begin{pmatrix} \sigma^2_{b_1} & \sigma_{a_1 b_1} \\ \sigma_{a_1 b_1} & \sigma^2_{b_2} \end{pmatrix} = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.6 \end{pmatrix}$; we set $\sigma_{a_1 b_1} = 0.2$ to simulate a low correlation between the two longitudinal measurements (i.e., $r = \frac{0.2}{\sqrt{0.6 \times 0.7}} = 0.3$); and $\mathbf{R} = \begin{pmatrix} \sigma^2_1 & 0 \\ 0 & \sigma^2_2 \end{pmatrix} = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.6 \end{pmatrix}$. For the bivariate survival outcome, we first generated the hazard function $h_{ik}(t)$ of subject k in pair i at time t from a Weibull proportional hazards model with a gamma frailty given by

$$h_{ik}(t) = \rho t^{\rho - 1} \mu_i \exp\{\mathbf{w}_{ik} \boldsymbol{\gamma}_k + \alpha_k (\beta_{0k} + \beta_{1k} t + b_{0ik})\},$$
$$\mu_i \sim\ GAM\left(\frac{1}{\theta}, \theta\right),$$

where the intercept term in the baseline covariate vector ($\mathbf{w}_{ik}$) corresponds to $\log \lambda_k$; $\rho = 6$ and (λ_1, λ_2) = (0.1, 0.2) so that the median survival time was between 1.1-1.5 years and the maximum survival time was ≤ 3.5 years among both genders; We set $\theta = 0.5$ to simulate a low overall dependence between the two survival times (i.e., Kendall's $\tau = 0.5/(0.5 + 2.0) = 0.2$); covariates ($w_{i1}$, w_{i2}) were both generated from a binomial distribution with probability 0.5 and their corresponding parameters (γ_1, γ_2) = (0.7, 0.6). We considered four sets of association parameters: (1) (α_1, α_2) = (0.1, 0.05) represents a small effect of longitudinal measurement on survival time for both genders (i.e., the hazard ratio (HR) in both genders (HR_1, HR_2) = (1.11, 1.05), per unit increase in longitudinal measurement); (2) (α_1, α_2) = (0.5, 0.3) represents a moderate effect for both genders (i.e., (HR_1, HR_2) = (1.65, 1.35)); (3) (α_1, α_2) = (1.0, 0.8) represents a large effect for both genders (i.e., (HR_1, HR_2) = (2.72, 2.23)); (4) (α_1, α_2) = (1.0, 0.05) represents a large effect for females but a small effect for males. Using $S_{ik}(t) = \exp\left\{ -\int_0^t h_{ik}(s)\, \mathrm{d}s \right\}$ and $S_{ik}(t) \sim U(0, 1)$, we generated event times by randomly generating $S_{ik}(t)$, then solving for t using the uniroot() and integrate() functions in R (Ihaka and Gentleman, 1996). Observations were censored with a probability of 0.2 for both genders. A censored subject's censoring time was chosen uniformly over the interval (0, t). Finally, the longitudinal measurements were censored when they were taken after the event times.

Simulation 2. Moderate dependence on both bivariate longitudinal and bivariate survival outcomes: With the same model settings in simulation 1, except here we set $\sigma_{a_1 b_1} = 0.4$

and $\theta = 2$ to simulate a moderate dependence on both longitudinal ($r = 0.6$) and survival (Kendall's $\tau = 0.5$) outcomes. We set $\rho = 12$ and $(\lambda_1, \lambda_2) = (1.0, 1.0)$ so that the median survival time was between 1.1-1.5 years and the maximum survival time was ≤ 3.5 years among both genders. Three sets of the association parameters were considered: $(\alpha_1, \alpha_2) = (0.1, 0.05)$, $(0.5, 0.3)$, and $(1.0, 0.8)$.

Simulation 3. High dependence on both bivariate longitudinal and bivariate survival outcomes: With the same model settings in simulation 1, except here we set $\sigma_{a_1 b_1} = 0.5$ ($r = 0.8$), $\theta = 6$ (Kendall's $\tau = 0.8$), $\rho = 40$, and $(\lambda_1, \lambda_2) = (1.2, 1.2)$. Three sets of the association parameters were considered: $(\alpha_1, \alpha_2) = \{(0.1, 0.05), (0.5, 0.3), (1.0, 0.8)\}$.

For each scenario, 1000 replications were conducted. We evaluate the model performance using the mean bias in the estimates, the mean standard error of the estimates, and the coverage probability of the estimated 95% confidence intervals. The results are given in Tables 1, 2, and 3. The simulation evidence suggests that overall our model performs well under all simulated circumstances from two perspectives. The biases of the estimates are all minimal and the coverage probabilities are close to or achieve the nominal level.

5 Application to the Cardiovascular Health Study (CHS)

We applied the proposed joint models to the spouse-pair data from the CHS (1) to investigate the association of both longitudinal depressive symptoms scores and mortality between husbands and wives in older adults, controlling for covariates associated with depressive symptoms and mortality separately, and (2) to characterize mortality in both genders based on their own longitudinal depressive symptoms score and other factors.

5.1 Study Population

The sample used in the study was obtained from the CHS, a prospective, observational study designated to identify the risk factors for and consequences of cardiovascular disease (CVD) in older adults. Adults 65 years and older were recruited from random samples of Medicare eligibility lists in four communities — Sacramento County, California; Washington County, Maryland; Forsyth County, North Carolina; and Pittsburgh, Pennsylvania — and from age-eligible participants in the same household. A total of 5201 men and women 65 years or older were enrolled in 1989 to 1990 (cohort 1), and a supplemental cohort of 687 African Americans was enrolled from 1992 through 1993 (cohort 2). Further details regarding CHS sampling and recruitment can be found in Fried et al. (1991) and Tell et al. (1993). Participants underwent annual clinical examinations and health assessments and were followed for coronary events and mortality. The follow-up length is 18 years for cohort 1 and 15 years for cohort 2. For the present study, a total of 1330 married couples from across the two cohorts (cohort 1, $n = 2520$; cohort 2, $n = 140$) were identified in the CHS sample.

5.2 Outcomes and Covariates

Mortality — The CHS has complete follow-up on mortality (18 years for cohort 1 and 15 years for cohort 2). Deaths were confirmed through reviews of obituaries, medical records, death certificates, and the Health Care Financing Administration healthcare database for hospitalizations. The survival time was defined as the time from enrollment to death.

Depressive symptoms — Depressive symptoms were assessed annually up to 10 years in cohort 1 and 6 years in cohort 2. Severity of depressive symptoms was evaluated using the previously validated 10-item Center for Epidemiological Studies Depression Scale (CES-D) (Andresen et al., 1994; Radloff, 1977) at baseline and yearly throughout the follow-up. The CES-D score was between 0 and 30 with a higher score indicating a greater severity of depressive symptoms.

Covariates — Sociodemographic factors included baseline age, race (white or non-white), education (the highest grade or year of school ever completed), stressful life events (total of 10 possible stressful life events in past 6 months), and annual income. Health behavior factors included smoking status (never, former, or current), alcohol consumption (drinks per week), and body mass index (BMI). Functional disability was assessed using difficulty with activities of daily living (ADL)/instrumental activities of daily (IADL). It was yes for the presence of any self-reported difficulty in walking, getting in and out of a bed or chair, eating, dressing, bathing, or using the toilet (ADL), or any difficulty with heavy housework, light housework, shopping, preparing meals, managing money, or using the telephone (IADL) "because of health or physical problems". Cognitive status was estimated using the modified Mini-Mental State Examination (3MS), with a higher score (range 0-30) indicating better functioning (Teng and Chui, 1897).

Caregiving status was evaluated by asking participants if they ever provided help with IADL. CVD was measured as follows: a) prevalent clinical disease including angina pectoris, myocardial infarction, bypass, congestive heart failure, intermittent claudication, stroke, and transient ischemic attack, and b) subclinical disease, indicative of risk for CVD but without clinical manifestations, including the Rose questionnaires for claudication and angina ratio of ankle to arm blood pressure, major electrocardiogram abnormality, and carotid stenosis (Kuller et al., 1995; Psaty et al., 1995). Antidepressant medication use was defined as taking any medication classified as an antidepressant (i.e., non-tricyclic antidepressants other than monoamine oxidase inhibitors (MAOIs), tricyclic anti-depressants, or tri-cyclic anti-depressants plus anti-psychotics).

5.3 Statistical Analysis

Longitudinal submodel: To approximate the normality assumption for longitudinal CES-D score, we took the square root transformation of CES-D scores. We considered quadratic models for fitting the longitudinal fixed effects of the $\sqrt{\text{CES-D}}$ scores but the quadratic terms of time were not significant. Hence, in the bivariate linear mixed-effects models we used a linear fixed effects bivariate longitudinal model and found that using

random intercepts only fit the data best with the smallest AIC and BIC values. Thus, the bivariate linear population model with random intercepts was used to model the longitudinal part of the joint model. In addition, we tried to identify baseline factors related to longitudinal $\sqrt{\text{CES-D}}$ scores. Education, income, stressful life events, BMI, smoking status, alcohol consumption, caregiving status, ADL/IADL difficulty, 3MS score, prevalent clinical CVD, subclinical CVD, and antidepressant medication use all were considered. Univariate analyses were first conducted to test the association of each individual baseline variable with $\sqrt{\text{CES-D}}$ score. Significant variables in the univariate analyses ($P<.10$) were then included in the multivariable models, controlling for age and race, and retained if statistically significant ($P<.05$).

Survival submodel: Using the same model building strategy, we fit Weibull proportional hazards models with a gamma frailty to identify baseline factors related to mortality. Education, income, stressful life events, BMI, smoking status, alcohol consumption, ADL/IADL difficulty, 3MS score, prevalent clinical CVD, and subclinical CVD were considered.

Joint model: The two submodels above were then linked together through the random effects used in the bivariate linear mixed-effects models. Parameter estimates in joint models were obtained by maximizing the joint likelihood for the two submodels using the EM algorithm. We first looked at unadjusted joint models, where no covariate was included in the bivariate linear mixed-effects model or the Weibull proportional hazards model with a gamma frailty. Then the adjusted joint models were built, controlling for the covariates independently associated with longitudinal $\sqrt{\text{CES-D}}$ scores in the bivariate linear mixed-effects model as well as the covariates independently associated with mortality in the Weibull proportional hazards model with a gamma frailty. The marginal correlation between husbands' and wives' $\sqrt{\text{CES-D}}$ scores was calculated and the dependence of husbands' and wives' mortality was measured by Kendall's τ.

Because the CHS sample includes two cohorts with different follow-up periods for mortality and depressive symptoms, we truncated both outcomes at 6 years in both cohorts to ensure a same minimal follow-up period for both outcomes in both cohorts. To understand the impact of ignoring the correlation between husbands and wives on the estimation of the association parameters of CES-D score and mortality, we also fit data using a joint model where the correlations were not considered.

5.4 Results

Baseline characteristics of the analysis sample stratified by sex are presented in eTable 1 of the Supplementary Material. Out of 1330 spouse pairs, 117(8.8%) wives and 334 (25.1%) husbands died within six years.

Table 4 shows the results of unadjusted and adjusted joint models. Without controlling for any covariate, an increase of one unit of the $\sqrt{\text{CES-D}}$ score was associated with 82%

(95%CI, 1.42-2.34) and 66% (95%CI, 1.40-1.96) higher risks of mortality in wives and husbands, respectively. A moderate correlation of $\sqrt{\text{CES-D}}$ scores (r=0.36) and low dependence of mortality (Kendall's τ=0.21) between husbands and wives were observed. The $\sqrt{\text{CES-D}}$ score was still associated with mortality after adjusting for the covariates, where the mortality increased 45% (95%CI, 1.10-1.91) in wives and 35% (95%CI, 1.13-1.61) in husbands with one unit of increase in the $\sqrt{\text{CES-D}}$ score. These estimates are obtained from taking $\exp(\alpha_1)$ and $\exp(\alpha_2)$, where α_1 and α_2 are the parameters associating the longitudinal and survival sub-models in equation (2.1). The $\sqrt{\text{CES-D}}$ score increased with time in both genders. Older age, less education, having stressful life events, having ADL/IADL difficulty, prevalent clinical CVD, and being on antidepressant medication were independently related to the longitudinal $\sqrt{\text{CES-D}}$ score in both genders. Non-white race was associated to the longitudinal $\sqrt{\text{CES-D}}$ scores in husbands only. Older age, and having prevalent clinical CVD and subclinical CVD were independently associated with mortality in both genders. Having ADL/IADL difficulty was associated with mortality in husbands only. The correlation of $\sqrt{\text{CES-D}}$ scores and the dependence of mortality between husbands and wives became smaller after adjusting for the covariates (r = 0.30; Kendall's τ=0.13). When the correlations between husbands and wives were ignored in the model, the estimate of the wives' association between the $\sqrt{\text{CES-D}}$ score and mortality didn't change much but had a larger standard error (i.e., HR(95%CI)=1.83(1.38-2.43)) and husbands' association estimate became smaller and also had a larger standard error (i.e., HR(95%CI) = 1.55(1.32-1.82)).

5.5 Discussion of Results

Utilizing our new joint modeling approach to simultaneously investigate the association of both longitudinal severity of depressive symptoms and mortality between husbands and wives, we found that longitudinal severity of depressive symptoms was a significant independent risk factor for mortality in both husbands and wives after adjusting for covariates. Our analyses also showed that the associations between longitudinal severity of depressive symptoms and mortality were slightly attenuated after adjusting for covariates. It is believed that depression-mortality effect is driven by an underlying psychological state that includes elements of health and functioning. We would expect this effect to be shared and diluted among a wide range of health and functional status factors. Our study also points out the importance of taking into account the correlations between husbands and wives in the joint models. Ignoring such correlations may result in a less accurate estimate of the true association between the longitudinal severity of depressive symptoms and mortality.

The longitudinal parameter estimates obtained from the joint models indicate that older age, less educated, and having more stressful life events, ADL/IADL difficulty, prevalent clinical CVD, and antidepressant medication use were independently associated with depressive symptoms changes over time in both genders. However, a significant association between antidepressant medication use and longitudinal $\sqrt{\text{CES-D}}$ score does not mean that depressive symptoms were worse among those taking medication. This association cannot be established to be causal, and we assert that antidepressant medication does not cause

depressive symptoms but rather that antidepressant medication use is a proxy for having clinical depression (i.e., severe depressive symptoms). From a clinical perspective, the current findings suggest that severity of depressive symptoms measured by existing screening tests should be taken seriously and further evaluated for possible treatment to stop the progression of depressive symptoms, thus enhancing quality of life and longevity in older people.

6 Conclusions

We have proposed a joint modeling approach for paired data which took into account the within-pair correlation, both in the longitudinal and in the time-to-event processes. Our method offers a feasible approach to connect the long-term course of psychiatric conditions to the time to mortality in paired subjects and simultaneously investigate the association of both longitudinal psychiatric conditions and mortality within pairs. Application of the methodology and simulation evidence show that it is accessible for routine use and provides reliable inference.

There is much potential future work related to this research. The advantage of our choosing parametric proportional hazards models with gamma frailties is that the marginal likelihood is fully parametric and we can rely on classical maximum likelihood techniques and the EM algorithm to estimate the parameters. However, we may not know what the appropriate baseline hazard distribution is and it may be more appropriate not to make any assumption on its distribution. Moreover, the gamma distribution was used in the study due to its simple interpretation and mathematical tractability. However, other distributions, such as the positive stable and the inverse Gaussian distributions proposed in the literature for frailty, may also be considered and compared with that used in our method.

The simulations we performed in this work largely assessed how our model performed with differing levels of dependence within and across the longitudinal and survival processes. However, even though the Weibull models are very flexible, future work would involve assessing, through simulation or otherwise, the robustness of the model under situations where either or both of the longitudinal or survival processes are misspecified. Due to the complexity of the model, many combinations of misspecifications could occur within either or both of the longitudinal and survival sub-processes. Furthermore, a thorough comparison of our model to that of Chi and Ibrahim (2006) would be helpful and is an area of future research.

Acknowledgements

This work was supported by the grant P30 MH090333 from the National Institute of Mental Health and by contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute

(NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at `CHS-NHLBI.org`. The authors declare no conflicts of interest.

The authors would like to thank the reviewers for their useful comments which helped us improve our manuscript.

Table 1: Results of Simulation 1: Low dependence on both bivariate longitudinal and bivariate survival outcomes with 20% censoring rate and 1000 replications.

	Scenario 1				**Scenario 2**				**Scenario 3**				**Scenario 4**			
Parameter	True	Bias	SE	CP	True	Bias	SE	CP	True	Bias	SE	CP	True	Bias	SE	CP
β_{01}	2.0	0.000	0.044	0.949	2.0	-0.002	0.045	0.948	2.0	-0.002	0.045	0.948	2.0	-0.001	0.045	0.938
β_{02}	1.0	-0.001	0.042	0.950	1.0	-0.001	0.043	0.949	1.0	-0.002	0.043	0.952	1.0	-0.002	0.042	0.950
β_{11}	0.2	-0.001	0.041	0.949	0.2	-0.005	0.049	0.943	0.2	-0.010	0.061	0.939	0.2	-0.012	0.061	0.936
β_{12}	0.1	-0.001	0.046	0.953	0.1	-0.002	0.048	0.959	0.1	-0.003	0.053	0.947	0.1	0.001	0.046	0.954
σ_1	0.77	-0.001	0.016	0.948	0.77	0.000	0.018	0.948	0.77	-0.002	0.020	0.953	0.77	-0.002	0.020	0.957
σ_2	0.77	-0.001	0.017	0.949	0.77	-0.001	0.018	0.949	0.77	-0.001	0.019	0.949	0.77	-0.001	0.017	0.944
$\sigma^2_{b_1}$	0.7	0.001	0.046	0.911	0.7	-0.001	0.048	0.915	0.7	-0.001	0.050	0.914	0.7	0.001	0.050	0.917
$\sigma_{b_1 b_2}$	0.2	0.000	0.040	0.975	0.2	0.000	0.041	0.970	0.2	0.002	0.041	0.972	0.2	0.001	0.041	0.975
$\sigma^2_{b_2}$	0.6	-0.001	0.046	0.946	0.6	-0.003	0.046	0.935	0.6	-0.002	0.048	0.929	0.6	-0.000	0.046	0.943
θ	0.5	0.015	0.070	0.948	0.5	0.017	0.071	0.948	0.5	0.015	0.075	0.949	0.5	0.009	0.073	0.947
ρ	6.0	0.113	0.211	0.933	6.0	0.114	0.214	0.925	6.0	0.125	0.230	0.934	6.0	0.100	0.222	0.951
$\log \lambda_1$	-2.30	-0.090	0.213	0.935	-2.30	-0.097	0.226	0.941	-2.30	-0.093	0.254	0.963	-2.30	-0.049	0.248	0.983
$\log \lambda_2$	-1.61	-0.081	0.140	0.915	-1.61	-0.082	0.144	0.913	-1.61	-0.089	0.158	0.933	-1.61	-0.077	0.142	0.936
γ_1	0.7	0.011	0.122	0.953	0.7	0.009	0.124	0.954	0.7	0.008	0.130	0.955	0.7	0.006	0.129	0.948
γ_2	0.6	0.008	0.121	0.944	0.6	0.007	0.122	0.950	0.6	0.010	0.127	0.950	0.6	0.008	0.122	0.950
α_1	0.1	-0.001	0.082	0.943	0.5	0.012	0.089	0.950	1.0	0.021	0.109	0.962	1.0	0.002	0.106	0.986
α_2	0.05	0.002	0.091	0.946	0.3	0.008	0.094	0.950	0.8	0.025	0.108	0.953	0.01	0.004	0.091	0.951

True, true value; Bias, the mean bias of estimates; SE, the mean standard error of estimates; CP, the coverage probability of the estimated 95% confidence intervals.

Table 2: Results of Simulation 2: Moderate dependence on both bivariate longitudinal and survival outcomes with 20% censoring rate and 1000 replications.

	Scenario 1				**Scenario 2**				**Scenario 3**			
Parameter	True	Bias	SE	CP	True	Bias	SE	CP	True	Bias	SE	CP
β_{01}	2.0	-0.000	0.045	0.945	2.0	-0.003	0.045	0.949	2.0	-0.001	0.045	0.957
β_{02}	1.0	-0.001	0.043	0.954	1.0	-0.002	0.043	0.949	1.0	-0.003	0.043	0.955
β_{11}	0.2	0.001	0.056	0.955	0.2	0.000	0.062	0.952	0.2	-0.006	0.069	0.948
β_{12}	0.1	0.000	0.054	0.951	0.1	0.000	0.056	0.947	0.1	-0.002	0.059	0.955
σ_1	0.77	-0.001	0.019	0.948	0.77	0.000	0.020	0.958	0.77	-0.001	0.021	0.958
σ_2	0.77	0.000	0.019	0.949	0.77	0.000	0.019	0.960	0.77	-0.001	0.020	0.957
$\sigma_{b_1}^2$	0.7	-0.002	0.049	0.886	0.7	-0.001	0.050	0.903	0.7	0.001	0.051	0.910
$\sigma_{b_1 b_2}$	0.4	-0.002	0.045	0.968	0.4	-0.002	0.045	0.967	0.4	-0.001	0.045	0.967
$\sigma_{b_2}^2$	0.6	-0.003	0.050	0.950	0.6	-0.003	0.051	0.961	0.6	-0.004	0.051	0.952
θ	2.0	0.032	0.161	0.958	2.0	0.045	0.163	0.949	2.0	0.035	0.167	0.951
ρ	12.0	0.121	0.450	0.943	12.0	0.145	0.458	0.950	12.0	0.114	0.484	0.961
$\log \lambda_1$	0.0	-0.041	0.279	0.950	0.0	-0.040	0.286	0.950	0.0	0.003	0.303	0.972
$\log \lambda_2$	0.0	-0.036	0.186	0.954	0.0	-0.035	0.188	0.951	0.0	-0.042	0.198	0.946
γ_1	0.7	0.013	0.149	0.950	0.7	0.008	0.150	0.952	0.7	0.002	0.155	0.956
γ_2	0.6	0.006	0.148	0.942	0.6	0.013	0.149	0.949	0.6	0.011	0.153	0.944
α_1	0.1	0.002	0.114	0.945	0.5	0.009	0.120	0.942	1.0	-0.009	0.138	0.963
α_2	0.05	-0.002	0.125	0.946	0.3	0.002	0.128	0.950	0.8	0.007	0.142	0.951

True, true value; Bias, the mean bias of estimates; SE, the mean standard error of estimates;

CP, the coverage probability of the estimated 95% confidence intervals.

Table 3: Results of Simulation 3: High dependence on both bivariate longitudinal and bivariate survival outcomes with 20% censoring rate and 1000 replications.

	Scenario 1				**Scenario 2**				**Scenario 3**			
Parameter	True	Bias	SE	CP	True	Bias	SE	CP	True	Bias	SE	CP
β_{01}	2.0	0.001	0.04	0.962	2.0	0.000	0.045	0.950	2.0	0.000	0.045	0.946
β_{02}	1.0	0.002	0.04	0.938	1.0	0.000	0.043	0.934	1.0	0.000	0.043	0.948
β_{11}	0.2	0.000	0.05	0.953	0.2	-0.001	0.054	0.955	0.2	0.000	0.056	0.950
β_{12}	0.1	-0.002	0.05	0.948	0.1	-0.001	0.051	0.939	0.1	-0.001	0.053	0.952
σ_1	0.77	-0.001	0.01	0.959	0.77	-0.001	0.019	0.956	0.77	-0.004	0.020	0.950
σ_2	0.77	0.000	0.01	0.948	0.77	-0.001	0.019	0.957	0.77	-0.002	0.019	0.952
$\sigma^2_{b_1}$	0.7	0.001	0.04	0.903	0.7	-0.001	0.049	0.913	0.7	0.008	0.049	0.932
$\sigma_{b_1 b_2}$	0.5	-0.001	0.04	0.968	0.5	-0.001	0.048	0.972	0.5	0.001	0.048	0.971
$\sigma^2_{b_2}$	0.6	-0.001	0.05	0.971	0.6	-0.001	0.053	0.967	0.6	0.000	0.053	0.974
θ	6.0	0.091	0.40	0.959	6.0	0.099	0.406	0.959	6.0	0.042	0.410	0.950
ρ	40.0	0.262	1.60	0.953	40.0	0.337	1.633	0.957	40.0	0.051	1.685	0.959
$\log \lambda_1$	0.18	-0.008	0.40	0.950	0.18	-0.015	0.407	0.957	0.18	0.175	0.415	0.960
$\log \lambda_2$	0.18	-0.016	0.26	0.946	0.18	-0.020	0.268	0.953	0.18	0.021	0.275	0.951
γ_1	0.7	0.004	0.16	0.954	0.7	-0.004	0.168	0.945	0.7	0.000	0.171	0.951
γ_2	0.6	0.007	0.16	0.946	0.6	0.008	0.167	0.945	0.6	-0.002	0.170	0.952
α_1	0.1	-0.003	0.16	0.950	0.5	0.008	0.170	0.954	1.0	-0.089	0.183	0.957
α_2	0.05	-0.003	0.17	0.950	0.3	0.005	0.182	0.952	0.8	-0.042	0.195	0.952

True, true value; Bias, the mean bias of estimates; SE, the mean standard error of estimates;

CP, the coverage probability of the estimated 95% confidence intervals.

Table 4: Results of joint modeling of bivariate longitudinal CES-D score and bivariate mortality in the CHS spouse pairs.

	Unadjusted (N=1330 spouse pairs)		Adjusted for covariates (N=1277 spouse pairs)	
	Estimate (95% CI)	HR (95% CI)	Estimate (95% CI)	AHR (95% CI)
Longitudinal model:				
Wives				
Intercept	1.87 (1.82, 1.92)		1.19 (0.46, 1.92)	
Time, per 1 y	0.07 (0.06, 0.08)		0.07 (0.06, 0.08)	
Age, per 1 y			0.01 (0.00, 0.02)	
White			-0.07 (-0.25, 0.10)	
Education, per 1 y			-0.03 (-0.04, -0.01)	
Stressful life event, per 1			0.12 (0.08, 0.16)	
Any ADL/IADL difficulty			0.42 (0.32, 0.52)	
Prevalent clinical CVD			0.18 (0.06, 0.30)	
Antidepressant medication use			0.39 (0.19, 0.60)	
σ_1	0.76 (0.75, 0.77)		0.76 (0.75, 0.77)	
Husbands				
Intercept	1.56 (1.51, 1.61)		0.94 (0.26, 1.61)	
Time, per 1 y	0.08 (0.07, 0.08)		0.08 (0.07, 0.09)	
Age, per 1 y			0.01 (0.00, 0.02)	
White			-0.21 (-0.38, -0.04)	
Education, per 1 y			-0.03 (-0.03, -0.02)	
Stressful life event, per 1			0.12 (0.07, 0.16)	
Any ADL/IADL difficulty			0.42 (0.31, 0.53)	
Prevalent clinical CVD			0.19 (0.10, 0.29)	
Antidepressant medication use			0.63 (0.33, 0.93)	
σ_2	0.76 (0.75, 0.77)		0.76 (0.75, 0.78)	
$\sigma^2_{b_1}$	0.68 (0.63, 0.73)		0.57 (0.52, 0.62)	
$\sigma_{b_1 b_2}$	0.24 (0.19, 0.28)		0.17 (0.13, 0.21)	
$\sigma^2_{b_2}$	0.66 (0.61, 0.71)		0.55 (0.50, 0.60)	
Survival model:				
Wives				
$\log \lambda_1$	-6.39 (-7.07, -5.70)		-15.11 (-17.87, -12.35)	
Age, per 1 y			0.11 (0.08, 0.15)	1.12 (1.08, 1.16)
White			0.39 (-0.43, 1.22)	1.48 (0.65, 3.38)
Any ADL/IADL difficulty			0.18 (-0.25, 0.60)	1.19 (0.78, 1.83)
Prevalent clinical CVD			0.63 (0.21, 1.05)	1.87 (1.23, 2.86)
Subclinical CVD			0.51 (0.07, 0.96)	1.67 (1.08, 2.60)
$\sqrt{\text{CES-D}}$ score, per 1	0.60 (0.35, 0.85)	1.82 (1.42, 2.34)	0.37 (0.09, 0.65)	1.45 (1.10, 1.91)
Husbands				
$\log \lambda_2$	-4.79 (-5.20, -4.37)		-10.00 (-11.72, -8.28)	
Age, per 1 y			0.06 (0.04, 0.09)	1.07 (1.04, 1.09)
White			-0.19 (-0.60, 0.22)	0.83 (0.55, 1.25)
Any ADL/IADL difficulty			0.70 (0.44, 0.97)	2.02 (1.55, 2.64)
Prevalent clinical CVD			0.27 (0.03, 0.51)	1.31 (1.03, 1.67)
Subclinical CVD			0.80 (0.46, 1.14)	2.23 (1.59, 3.14)
$\sqrt{\text{CES-D}}$ score, per 1	0.51 (0.34, 0.68)	1.66 (1.40, 1.96)	0.30 (0.12, 0.48)	1.35 (1.13, 1.61)
θ	0.52 (0.08, 0.97)		0.29 (-0.06, 0.65)	
ρ	1.41 (1.27, 1.54)		1.47 (1.33, 1.61)	

Abbreviation: CI, confidence interval; HR, hazard ratio; AHR, adjusted hazard ratio; ADL, Activities of Daily Living IADL, Instrumental Activities of Daily; CVD, cardiovascular disease; CES-D, Center for Epidemiological Studies Depression Scale.

References

Andresen, E., Malmgren, J., and Carter, W. (1994), "Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale)," *American Journal of Preventive Medicine*, 10, 77–84.

Blazer, D., Hughes, D., and George, L. (1987), "The epidemiology of depression in an elderly community population," *The Gerontologist*, 27, 281–287.

Brown, E., Ibrahim, J., and DeGruttola, V. (2005), "A flexible B-spline model for multiple longitudinal biomarkers and survival," *Biometrics*, 61(1), 64–73.

Chi, Y. and Ibrahim, J. (2006), "Joint models for multivariate longitudinal and multivariate survival data," *Biometrics*, 62(2), 432–445.

Choi, J., Anderson, S., Richardsc, T., and Thompsond, W. (2014), "Prediction of transplant-free survival in idiopathic pulmonary fibrosis patients using joint models for event times and mixed multivariate longitudinal data," *Journal of Applied Statistics*, 41(10), 2192–2205.

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

Clayton, D. and Cuzick, J. (1985), "Multivariate generalizations of the proportional hazards model," *Journal of the Royal Statistical Society: Series A*, 148, 82–117.

Dang, Q., Anderson, S., Tan, L., and Mazumdar, S. (2005), "Modeling unequally spaced bivariate growth curve data using a Kalman filter approach," *Communications in Statistics, Theory and Methods*, 34(8), 1821–1831.

Elashoff, R., Li, G., and Li, N. (2007), "An approach to joint analysis of longitudinal measurements and competing risks failure time data," *Stat. Med*, 26(14), 2813–2835.

Fisher, J., Zeiss, A., and Carstensen, L. (1993), "Psychopathology in the aged," in *Comprehensive handbook of psychopathology*, eds. Sutker, P. and Adams, H., New York: Plenum, 2nd ed., p. 815–842.

Freund, J. (1961), "A bivariate extension of the exponential distribution," *Journal of the American Statistical Association*, 56, 971–977.

Fried, L., Borhani, N., Enright, P., Furberg, C., Gardin, J., Kronmal, R., Kuller, L., Manolio, T., Mittelmark, M., and Newman, A. (1991), "The cardiovascular health study: design and rationale," *Ann Epidemiol*, 1(3), 263–276.

Ghosh, S. and Gelfand, A. (1998), "Latent waiting time models for bivariate event times with censoring," *Sankhya*, 60(1), 34–47.

Gumbel, E. (1960), "Bivariate exponential distributions," *Journal of the American Statistical Association*, 55, 698–707.

Guo, G. and Rodriguez, G. (1992), "Estimating a Multivariate Proportional Hazards Model for Clustered Data Using the EM Algorithm, with an Application to Child Survival in Guatemala," *Journal of the American Statistical Association*, 87(420), 969–976.

Hanagal, D. (2006), "A gamma frailty regression model in bivariate survival data," *IAPQR Transactions*, 31, 73–83.

— (2007), "Gamma frailty regression models in mixture distributions," *Economic Quality Control*, 22(2), 295–302.

Hanagal, D. and Dabade, A. (2013), "Bayesian estimation of parameters and comparison of shared gamma frailty models. Comm. Statist," *Communications in Statistics - Simulation and Computation*, 42(4), 910–931.

Hanagal, D. and Pandey, A. (2014), "Gamma shared frailty model based on reversed hazard rate for bivariate survival data," *Statistics and Probability Letters*, 88(1), 190–196.

Henderson, R., Diggle, P., and Dobson, A. (2000), "Joint modelling of longitudinal measurements and event time data," *Biostatistics*, 1(4), 465–480.

Ihaka, R. and Gentleman, R. (1996), "R: A language for data analysis and graphics," *Journal of Computational and Graphical Statistics*.

Kennedy, G., Kelman, H., Thomas, C., Wisniewsky, W., Metz, H., and Bijur, P. (1989), "Hierarchy of characteristics associated with depressive symptoms in an urban elderly sample," *American Journal of Psychiatry*, 146, 220–225.

Klein, J., Keiding, N., and Kamby, C. (1989), "Semiparametric Marshall-Olkin models applied to the occurrence of metastases at multiple sites after breast cancer," *Biometrics*, 45, 1073–1086.

Kuller, L., Shemanski, L., Psaty, B., Borhani, N., Gardin, J., Haan, M., O'Leary, D., Savage, P., Tell, G., and Tracy, R. (1995), "Subclinical disease as an independent risk factor for cardiovascular disease," *Circulation*, 92, 720–726.

Lancaster, T. (1979), "Econometric methods for the duration of unemployment," *Econometrica*, 47, 939–956.

Lin, H., McCulloch, C., and Mayne, S. (2002), "Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables," *Statistics in Medicine*, 21(16), 2369–2382.

Liu, L., Huang, X., and O'Quigley, J. (2008), "Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data," *Biometrics*, 64(3), 950–958.

Marshall, A. and Olkin, I. (1967), "A multivariate exponential distribution," *Journal of the American Statistical Association*, 62, 30–44.

Morrell, C., Brant, L., Sheng, S., and Najjar, S. (2003), "Using multivariate mixed effect model to predict hypertension," *In Joint Statistical Meetings, Biometrics Section*.

Murrell, S., Himmelfarb, S., and Wright, K. (1983), "Prevalence of depression and its correlates in older adults," *American Journal of Epidemiology*, 111, 173–185.

Oakes, D. (1982), "A model for association in bivariate survival data," *Journal of the Royal Statistical Society: Series B*, 44(3), 414–422.

Potthoff, R. and Roy, S. (1964), "A generalized multivariate analysis of variance model useful especially for growth curve problems," *Biometrika*, 51 (3/4), 313–326.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (2007), *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New York, 3rd ed.

Psaty, B., Kuller, L., Bild, D., Burke, G., Kittne, r. S., Mittelmark, M., Price, T., Rautaharju, P., and Robbins, J. (1995), "Methods of assessing prevalent cardiovascular disease in the Cardiovascular Health Study," *Ann Epidemiol*, 5, 270–277.

Radloff, L. (1977), "The CES-D scale: a self-report depression scale for research in the general population," *Applied Psychological Measurement*, 1, 385–401.

Reinsel, G. (1982), "Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure," *Journal of the American Statistical Association*, 77, 190–195.

— (1984), "Estimation and prediction in a multivariate random effect generalized linear model," *Journal of the American Statistical Association*, 79, 406–414.

Rizopoulos, D. (2012), "Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule," *Computational Statistics and Data Analysis*, 56(3), 491–501.

Rizopoulos, D. and Ghosh, P. (2011), "A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event," *Statistics in Medicine*, 30(12), 1366–1380.

Ruppert, D., Wand, W., and Carroll, R. (2003), *Semiparametric regression*, Cambridge: Cambridge University Press.

Schafer, J. (1997), "Imputation of missing covariates under a multivariate linear mixed model," Tech. rep., Dept. of Statistics, Penn State University.

Schafer, J. and Yucel, R. (2002), "Computational strategies for multivariate linear mixed-effects models with missing values," *Journal of Computational and Graphical Statistics*, 11, 437–457.

Schulz, R., Beach, S., Ives, D., Martire, L., Ariyo, A., and Kop, W. (2000), "Association between depression and mortality in older adults," *Archives of Internal Medicine*, 160(12), 1761–1768.

Schulz, R., Drayer, R., and Rollman, B. (2002), "Depression as a risk factor for non-suicide mortality in the elderly," *Biological Psychiatry*, 52(3), 205–225.

Shah, A., Laird, N., and Schoenfeld, D. (1997), "A random effects model with multiple characteristics with possibly missing data," *Journal of the American Statistical Association*, 92, 775–779.

Song, X., Davidian, M., and Tsiatis, A. (2002), "An estimator for the proportional hazards model with multiple longitudinal covariates measured with errors," *Biostatistics*, 3, 511–528.

Tell, G., Fried, L., and Hermanson, B. (1993), "Recruitment of adults 65 years and older as participants in the Cardiovascular Health Study," *Ann Epidemiol*, 3, 358–366.

Teng, E. and Chui, H. (1897), "The modified mini-mental state (3MS) examination," *J Clin Psychiatry*, 48, 314 –318.

Thiébaut, R., Jacqmin-Gadda, H., Babiker, A., Commenges, D., and Collaboration, C. (2005), "Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection," *Statistics in Medicine*, 24(1), 65–82.

Vaupel, J., Manton, K., and Stallard, E. (1979), "The impact of heterogeneity in individual frailty and the dynamics of mortality," *Demography*, 16, 439–454.

Wienke, A., Lichtenstein, P., and Yashin, A. (2003), "A bivariate frailty model with a cure fraction for modeling familial correlations in diseases," *Biometrics*, 59(4), 1178–1183.

Wulsin, L., Vaillant, G., and Wells, V. (1999), "A systematic review of the mortality of depression," *Psychosomatic Medicine*, 61(1), 6–17.

Zhang, J., Kahana, B., Kahana, E., Hu, B., and Pozuelo, L. (2009), "Joint modeling of longitudinal changes in depressive symptoms and mortality in a sample of community-dwelling elderly people," *Psychosom Med*, 71(7), 704–714.

Received: May 19, 2019

Accepted: June 16, 2019