JOINT MODELING OF BIVARIATE LONGITUDINAL AND SURVIVAL DATA IN SPOUSE PAIRS

by

Jia-Yuh Chen

B.S., Environmental Engineering, National Cheng Kung University, Taiwan, 2003

M.S., Environmental Health, National Taiwan University, Taiwan, 2005

M.S., Biostatistics, University of Pittsburgh, 2009

Submitted to the Graduate Faculty of

Department of Biostatistics

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Jia-Yuh Chen

It was defended on

May 20, 2016

and approved by

Stewart J. Anderson, PhD, Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Ying Ding, PhD, Assistant Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Wendy C. King, PhD, Associate Professor, Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh

Richard Schulz, PhD, Professor, Department of Psychiatry, School of Medicine, University of Pittsburgh

Ada O. Youk, PhD, Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Dissertation Director: Stewart J. Anderson, PhD, Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

JOINT MODELING OF BIVARIATE LONGITUDINAL AND SURVIVAL DATA IN SPOUSE PAIRS

Jia-Yuh Chen, PhD

University of Pittsburgh, 2016

ABSTRACT

Joint modeling of longitudinal and survival data has become increasingly useful for analyzing clinical trials data. Recent multivariate joint models relate one or more longitudinal outcomes to one or more failure times (e.g., competing risks) in the same subject. We consider a case where longitudinal and survival outcomes are measured in subject pairs (e.g., married couples). In this dissertation, we propose a joint model incorporating within-pair correlations, both in the longitudinal and survival processes. We use a bivariate linear mixed-effects model for the longitudinal process, where the random effects are used to model the temporal correlation among longitudinal outcomes and the correlation between different outcomes. For the survival process, we incorporate a gamma frailty into a Weibull proportional hazards model to account for the correlation between survival times within pairs. The two sub-models are then linked through the shared random effects, where the longitudinal and survival processes are conditionally independent given the random effects. Parameter estimates are obtained by maximizing the joint likelihood for the bivariate longitudinal and bivariate survival data using the EM algorithm.

The proposed methodology is applied to the spouse data from the Cardiovascular Health Study (CHS) to investigate the association of both longitudinal depression scores and survival times between husbands and wives, and to quantify the association of mortality and longitudinal depression with other covariates in husbands and wives after accounting for the within-spouse correlation. **Public Heath Significance:** Spouse studies seek to reveal the

importance of both environmental and genetic influences on individuals. The analysis of such information is useful in assessing long term health effects in spouse pairs and/or individuals living together. The methodology we propose provides a valid statistical inference on the association of longitudinal measurements and the time-to-events among paired subjects. This methodology will contribute to the analysis of public health studies by ensuring that proper prediction and inference are made when pairs of individuals are measured longitudinally.

Keywords: Joint models, spouse pairs, bivariate longitudinal data, bivariate survival data, bivariate linear mixed-effects model, Weibull proportional hazards model with gamma frailty, depression, mortality.

TABLE OF CONTENTS

LIST OF TABLES

1.0 INTRODUCTION

Joint modeling of longitudinal and survival data has become a valuable tool for analyzing clinical trials data. The motivating idea behind this approach is to couple the survival model, which is of primary interest, with a suitable model for the repeated measurements of the endogenous outcome that will account for its special features. A well-known example where this methodology is used is in immunodeficiency virus (HIV) clinical trials where longitudinally measured immunologic and physiological status such as CD4 (a glycoprotein found on the surface of immune cells such as T helper cells) count and viral ribonucleic acid (RNA) copy number are considered as the predictors for time to progression to acquired immunodeficiency syndrome (AIDS) or death (Thiébaut *et al.*, 2005 [1]).

Most joint models developed so far have focused on relating a single or multiple longitudinal measurements to a time-to-event (Tsiatis *et al.*, 1995 [2]; Henderson *et al.*, 2000 [3]; Song *et al.*, 2002 [4]; Lin *et al.*, 2002 [5]; Guo and Carlin, 2004 [6]; Thiébaut *et al.*, 2005 [1]; Rizopoulos and Ghosh, 2011 [7]; Choi *et al.*, 2014 [8]), to multiple time-to-events, for example, competing risks (Elashoff, 2007 [9]), or recurrent events (Liu *et al.*, 2008 [10]; Liu and Huang, 2009 [11]). Furthermore, all the subjects in these models are assumed to be independent. However, this assumption is not adequate when the subjects are paired or related in some way. A prototype example is twins, who have the same genes and also share a similar childhood environment. Another example is married couples. The individuals in a married couple do not have common genetic traits like twins but they might 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 choses a smoker, they will both have the risk from the smoke, one as an active smoker, one as passive smoker. Furthermore, couples usually have

similar diets and living environments. When the subjects are no longer independent, a more complicated model is required to further account for the correlation between the subjects within a pair.

It is of interest to understand the lifespan and the mental health in such a paired population, especially in married couples. Ciocco [12] first reported high correlations between lifespan in married couples in 1940. A more recent refinement of assessing lifespan in married couples involves investigating the effects of bereavement on a widowed spouse. In general, studies report an increased risk of mortality in a widowed spouse, compared with non-widowed spouses (Kaprio *et al.*, 1987 [13]; Martikainen and Valkonen, 1996 [14]; Manor and Eisenbach, 2003 [15]). For example, Hart *et al.* (2007) [16] studied how loss of a spouse affects mortality risk in the widowed partner among married couples in Scotland. They found widowed participants were at higher risk than non-widowed participants of dying from any cause. Stahl *et al.* (2016) [17] assessed the associations among bereavement, cardiovascular disease (CVD), and depressive symptoms on mortality in older spouses from the large Cardiovascular Health Study (CHS) population-based cohort. They found the relation between bereavement and mortality was different in men and women and varied by CVD status. Bereavement decreased mortality in women with CVD and increased mortality in men without CVD. Significant relationships between spouses' psychological well-being have also been reported in many studies. The idea that the mood or affective state of one individual might influence that of another was first demonstrated by Coyne (1976) [18] in an experimental study. He found that participants who spoke with a depressed person over the telephone expressed similar feelings such as depression, anxiety, and hostility following the conversation. In contrast, those who spoke to a non-depressed person did not experience these negative emotions. Coyne *et al.* (1987) [19] later studied the psychological distress in a sample of adults (mostly spouses) who lived with a depressed patient. They found that those living with a person experiencing depressive symptoms were quite depressed themselves. Galbaud du Fort *et al.* (1994) [20] also found an significant relationship in symptoms of psychological distress and feelings of general well-being between the spouses in a large population study of Canadian adults. Tambs (1991) [21] examined similarities in psychological well-being in a large sample of Norwegian nuclear families and found a significant correlation in feelings of

anxiety and depression between husbands and wives in the entire sample as well as within different age groups. Regarding the interdependence involved in marriage, Kelley (1981) [22] reported that when one spouse experienced depressive symptoms, the other spouse's risk increased. Bookwala and Schulz (1996) [23] studied the spousal similarity of subjective well-being in a CHS sample of older adults. They found that one spouse's well-being and depression predicted the other's well-being even after controlling for known predictors of the these outcomes. Townsend *et al.* (2001) [24] investigated correlation between spouses' depressive symptomatology in middle-aged and older married couples in the Health and Retirement Study (HRS) and the Study of Asset and Health Dynamics Among the Oldest Old (AHEAD). They found that husbands' and wives' depressive symptoms were moderately correlated.

The proposed joint model developed in this dissertation was primarily motivated by the Cardiovascular Health Study (CHS) (Fried *et al.*, 1991 [25]). 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 [26]). Epidemiological studies have found that 10–20% of community-dwelling elderly persons report clinically significant depressive symptomatology (Blazer *et al.*, 1987 [27]; Murrell *et al.*, 1983 [28]; Kennedy *et al.*, 1989 [29]). 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 (Wulsin *et al.*, 1999 [30]; Schulz *et al.*; 2000 [31]; Schulz *et al.*; 2002 [32]). Methodological limitations may have contributed to inconsistent findings about the relationship between depression and mortality (Zhang *et al.*; 2009 [33]). First, most studies only have a one-time assessment of depression. However, depression status is a dynamic process and the change in severity of depression may have temporal effects on mortality, which cannot be captured by a single assessment. Second, most studies had relatively

short follow-up periods to ascertain mortality, which captured a small number of deaths and hence, produced potentially truncated and biased results. Therefore, it is necessary to examine the association between depression and mortality in a large community sample that utilizes longitudinal depression measures, while controlling for possible confounding factors. The traditional time-varying 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.

In this dissertation, we propose a joint modeling methodology for paired data where we extend the current joint models to take into account the within-pair correlation, both in the longitudinal and in the time-to-event processes. 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.

2.0 LITERATURE REVIEW

2.1 MODELS FOR BIVARIATE LONGITUDINAL DATA

Multivariate longitudinal data arise when a set of dependent outcomes is measured repeatedly over time. Bivariate longitudinal data refers to when paired outcomes are measured repeatedly. The primary example we consider in this dissertation involves depression scores of married couples in the CHS. There are two sources of correlations in such data: (1) serial correlation of repeated observations on any given response in a pair and (2) cross correlation of paired responses measured at a given time point. A popular approach to model such data is to use a bivariate random-effects model where a random effect is assumed for each outcome process and the two different processes are associated through a joint bivariate distribution on the random effects. Let $X_{ik}(t)$ be a measurement of the response of the kth subject in the ith pair at time $t \geq 0$, where $i = 1, \ldots, n$ and $k = 1, 2$. Assume each longitudinal process $X_{ik}(t)$ satisfies

$$\begin{aligned} X_{ik}(t) &= b_{0ik} + b_{1ik}t + \cdots + b_{(q_k-1)ik}t^{q_k-1} \\ &= \begin{bmatrix} 1 & t & \cdots & t^{q_k-1} \end{bmatrix} \begin{bmatrix} b_{0ik} \\ b_{1ik} \\ \vdots \\ b_{(q_k-1)ik} \end{bmatrix} \\ &= \mathbf{z}_{ik}(t)\boldsymbol{b}_{ik} \end{aligned} \tag{2.1}$$

where $\mathbf{z}_{ik}(t)$ is a $(1 \times q_k)$ vector of functions of time t, $\boldsymbol{b}_{ik}$ is a $(q_k \times 1)$ random effects, and $\mathbf{z}_{ik}(t)$ and $\boldsymbol{b}_{ik}$ may be different for each subject k. This allows flexibility in presenting the time trajectory of each response via polynomial. The longitudinal responses $X_{ik}(t)$ are not

observed directly; rather longitudinal measurements $y_{ik}(t)$ on the kth response are taken at times t, for each pair i, where

$$y_{ik}(t) = X_{ik}(t) + e_{ikj} \tag{2.2}$$

and $e_{ik} \sim N(0, \sigma_k^2)$ that reflect both biological variation and measurement error. Thus, together, Equations (2.1) and (2.2) contribute the bivariate random-effects model, and $X_{ik}(t)$ can be regarded as the "inherent" or "latent" trajectory for the response of subject k in pair i.

This approach has many advantages and it is applicable in a wide variety of situations. An earlier model proposed by Reinsel (1982 [34], 1984 [35]) 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 has been extended by Shah *et al.* (1997) [36] to accommodate the case of arbitrary measurement times. Their approach employed the EM algorithm for the parameter estimates. Schafer (1997) [37] and Schafer and Yucel (2002) [38] developed a similar model which allowed for multiple imputation in cases when there was missing data. Morrell *et al.* (2003) [39] used the multivariate linear mixed-effects model in a Bayesian framework to predict hypertension based on Body Mass Index, systolic blood pressure and triglyceride levels from the Baltimore longitudinal study of aging. Lin *et al.* (2002) [5], Thiébaut *et al.* (2005) [1], and Chi *et al.* (2006) [40] also employed this approach.

The approach described above can be used directly even when the underlying process is nonlinear in time. This is achieved via employing the use of polynomials in the inherent process. The newer approach is also applicable but must be modified somewhat when the outcome processes are nonlinear functions of parameters or when splines are used. Compared to polynomials, splines are usually preferred due to their local nature and better numerical properties (Ruppert *et al.*, 2003 [41]). For example, Song *et al.* (2002) [4] investigated formally whether a quadratic pattern of outcome processes provides a better characterization using a conditional F-test; Brown *et al.* (2005) [42] 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) [7] considered natural cubic splines. An alternative approach to model highly nonlinear shapes of subject-specific evolutions is to incorporate an additional stochastic term in the bivariate linear random-effects model to capture the remaining serial correlation in the observed measurements not captured by the random effects. The model is considered to be of the form

$$X_{ik}(t) = \mathbf{z}_{ik}(t)\boldsymbol{b}_{ik} + U_{ik}(t), \tag{2.3}$$

where $U_{ik}(t)$ is a mean-zero stochastic process, usually taken to be independent of $\boldsymbol{b}_{ik}$. Sy *et al.* (1997) [43] used the Fisher scoring to fit the model (2.3) and specified $U_{ik}(t)$ to be an integrated Orstein-Uhlenbeck (IOU) process which includes Brownian motion as a special limiting case. In many epidemiological studies, outcomes are related to covariates in a nonlinear fashion. Coull and Staudenmayer (2004) [44] presented a self-modeling regression model (SEMOR) (Lawton and Sylvestre, 1971 [45]) for flexible nonparametric modeling of multiple longitudinal outcomes, which allows for such nonlinear relationship. The bivariate random-effects model can be constructed joining different types of random-effects models such as a combination of a linear random-effects model for a continuous outcome and a generalized linear random-effects model for a binary outcome (Gueorguieva, 2001 [46]; Liu *et al.*, 2010 [47]; Choi *et al.*, 2014 [8]).

2.2 MODELS FOR BIVARIATE SURVIVAL DATA

The traditional techniques to analyze survival data are based on the assumption that the survival times of distinct individuals are independent of each other. However, this assumption is violated when the study units are paired such as child and parents, twins, or married couples. In the presence of the dependence between the event times, a joint survival model is considered.

Suppose there are two survival times, T_1 and T_2. Let $S_1(t_1)$ and $S_2(t_2)$ denote the survival function of T_1 and T_2, respectively. The general form of a joint survival function of two survival times can be written as

$$S(t_1, t_2) = Pr\{T_1 \geq t_1, T_2 \geq t_2\} \tag{2.4}$$

2.2.1 Bivariate Exponential and Weibull Model

The exponential distribution plays a prominent role in statistics. The reason is that it has a lot of interesting properties and it can be justified by many different mathematical constructions. Many authors have extended the univariate exponential distribution to the multidimensional case because the dependence structure can be addressed directly in the model. Gumbel (1960) [48] first proposed two bivariate distributions whose marginal distributions are exponential:

$$\begin{aligned} S(t_1, t_2) &= 1 - e^{-t_1} - e^{-t_2} + e^{-(t_1+t_2+\theta t_1 t_2)} \\ S(t_1, t_2) &= (1 - e^{-t_1})(1 - e^{-t_2})(1 + e^{-(t_1+t_2+\theta t_1 t_2)}). \end{aligned} \tag{2.5}$$

Freund (1961) [49] pointed out that Gumbel (1960) [48] did not discuss the appropriateness of these models to particular physical situations. Thus, he presented a different bivariate extension of the exponential distribution, which was particularly designed for the life testing of two-component systems that can function even after one of the components has failed. However, the margins in Freund's bivariate model are not exponential but mixtures of exponentials. Marshall and Olkin (1967) [50] considered a two-component system where the system survives or dies according to the occurrences of shocks to each or both of the components. The occurrences of shocks are governed by three independent Poisson processes, $N_1(t)$, $N_2(t)$ and $N_{12}(t)$ with intensities λ_1, λ_2 and λ_{12}. The first two processes, N_1 and N_2, control shocks to individual components, whereas N_{12} controls shocks to both components. The bivariate distribution they formulated is

$$S(t_1, t_2) = \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_{12} \max(t_1, t_2)\}. \tag{2.6}$$

This distribution is the most known bivariate exponential distribution (BVE). Block and Basu (1974) [51] presented an absolutely continuous bivariate extension of the exponential distribution, and Proschan and Sullo (1974) [52] investigated the parameter estimation for the three parameter version of bivariate exponential distribution proposed by Marshall and Olkin [50].

The exponential models can be generalized to the more flexible Weibull case where it allows for either proportional hazards or accelerated failure time models (Klein *et al.*, 1989 [53]; Ghosh and Gelfand, 1998 [54]). However, in the exponential case, there does not exist a unique natural extension of the univariate exponential distribution to bivariate or multivariate case. Thus, there are different versions of bivariate or multivariate Weibull distributions extended from the bivariate or multivariate exponential cases. For example, Hanagal developed various versions of multivariate Weibull distribution (MVW) and bivariate Weibull distribution (BVW). He first proposed a MVW which was the extension of the multivariate exponential of Marshall-Olkin (1967) (Hanagal, 1996 [55]) and a BVW by taking simple transformation to the bivariate exponential of Freund (Hanagal, 2005 [56]). He applied these models to censored samples with common covariates and derived maximum likelihood estimation (MLE) and a significance test for the covariates (Hanagal, 2004 [57], 2005 [56]). He later proposed another new bivariate Weibull regression model that was an extension of bivariate exponential of Proschan-Sullo (1974) with a frailty (a random effect shared by the dependent subjects and will be described in detail in Section 2.2.2) generated from several different distributions. This model was then applied to censored samples with covariates in order to investigate whether the frailty had no effect when the covariates were included in the model. The frailty distributions he considered include gamma, positive stable, and power variance functions (Hanagal, 2009a [58], 2009b [59]). In 2010 [60], he developed a BVW with a frailty generated by Weibull distribution and derived two-stage MLE procedure for the parameters.

2.2.2 Shared Frailty Model

Vaupel *et al.* (1979) [61] and Lancaster (1979) [62] first introduced the notation of frailty, a random effect, to represent the unobserved population heterogeneity. A frailty then was defined as an unobservable random effect shared by all subjects within a cluster. This type of model is becoming increasingly popular for modeling the association between individual survival times within clusters. In the bivariate case, the model assumes given the frailty μ, T_1 and T_2 are independent, that is,

$$S(t_1, t_2|\mu) = S_1(t_1|\mu)S_2(t_2|\mu). \tag{2.7}$$

This approach can be seen as a strategy to model the dependence in bivariate or multivariate data using univariate distributions (Sahu and Dey, 2000 [63]). A frailty is assumed to act either multiplicatively or additively on the hazard rates of all subjects within a cluster. In such models, clusters with large values of the frailty will experience the event at earlier times than clusters with small values. Hence, the larger the frailty value, the more "frail" a population tends to be.

Usually the frailty, μ, is assumed to act multiplicatively on proportional hazards, that is, $h(t|\mu) = \mu h_0(t)$, where $h_0(t)$ is the baseline hazard function. Under this assumption the cumulative hazard function is $H(t|\mu) = \mu H_0(t)$ and survival function is $S(t|\mu) = S_0(t)^\mu$ with corresponding baseline cumulative hazard function $H_0(t)$ and baseline survival function $S_0(t)$, respectively. Suppose that μ has a density $g(\mu)$. Then the joint survival function can be obtained by integrating out μ in the conditional bivariate survival function

$$\begin{aligned} S(t_1, t_2) &= \int_0^\infty S(t_1, t_2|\mu)g(\mu)\mathrm{d}\mu \\ &= \int_0^\infty S_{01}(t_1)^\mu S_{02}(t_2)^\mu g(\mu)\mathrm{d}\mu \\ &= \int_0^\infty e^{-\mu[H_{01}(t_1)+H_{02}(t_2)]}g(\mu)\mathrm{d}\mu \end{aligned} \tag{2.8}$$

Since the Laplace transform of a function, $f(t)$, can be written as $\mathcal{L}(s) = \int_0^\infty e^{-st}f(t)\mathrm{d}t$, Equation (2.8) can be recognized as the Laplace transform of $g(\mu)$ evaluated at $s = H_{01}(t_1) + H_{02}(t_2)$. Thus,

$$S(t_1, t_2) = \mathcal{L}_g\{H_{01}(t_1) + H_{02}(t_2))\} \tag{2.9}$$

There are some common distributions proposed in the literature for frailty. Gamma distributions have been extensively studied due to their simple interpretation and mathematical tractability. For example, they fit well to survival models due to the simplicity of the derivatives of the Laplace transform. Suppose frailty, μ, has a one-parameter gamma distribution with mean 1 and variance θ. Thus, μ has a density

$$g(\mu) = \frac{\mu^{\frac{1}{\theta}-1}\exp(-\frac{\mu}{\theta})}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}}. \tag{2.10}$$

The Laplace transform of $g(\mu)$ is $\mathcal{L}(s) = (1/\theta(1/\theta + s))^{1/\theta}$. A joint survival function can be obtained by using the result

$$S(t_1, t_2) = \left(\frac{1}{1 + \theta H_{01}(t_1) + \theta H_{02}(t_2)}\right)^{1/\theta} \tag{2.11}$$

The marginal survival function of T_1 is $S(t_1) = S(t_1, 0) = (\frac{1}{1+\theta H_{01}(t_1)})^{1/\theta}$. After doing some algebra, we can obtain $S(t_1)^{-\theta} - 1 = \theta H_{01}(t_1)$. Similarly, $S(t_2)^{-\theta} - 1 = \theta H_{02}(t_2)$. Plugging the results into Equation (2.11), we obtain the joint survival function as a function of the two marginal survival functions:

$$S(t_1, t_2) = (S(t_1)^{-\theta} + S(t_2)^{-\theta} - 1)^{-1/\theta} \tag{2.12}$$

As θ approaches 0, the marginal distributions $S(t_1)$ and $S(t_2)$ become independent. As θ approaches ∞, $S(t_1, t_2)$ approximates the Fréchet-Hoeffding bound on the maximum possible positive association between two distributions with given marginals. In other words, the model can range its association structure from independence to positive association but can't account for negative association. Clayton (1978) [64] proposed a continuous bivariate survival model where the conditional hazard for subject 1 at time t_1 given that the other 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,

$$\frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 \geq t_2)} = 1 + \Phi, \tag{2.13}$$

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) [65] considered the bivariate case without covariates and showed that θ (or Φ) is closely related to a measure of dependence, Kendall's τ, where $\tau = \frac{\theta}{\theta+2}$ 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) [66] 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) [67] generalized Clayton's model to the multivariate case and fitted the model using an accelerated EM algorithm. Hanagal also has published many works on the gamma shared frailty model (Hanagal, 2006 [68], 2007 [69]; Hanagal and Dabade, 2013 [70]; Hanagal and Pandey, 2014 [71]). In practice, however, the gamma frailty may not be suitable (Shih, 1998 [72]; Glidden, 1999 [73]; Fan *et al.*, 2000 [74]). Positive stable distributions (Hougaard, 1986a [75]; Fine *et al.*, 2003 [76]) and inverse Gaussian distributions (Hougaard, 1986b [77]; Whitmore and Lee, 1991 [78]) are the useful alternatives because they have an attractive feature that the hazard ratios decrease over time. However, the derivatives of their Laplace transforms are more complicated, which make some calculations more difficult. Laird (1978) [79], Heckman and Singer (1984) [80], and Guo and Rodriguez (1992) [67] considered estimating the distribution of the frailty by non-parametric maximum likelihood. That strategy leads to a finite mixture model where the frailty is assumed to have a discrete distribution. The nonparametric maximum likelihood approach has the advantage of flexibility, but it does not share some of the convenient properties of models with gamma frailties. Manda (2011) [81] considered a nonparametric frailty distribution modeled completely with a Dirichlet process prior.

The hazard ratio can also be estimated as a function of time. Many approaches have been developed to estimate a piecewise constant hazard ratio (Shih and Louis, 1995a [82], 1995b [83]; Nan *et al.*, 2006 [84]; Li *et al.*, 2008 [85]). Hu *et al.* (2011) [86] proposed a time-dependent hazard ratio motivated by Clayton's model and developed a pseudo-partial likelihood approach to estimate the hazard ratio that is a continuous function of the bivariate survival times.

Yashin *et al.* (1995) [87] noted that the shared frailty model does not satisfy some natural conditions (e.g., combining identical and fraternal twin data sets) and suggested a correlated individual frailty model for the analysis of twins data where the frailties of two twins in a

pair are not necessarily the same as they are in the shared frailty. Thus, the specific feature of the association of twins is captured by the correlation. Wienke *et al.* (2003) [88] 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.

2.2.3 Marginal and Copula Model

For bivariate survival data, a marginal model is useful when the main interest is to compare the survival times of individuals across the pairs and to estimate the effect of covariates on survival. Marginal approaches model the effect of covariates on the hazards of the individual events, taking into account the fact that observed event times are correlated but without explicitly modeling this correlation. Specifically, the correlation is often ignored when estimating the covariate effects and the uncertainty of the parameter estimates is evaluated by a "sandwich estimator" (Hardin and Hilbe, 2003 [89]) to ensure correct inference. It is closely related to the so-called generalized estimating equation (GEE) methodology (Liang and Zeger, 1986 [90]) and has mostly been considered in the context of proportional hazards models. For example, Lee *et al.* (1995) presented marginal Cox models and Wei *et al.* (1989) [91] considered the stratified Cox models.

When the correlation within pairs is also of interest, the marginal approach is not useful because it does not provide any information on the dependence between the failure times in pairs. Therefore, we need another approach to model the dependence structure, that is, the copula. The idea behind a copula approach is to study the dependence when the effect of the marginal distribution as function of time and covariates is removed. This removal is obtained by assuming the margins are known (fixed). Copula models combine the marginal model and the copula so that the margins are modeled by standard Cox models and the dependence is modeled by some sort of copula. The copulas used are those corresponding to the frailty models mentioned in Section 2.2.2. It is a multivariate approach that is more consistent with the univariate model than the frailty model. This approach gives us frailty models in simple cases. Copula models offer a way that the joint survival function of two

failure times in a pair is modeled as a function of the marginal survival functions through a copula. The copula, used to couple the marginal survivals functions and joint survival function, determines the type of dependence.

A two-stage estimation procedure is usually used in the copula models. In the first stage, the marginal parameters are estimated assuming independence. In the second stage, the estimated margins are used to estimate the association parameter in the copula by maximizing the likelihood with respect to the copula parameter with the estimated marginal parameters. This strategy was first suggested by Hougaard (1986a) [75] in the stable copula case. Shih and Louis (1995b) [83] investigated two-stage parametric and two-stage semi-parametric estimation procedures in the case where each margin is modeled separately. Glidden (1999) [73] studied the case where the margins are modeled by a stratified Cox model and the association parameter is modeled by the gamma copula. Anderson (2005) [92] generalized the approaches of Shih and Louis (1995b) [83] and Glidden (1999) [73] to allow one to estimate the association while modeling the effect of covariates in the marginal model.

2.3 JOINT MODELS FOR LONGITUDINAL AND SURVIVAL DATA

In longitudinal studies and clinical trials it is common to collect repeated measurements on a continuous or discrete response, times to events of interest and additional covariate information on each subject simultaneously. These longitudinal responses are measured with error and often incomplete after a long follow-up period. It is often of interest to investigate the association between the longitudinal response and the time-to-event in such studies. The joint modeling approach has been developed to handle this association.

2.3.1 Time-Dependent Cox Model

When the interest is on inference for the model parameters of a time-to-event process, a naïve method is to treat the longitudinal process as a time-varying covariate in a Cox model (Cox, 1972 [93]; Rizopoulos, 2010 [94]). The hazard function for individual i at time t is

then given by

$$h_i(t) = h_0(t) \exp\{\mathbf{w}_i\boldsymbol{\gamma} + \alpha y_i(t)\} \ , \tag{2.14}$$

where h_0 is the baseline hazard function, $y_i(t)$ is the observed longitudinal response with corresponding regression parameter α, and $\mathbf{w}_i$ is a vector of the fixed covariates with corresponding regression parameter vector $\boldsymbol{\gamma}$.

However, this Cox model approach assumes the covariates are external (not related to the failure mechanism) (Kalbfleisch and Prentice, 2002 [95]; Rizopoulos, 2010 [94]) and requires complete values of each covariate at each failure time for all individuals. Often, the longitudinal measurements do not satisfy these conditions due to the fact that they are collected with time-to-event on the same individual and are usually measured intermittently, with error, and might not be available at all the observed event times. Moreover, the Cox model is not able to take into account the measurement error of the covariates and thus can introduce bias.

2.3.2 Linear Mixed-Effects Model

The most commonly used sub-model to model the longitudinal process in the joint modeling is a linear mixed-effects model. This model incorporates random effects into a linear model to account for the within-individual correlation caused by the repeated measures of individuals collected over time. Let $y_i(t)$ be the longitudinal response of subject i at time t. One linear mixed-effects model for the observed response $y_i(t)$ is given by

$$y_i(t) = \mathbf{x}_i(t)\boldsymbol{\beta} + b_i(t) + \varepsilon_i(t) \ , \tag{2.15}$$

where $\mathbf{x}_i(t)$ is the design matrix for the fixed effects with corresponding regression parameters $\boldsymbol{\beta}$; $b_i(t)$ is the subject-specific random effects, typically given by $b_i(t) = b_{0i} + b_{1i}t$, where b_{0i} and b_{1i} represent the random intercept and slope effects, respectively (e.g., Self and Pawitan, 1992 [96]; Tsiatis *et al.*, 1995 [2]; Wulfsohn and Tsiatis, 1997 [97]). Also, $\varepsilon_i(t) \sim N(0, \sigma^2)$ is the measurement error, assumed to be independent of $b_i(t)$. Philipson *et al.* (2012) [98] considered a random intercept only form and a random quadratic form for $b_i(t)$.

When the subjects show highly nonlinear longitudinal trajectories over time, high-order polynomials or splines formulation of $b_i(t)$ is considered (Brown *et al.*, 2005 [42]; Ding and Wang, 2008 [99]; Rizopoulos *et al.*, 2009 [100]; Rizopoulos and Ghosh, 2011 [7]). $b_i(t)$ may also incorporate a stochastic component, which accounts for the within-individual biological fluctuations in the observed measurements not captured by the random effects (Henderson *et al.*, 2000 [3]; Wang and Taylor, 2001 [101]). By doing this, the model characterizes better the true overall longitudinal process. However, a stochastic component may cause computational issues because of the increase in model complexity. Hence, a random effects approach is relatively easier to implement in practice because it only requires an appropriate specification of the random effects, $b_i(t)$.

The linear mixed-effects model by itself, of course, ignores any association between the longitudinal and survival processes in the estimation of parameters. When there is an association between them, this approach can cause biased estimates due to ignoring possible informative censoring induced by the occurrence of an event (e.g., death) in the longitudinal process (Ratcliffe *et al.*, 2004 [102]; Ibrahim *et al.*, 2010 [103]; Sweeting and Thompson, 2011 [104]).

2.3.3 Two-Stage Approaches

Tsiatis *et al.* (1995) [2] proposed a two-stage procedure with the attempt of reducing the bias of the parameter estimates introduced by the use of the time-dependent Cox model that incorporates a longitudinal covariate measured with error. In the first stage, a linear mixed-effects model is fitted to estimate the true longitudinal process without measurement error. In the second stage, the estimates obtained by the linear mixed-effects model are plugged into a Cox proportional hazards model as covariates. However, this is not an unbiased approach because no survival information is utilized when estimating the true longitudinal response and possible selection bias due to informative dropout is not taken into account.

2.3.4 Joint Likelihood Approach

An increasingly popular alternative is the use of a joint likelihood approach, where the estimates of the parameters are obtained by maximizing the joint likelihood of the longitudinal process and the survival time. The joint model is comprised of two linked sub-models, one for the "true" longitudinal process and the other for the survival time along with additional specifications and assumptions that allow ultimately a full representation of the joint distribution of the observed data. This approach uses all available information optimally because both the longitudinal and survival data are utilized simultaneously. There are different ways to parameterize the joint likelihood of the longitudinal and survival processes, including selection models, pattern-mixture models, and random effects models (Diggle *et al.*, 2008 [105]). Sousa (2011) [106] gives a good overview on selection and pattern-mixture models and McCrink *et al.* (2013) [107] give a good overview on random effects models. We focus on the random effects models in this dissertation because this kind of model has been widely used in recent studies.

Joint random effects models are constructed by assuming conditional independence of survival and longitudinal data, given the shared, latent random effects (Schluchter, 1992 [108]; Faucett and Thomas, 1996 [109]; Wulfsohn and Tsiatis, 1997 [97]; Henderson *et al.*, 2000 [3]; Tsiatis and Davidian, 2004 [110]; Diggle *et al.*, 2008 [105]; Rizopoulos, 2010 [94]). They are also known as "shared parameter" models. Let Y denote the longitudinal response, T denote the failure time, and b denote the random variables shared by both Y and T and thus b accounts for the association between both outcomes. The joint distribution of Y and T takes the form

$$p(Y,T) = \int_b p(b)p(Y|b)p(T|b)\,\mathrm{d}b \tag{2.16}$$

Depending on the focus of the analysis, random effects joint models can be formulated to handle: (1) a survival process with longitudinal covariates that are measured with error; (2) a longitudinal process with informative censoring; or (3) the joint evolution of the longitudinal and survival processes. In conjunction with these different formulations of joint models, there are various types of joint models depending on the choice of sub-models used to link the two processes.

The most widely used random effects joint models combine a linear mixed-effects model for the longitudinal process and a Cox proportional hazards model for the survival process. The two sub-models are then linked through their shared random effects, which account for the association between both outcomes. The formulation of a link (shared random effects) between the longitudinal and survival processes depends on the focus on the joint model.

When the interest is on the survival process where we want to incorporate a time-dependent covariate measured with error into a survival (also the primary interest in this dissertation), the joint models can be formulated as follows. First, assume we know $m_i(t)$, the true and unobserved value of the longitudinal response for subject i at time t, and a hazard model can be defined:

$$h_i(t|M_i(t)) = h_0(t) \exp\left\{\mathbf{w}_i\boldsymbol{\gamma} + \alpha m_i(t)\right\} , \tag{2.17}$$

where $M_i(t) = \{m_i(s), 0 \leq s < t\}$ represents the longitudinal response history up to time t. Assume that the hazard is taken to depend linearly on longitudinal response history through the current value, $m_i(t)$, and α quantifies the effect of the true underlying longitudinal response, $m_i(t)$, on the hazard for an event. Also, $h_i(t)$ is the hazard function, $h_0(t)$ is the baseline hazard, $\mathbf{w}_i$ is a vector of the baseline covariates with corresponding vector of regression parameters, $\boldsymbol{\gamma}$. Next, a linear mixed model is used to obtain a less biased and thus more precise estimate of the longitudinal process, $m_i(t)$:

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) \\ &= \mathbf{x}_i(t)\boldsymbol{\beta} + b_i(t) + \varepsilon_i(t) \ , \end{aligned} \tag{2.18}$$

where $\mathbf{x}_i(t)$ is the design matrix for the fixed effects with corresponding regression parameter $\boldsymbol{\beta}$. $b_i(t)$ is the subject-specific random effects, and $\varepsilon_i(t) \sim N(0, \sigma^2)$ is the measurement error, assumed to be independent of $b_i(t)$. Finally, the two processes are linked through the random effects, $b_i(t)$, where the longitudinal and survival processes are conditionally independent given $b_i(t)$. This means that these random effects account for both the association between the longitudinal and survival processes, and the correlation between the repeated measurements in the longitudinal process (Tsiatis and Davidian, 2004 [110]; Rizopoulos, 2010 [94]).

When the interest is on the longitudinal process, where we want to take into account the informative censoring due to a time-to-event or when the focus is on both processes, where we want to make an inference regarding the relationship between the two processes, the joint model can be formulated as follows. First, a linear mixed model is used to obtain the estimate of the longitudinal process, $y_i(t)$.

$$y_i(t) = \mathbf{x}_i(t)\boldsymbol{\beta} + b_i(t) + \varepsilon_i(t) \ , \tag{2.19}$$

where $\mathbf{x}_i(t)$ is the design matrix for the fixed effects with corresponding regression parameters $\boldsymbol{\beta}$, $b_i(t)$ is the subject-specific random effects, and $\varepsilon_i(t) \sim N(0, \sigma^2)$ is the measurement error, assumed to be independent of $b_i(t)$. Next, a proportional hazards model is used for the survival data

$$h_i(t) = h_0(t) \exp\left\{\mathbf{w}_i\gamma + \delta b_i(t)\right\} \ , \tag{2.20}$$

where δ represents the influence of the longitudinal random effects on the survival process (Faucett and Thomas, 1996 [109]; Wulfsohn and Tsiatis, 1997 [97]; Diggle *et al.*, 2008 [105]). A separate association for the influence of the random effects on the survival process can also be assumed by allowing different coefficients for the random intercept and slope (e.g., $\delta_0 b_{0i} + \delta_1 b_{1i} t$). $b_i(t)$ can have other forms as was mentioned earlier in Section 2.3.2. Finally, assume that given the random effects, $b_i(t)$, the longitudinal process and the survival time are conditionally independent. The informative censoring can be accounted for through a joint likelihood, and thus reduces the bias in the estimates in the longitudinal sub-model.

Because the parameters that describe the longitudinal process and the parameters that describe the time-to-event as a function of the longitudinal process are estimated at the same time, the joint model uses both the observed longitudinal data and survival information to obtain estimates of the true longitudinal value at any time. Therefore, we are able to get more precise and accurate estimates of the strength of the relationship between the longitudinal process and the survival time.

2.3.5 Estimation of Joint Models

Maximum likelihood (ML) is the main estimation method used to estimate the parameters of joint models (Schluchter, 1992 [108]; Wulfsohn and Tsiatis, 1997 [97]; Henderson *et al.*, 2000 [3], 2002 [111]; Tsiatis and Davidian, 2004 [110]; Diggle *et al.*, 2008 [105]; Rizopoulos, 2010 [94]). This method involves maximizing the log likelihood of the joint distribution of the longitudinal and survival processes given the shared random effects that are assumed to link both processes. Let the observed data for each individual is $\{Y_i, T_i, \Delta_i\}$ where Y_i is the longitudinal response for subject i, T_i the event time, and Δ_i the event indicator ($\Delta_i = 1$ if the event occurs and $\Delta_i = 0$, otherwise). We do not observe the random effects b_i. The joint likelihood, $L(\boldsymbol{\phi})$ is given by

$$L(\boldsymbol{\phi}) = \prod_{i=1}^{n} \int p(Y_i|b_i, \boldsymbol{\phi}) p(T_i, \Delta_i|b_i, \boldsymbol{\phi}) p(b_i|\boldsymbol{\phi}) \, \mathrm{d}b_i \tag{2.21}$$

where $p(Y_i|b_i, \boldsymbol{\phi})$, $p(T_i, \Delta_i|b_i, \boldsymbol{\phi})$, and $p(b_i|\boldsymbol{\phi})$ are the densities of the longitudinal and survival processes, and random effects, respectively; $\boldsymbol{\phi} = (\boldsymbol{\phi}_y, \boldsymbol{\phi}_t, \boldsymbol{\phi}_b)^T$ denote the full parameter vector, with $\boldsymbol{\phi}_y$ denoting the parameters for the longitudinal process, $\boldsymbol{\phi}_t$ the parameters for the survival process, and $\boldsymbol{\phi}_b$ the parameters for the random effects covariance matrix.

The expectation–maximization (EM) algorithm is commonly used to maximize the joint log-likelihood function. This is done by iterating between an E-step, where the expected log-likelihood of the complete data conditional on the observed data and the current estimate of the parameters is computed, and an M-step, where new parameter estimates are computed by maximizing this expected log-likelihood. The EM algorithm has been traditionally preferred in the literature mainly due to the fact that in the E-step some of the parameters have closed-form update (Dempster *et al.*, 1977 [112]). However, a major disadvantage of using EM algorithm is its linear convergence rate that results in slow convergence especially near the maximum.

To improve the slow linear convergence rates when using the ML techniques, Bayesian estimation of joint models using Markov chain Monte Carlo (MCMC) techniques, first proposed by Faucett and Thomas (1996) [109], has been considered by many authors (Xu and Zeger, 2001a [113], 2001b [114]; Wang and Taylor, 2001 [101]; Song, Davidian, and Tsiatis.,

2002 [4]; Guo and Carlin, 2004 [6]). However, Bayesian approaches are based on prior distributions for the model parameters and thus the need to choose such prior distributions leads to the requirement of sensitivity analysis during model validation. In contrast, Tsiatis and Davidian (2001) [115] proposed an alternative estimation technique, the conditional score approach, that requires no assumption on the distribution of random effects, and they developed a set of unbiased estimating equations to determine the parameter estimates, which is less intensive than likelihood methods.

2.3.6 The Submodels for the Survival Data

A proportional hazards model is commonly used in joint models to represent the survival process, where the form of the baseline hazard is either parametric or unspecified. In the survival analysis context, the typical used parametric distributions for the baseline hazard include the Weibull, the log-normal (Schluchter, 1992 [108]), and the Gamma. However, it is customary in semi-parametric models to leave the baseline hazard unspecified in order to avoid the mis-specification for the distribution of survival time. Recently, however, it has been found that a completely unspecified baseline hazard can lead to an underestimation of the standard errors of the parameter estimates (Hsieh *et al.*, 2006 [116]) and that the piecewise-constant hazard (Brown and Ibrahim, 2003 [117]) can increase efficiency compared with an unspecified baseline (Slasor and Laird, 2003 [118]).

The proportional hazards model assumes not only that the covariates have a multiplicative effect on the hazard for an event and but also that the hazard of the event at a certain time only depends on the current value of the covariates and not on the history of the covariates. When these assumptions are violated, the joint models commonly incorporate the accelerated failure time model to model the survival process (Tseng *et al.*, 2005 [119]).

2.3.7 Joint Models for Multiple Longitudinal Outcomes

The extension of a joint model to handle more than one longitudinal outcome (either continuous or categorical) has been studied by several statisticians (Song *et al.*, 2002 [4]; Xu and Zeger, 2001b [114]; Lin *et al.*, 2002 [5]; Thiébaut *et al.*, 2005 [1]). The multivariate

random effects models as we mentioned previously in Section 2.1 are widely used to model the multiple longitudinal outcomes within the joint models. Estimation of the multivariate joint models proceeds by maximizing the corresponding log-likelihood function, which is very similar to the one presented in Equation 2.21. Assume we have the bivariate longitudinal outcome, and the only difference is the density function for the longitudinal part that under the bivariate model takes the form:

$$p(\boldsymbol{Y}_i|b_i,\boldsymbol{\phi}) = p(Y_i^1, Y_i^2|b_i,\boldsymbol{\phi}) = \Big\{\prod_{j=1}^{n_{i1}} f(Y_i^1|b_i,\boldsymbol{\phi})\Big\}\Big\{\prod_{j=1}^{n_{i2}} f(Y_i^2|b_i,\boldsymbol{\phi})\Big\} \tag{2.22}$$

where $\boldsymbol{Y}_i = \begin{bmatrix} Y_i^1 \\ Y_i^2 \end{bmatrix}$, Y_i^1 and Y_i^2 represent two longitudinal responses for subject i.

The main practical problem in fitting multivariate joint models is their computational complexity due to the requirement for the numerical integration with respect to the random effects. In particular, since we assume a different set of random effects per outcome, it is obvious that the dimensionality of the random effects vector is considerably increased with an increase in the number of longitudinal outcomes. When the subject-specific longitudinal profiles are highly nonlinear, the high-order polynomials or splines forms for the random effects have been considered by Rizopoulos and Ghosh (2011) [7], Chi and Ibrahim (2006) [40], Brown *et al.* (2005) [42], and Lin *et al.* (2002) [5].

3.0 DISSERTATION STATEMENT

The vast majority of the joint models developed so far in the literature have assumed that the experimental units are independent. However, it is of interest to jointly model survival data with longitudinal information in *paired* subjects, such as married couples. This latter focus will drive the primary aim of this dissertation.

There are three objectives in this dissertation. First, we develop a new joint modeling methodology for paired data where we extend the current joint models to take into account the within-pair correlation, both in the longitudinal process and in the time-to-event process. We use a bivariate linear mixed-effects model for the longitudinal process where the random effects are used to model the temporal correlation among continuous longitudinal outcomes and the correlation between different outcomes. For the survival process, we use a Weibull proportional hazards model with a gamma frailty to account for the correlation between survival times within pairs. The 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 by maximizing the joint likelihood for the bivariate longitudinal and bivariate survival data using the Expectation-Maximization (EM) algorithm. Second, we conduct simulations to evaluate the performance of the proposed model. Finally, we apply the proposed joint modeling approach to the CHS spouse sample with the following goals:

1. To investigate the association of both longitudinal depression score and mortality between husbands and wives, controlling for covariates associated with depression and mortality, respectively;
2. To quantify the husband- and wife- specific association of mortality with their longitudinal depression scores.

4.0 JOINT MODELING FRAMEWORK

4.1 THE BIVARIATE LONGITUDINAL PROCESS

When paired responses are measured over time, two levels of correlation structure need to be considered in the observations of a pair. The first level is the correlation over time for each response and the second level is the correlation between the two responses. We propose a bivariate linear mixed-effects model to explicitly model the two sources of correlations.

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 using a linear mixed-effects model:

$$\begin{aligned} y_{i1}(t) &= \mu_1(t) + v_{i1}(t) + \varepsilon_{i1}(t) \\ y_{i2}(t) &= \mu_2(t) + v_{i2}(t) + \varepsilon_{i2}(t) \ , \end{aligned} \tag{4.1}$$

where $\mu_1(t)$ and $\mu_2(t)$ refer to the average response trajectories, $v_{i1}(t)$ and $v_{i2}(t)$ are the random effects to capture the subject deviation from the average profile, and $\varepsilon_{i1}(t)$ and $\varepsilon_{i2}(t)$ are the measurement errors.

The average response trajectories can also be described by $\mu_1(t) = \mathbf{x}_{i1}(t)\boldsymbol{\beta}_1$ and $\mu_2(t) = \mathbf{x}_{i2}(t)\boldsymbol{\beta}_2$, in which $\mathbf{x}_{i1}(t)$ and $\mathbf{x}_{i2}(t)$ represent the covariates (which can be either time-invariant or time-varying) considered to be associated with the response, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are their corresponding regression parameters, respectively.

We assume the subject-specific random effects have a form of

$$\begin{aligned} v_{i1}(t) &= \mathbf{z}_{i1}(t)\boldsymbol{b}_{i1} = b_{0i1} + b_{1i1}t \\ v_{i2}(t) &= \mathbf{z}_{i2}(t)\boldsymbol{b}_{i2} = b_{0i2} + b_{1i2}t \ , \end{aligned} \tag{4.2}$$

where $\mathbf{z}_{i1}(t)$ and $\mathbf{z}_{i2}(t)$ are the design matrix of random effects with vector of corresponding regression parameters $\boldsymbol{b}_{i1}$ and $\boldsymbol{b}_{i2}$; b_{0i1} and b_{0i2} represent the random intercept effects, and b_{1i1} and b_{1i2} represent the random slope effects. Here, we allow both intercepts and slopes to be random.

Thus, the two linear mixed-effects models can be re-written as

$$\begin{aligned} y_{i1}(t) &= \mathbf{x}_{i1}(t)\boldsymbol{\beta}_1 + \mathbf{z}_{i1}(t)\boldsymbol{b}_{i1} + \varepsilon_{i1}(t) \\ y_{i2}(t) &= \mathbf{x}_{i2}(t)\boldsymbol{\beta}_2 + \mathbf{z}_{i2}(t)\boldsymbol{b}_{i2} + \varepsilon_{i2}(t) \end{aligned} \tag{4.3}$$

We propose a bivariate linear mixed-effects model to characterize the two responses jointly, where the two response trajectories are tied together through a joint distribution for the random effects

$$\begin{bmatrix} \boldsymbol{b}_{i1} \\ \boldsymbol{b}_{i2} \end{bmatrix} = \begin{bmatrix} b_{0i1} \\ b_{1i1} \\ b_{0i2} \\ b_{1i2} \end{bmatrix} \sim N(\mathbf{0}, \mathbf{D}) \ , \tag{4.4}$$

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

$$\mathbf{D} = \begin{bmatrix} \sigma^2_{b_{01}} & \sigma_{b_{01}b_{11}} & \sigma_{b_{01}b_{02}} & \sigma_{b_{01}b_{12}} \\ & \sigma^2_{b_{11}} & \sigma_{b_{11}b_{02}} & \sigma_{b_{11}b_{12}} \\ & & \sigma^2_{b_{02}} & \sigma_{b_{02}b_{12}} \\ & & & \sigma^2_{b_{12}} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_2 \end{bmatrix} \tag{4.5}$$

$\mathbf{D}$ 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 $\mathbf{D}_{12} = \mathbf{D}_{21}$ have all entries equal to 0, both responses are assumed to be completely independent at any time.

The two measurement errors are assumed to follow a joint distribution given by

$$\begin{bmatrix} \varepsilon_{i1}(t) \\ \varepsilon_{i2}(t) \end{bmatrix} \sim N(\mathbf{0}, \mathbf{R}) \ , \tag{4.6}$$

where $\mathbf{R}$, the covariance matrix of the measurement errors, is assumed 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. The structure of $\mathbf{R}$ characterizes the correlation between the two responses at the same time and is assumed to be the same across time and pairs. 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. We can combine the models above into one single bivariate longitudinal model:

$$\begin{aligned} y_{ik}(t) &= \mathbf{x}_{ik}(t)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(t)\boldsymbol{b}_{ik} + \varepsilon_{ik}(t) \\ &= m_{ik}(t) + \varepsilon_{ik}(t), \boldsymbol{b}_i \sim N(\mathbf{0}, \mathbf{D}), \boldsymbol{\varepsilon}_i(t) \sim N(\mathbf{0}, \mathbf{R}) \ , \end{aligned} \tag{4.7}$$

where $m_{ik}(t)$ denotes the *true* and *unobserved* value of longitudinal response; $\mathbf{x}_{ik}(t)$ and $\mathbf{z}_{ik}(t)$ are the design matrices of fixed and random effects, respectively, with vectors of corresponding fixed and random effects parameters $\boldsymbol{\beta}_k$ and $\boldsymbol{b}_{ik}$; $\varepsilon_{ik}(t)$ is the measurement error.

4.2 THE BIVARIATE SURVIVAL PROCESS

We propose a Weibull proportional hazards model with gamma frailty to jointly characterize the two event times. We assume that the event times are conditionally independent given the pair-specific random effect (the frailty), μ_i. The conditional hazard function at time t for subject k $(k = 1, 2)$ in pair i $(i = 1, \ldots, n)$ is as follows:

$$h_{ik}(t|\mu_i) = h_0(t)\mu_i \exp\left\{\mathbf{w}_{ik}^*\boldsymbol{\gamma}_k^* + \alpha_k m_{ik}(t)\right\} \ , \tag{4.8}$$

where $h_0(t)$ is the baseline hazard; μ_i is the frailty for pair i; $\mathbf{w}_{ik}^*$ is the baseline covariate vector considered 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.

Equation (4.8) 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 a flexible Weibull (parametric) distribution for the baseline hazard function because it is very general with respect to characterizing different shapes of the associated hazard function. We assume that the shape parameter of the Weibull baseline hazard is the same for the two subjects. Thus, the conditional hazard function can be written as

$$\begin{aligned} h_{ik}(t|\mu_i) &= \rho\lambda_k t^{\rho-1}\mu_i \exp\left\{\mathbf{w}_{ik}^*\boldsymbol{\gamma}_k^* + \alpha_k m_{ik}(t)\right\}, \rho > 0, \lambda_k > 0 \\ &= \rho t^{\rho-1}\mu_i \exp\left\{\log\lambda_k + \mathbf{w}_{ik}^*\boldsymbol{\gamma}_k^* + \alpha_k m_{ik}(t)\right\} \\ &= \rho t^{\rho-1}\mu_i \exp\left\{\mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(t)\right\}\ , \end{aligned} \tag{4.9}$$

where ρ is the Weibull shape parameter, λ_k is the Weibull scale parameter for subject k, and the intercept term in the baseline covariate vector ($\mathbf{w}_{ik}$) corresponds to $\log\lambda_k$.

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 $g(\mu_i) = \frac{\mu_i^{1/\theta-1}\exp(-\mu_i/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}}$. 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 $\tau = \frac{\theta}{\theta+2}$. When $\theta = 0$, both event times are assumed to be independent.

The cumulative conditional hazard is given by

$$\begin{aligned} H_{ik}(t|\mu_i) &= \int_0^t h_{ik}(s|\mu_i)\mathrm{d}s \\ &= \int_0^t \rho s^{\rho-1}\mu_i \exp\left\{\mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(s)\right\}\mathrm{d}s \end{aligned} \tag{4.10}$$

Thus, the joint survival function for pair i can be obtained by integrating out μ_i

$$\begin{aligned} S(t_{i1}, t_{i2}) &= \int_0^\infty S(t_{i1}, t_{i2}|\mu_i)g(\mu_i)\mathrm{d}\mu_i \\ &= \int_0^\infty \exp\Big[-H(t_{i1}, t_{i2}|\mu_i)\Big]g(\mu_i)\mathrm{d}\mu_i \\ &= \int_0^\infty \exp\Big\{-\Big[H_{i1}(t|\mu_i) + H_{i2}(t|\mu_i)\Big]\Big\}g(\mu_i)\mathrm{d}\mu_i \\ &= \int_0^\infty \exp\left\{-\sum_{k=1}^{2}\int_0^t \rho s^{\rho-1}\mu_i \exp\Big[\mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(s)\Big]\mathrm{d}s\right\}g(\mu_i)\mathrm{d}\mu_i \end{aligned} \tag{4.11}$$

4.3 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 for pair i, $\boldsymbol{T}_i$ the event time, and Δ_i the event indicator ($\boldsymbol{\Delta}_i = 1$ if the event occurs and $\boldsymbol{\Delta}_i = 0$, otherwise). Assume censoring is independent of the frailty, μ_i. As previously, we assume that given the shared random effects $\boldsymbol{b}_i$, the bivariate longitudinal response and the bivariate event time are conditionally 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

$$\begin{aligned} p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i | \boldsymbol{b}_i; \boldsymbol{\phi}) &= p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}) p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}), \quad \text{and} \\ p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}) &= \prod_{k=1}^{2} \prod_{j=1}^{n_{ik}} p[y_{ik}(t_{ikj}) | \boldsymbol{b}_i, \boldsymbol{\phi}] \ , \end{aligned} \tag{4.12}$$

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 process and bivariate longitudinal process, 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; k represents the kth subject in a pair ($k = 1, 2$); 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 bivariate longitudinal response and bivariate event time for pair i can be formulated as follows

$$\begin{aligned} \log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) &= \log \int p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i \\ &= \log \int p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) p(\boldsymbol{b}_i | \boldsymbol{\phi}_b) \mathrm{d}\boldsymbol{b}_i \\ &= \int \Big[\log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) + \log p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) + \log p(\boldsymbol{b}_i | \boldsymbol{\phi}_b) \Big] \mathrm{d}\boldsymbol{b}_i \end{aligned} \tag{4.13}$$

with the conditional log-density of the survival part given by

$$
\begin{aligned}
&\log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) \\
&= \log \int_0^\infty \prod_{k=1}^{2} \left\{ \rho {T_{ik}}^{\rho-1} \mu_i \exp[\mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(T_{ik})] \right\}^{\Delta_{ik}} \\
&\quad \exp\left\{ -\int_0^{T_{ik}} \rho s^{\rho-1} \mu_i \exp\left[\mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(s)\right] \mathrm{d}s \right\} \\
&\quad \times \frac{\mu_i^{1/\theta-1} \exp(-\mu_i/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}} \mathrm{d}\mu_i \\
&= D_i \log\theta + \log \frac{\Gamma(D_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} + \sum_{k=1}^{2} \Delta_{ik} \Big[\log\left(\rho T_{ik}^{\rho-1}\right) + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k m_{ik}(T_{ik}) \Big] \\
&\quad - (D_i + \frac{1}{\theta}) \log \left\{ 1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp[\alpha_k m_{ik}(s)] \mathrm{d}s \right\},
\end{aligned}
\tag{4.14}
$$

where $D_i = \sum_{k=1}^{2} \Delta_{ik}$. The joint log-density of the bivariate longitudinal response together with the random effects takes the form

$$
\begin{aligned}
&\log \Big[p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) p(\boldsymbol{b}_i | \boldsymbol{\phi}_b) \Big] \\
&= \sum_{k=1}^{2} \left[-\frac{n_{ik}}{2} \Big(\log 2\pi + \log |\mathbf{R}| \Big) \right] - \frac{\Big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\boldsymbol{b}_i\Big)^T \mathbf{R}^{*-1} \Big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{b}_i\Big)}{2} \\
&\quad - \frac{q_b}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{D}| - \frac{\boldsymbol{b}_i^T \mathbf{D}^{-1} \boldsymbol{b}_i}{2},
\end{aligned}
\tag{4.15}
$$

where n_{ik} is the number of repeated measurements of subject k in pair i; $\mathbf{R}^*$ is a $(n_{i1} + n_{i2})$ $\times$ $(n_{i1} + n_{i2})$ dimensional square diagonal matrix with the elements consisting of variance of measurement error of the corresponding response; q_b denotes the dimensionality of the random-effects vector, and other quantities are as defined earlier.

Finally, the log-likelihood of bivariate longitudinal and bivariate survival data can be formulated as follows:

$$
\begin{aligned}
\ell(\boldsymbol{\phi}) &= \sum_{i=1}^{n} \log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \\
&= \sum_{i=1}^{n} \log \int p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \log \int p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) p(\boldsymbol{b}_i | \boldsymbol{\phi}_b) \mathrm{d}\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \int \Big[\log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) + \log p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}_y) + \log p(\boldsymbol{b}_i | \boldsymbol{\phi}_b) \Big] \mathrm{d}\boldsymbol{b}_i \\
&= \sum_{i=1}^{n} \int \Bigg\{ D_i \log \theta + \log \frac{\Gamma(D_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} + \sum_{k=1}^{2} \Delta_{ik} \Big[\log (\rho T_{ik}^{\rho-1}) + \mathbf{w}_{ik} \boldsymbol{\gamma}_k + \alpha_k m_{ik}(T_{ik}) \Big] \\
&\quad - (D_i + \frac{1}{\theta}) \log \Bigg\{ 1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik} \boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k m_{ik}(s)\right] \mathrm{d}s \Bigg\} \\
&\quad + \sum_{k=1}^{2} \Bigg[- \frac{n_{ik}}{2} \Big(\log 2\pi + \log |\mathbf{R}| \Big) \Bigg] - \frac{\Big(\boldsymbol{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k - \mathbf{Z}_i \boldsymbol{b}_i\Big)^T \mathbf{R}^{*-1} \Big(\boldsymbol{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k - \mathbf{Z}_i \boldsymbol{b}_i\Big)}{2} \\
&\quad - \frac{q_b}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{D}| - \frac{\boldsymbol{b}_i^T \mathbf{D}^{-1} \boldsymbol{b}_i}{2} \Bigg\} \mathrm{d}\boldsymbol{b}_i.
\end{aligned}
\tag{4.16}
$$

4.4 PARAMETER ESTIMATION USING THE EM ALGORITHM

In the joint modeling literature, the EM algorithm has been traditionally preferred (treating the random effects as 'missing data'), mainly due to the fact that in the M-step some of the parameters have closed-form updates. In this dissertation 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. This was done by iterating between the following steps until convergence.

1. An E-step, where we computed the expected log-likelihood (4.16) of the complete data conditional on the observed data and the current estimate of the parameters. The complete data for each individual were $(\boldsymbol{y}_{ik}, \mathbf{X}_{ik}, T_{ik}, \Delta_{ik}, \mathbf{w}_{ik}, \boldsymbol{\phi})$. All components except $\boldsymbol{\phi}$ were observed, and

2. An M-step, where new parameter estimates were computed by maximizing this expected joint log-likelihood.

The E-step for joint models

The joint models we proposed in Sections 4.1 and 4.2 are as follows:

$$\begin{aligned} h_{ik}(t) &= \rho t^{\rho-1}\mu_i \exp\Big\{\mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k[\mathbf{x}_{ik}(t)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(t)\boldsymbol{b}_{ik}]\Big\} \\ \mu_i &\sim\ GAM(\frac{1}{\theta},\theta), \\ y_{ik}(t) &= \mathbf{x}_{ik}(t)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(t)\boldsymbol{b}_{ik} + \varepsilon_{ik}(t) \\ \boldsymbol{b}_i &\sim N(\mathbf{0},\mathbf{D}), \quad \boldsymbol{\varepsilon}_i(t) \sim N(\mathbf{0},\mathbf{R}), \end{aligned} \tag{4.17}$$

$\boldsymbol{\phi} = (\boldsymbol{\phi}_y^T, \boldsymbol{\phi}_t^T, \boldsymbol{\phi}_b^T)^T$ with $\boldsymbol{\phi}_y = (\boldsymbol{\beta}_k, \sigma_1^2, \sigma_2^2)$, $\boldsymbol{\phi}_t = (\theta, \rho, \boldsymbol{\gamma}_k, \alpha_k)$ ($\boldsymbol{\gamma}_k$ contains $\log\lambda_k$), $\boldsymbol{\phi}_b = vec(\mathbf{D})$. 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.

$$\begin{aligned} &\varrho(\boldsymbol{\phi}|\boldsymbol{\phi}^{(it)}) \\ &= \sum_{i=1}^{n}\int \log p(\boldsymbol{T}_i,\boldsymbol{\Delta}_i,\boldsymbol{y}_i,\boldsymbol{b}_i;\boldsymbol{\phi})p(\boldsymbol{b}_i|\boldsymbol{T}_i,\boldsymbol{\Delta}_i,\boldsymbol{y}_i;\boldsymbol{\phi}^{(it)})\mathrm{d}\boldsymbol{b}_i \\ &= \sum_{i=1}^{n}\int \Big[\log p(\boldsymbol{T}_i,\boldsymbol{\Delta}_i|\boldsymbol{b}_i,\boldsymbol{\phi}_t,\boldsymbol{\beta}) + \log p(\boldsymbol{y}_i|\boldsymbol{b}_i,\boldsymbol{\phi}_y) + \log p(\boldsymbol{b}_i|\boldsymbol{\phi}_b)\Big]p(\boldsymbol{b}_i|\boldsymbol{T}_i,\boldsymbol{\Delta}_i,\boldsymbol{y}_i;\boldsymbol{\phi}^{(it)})\mathrm{d}\boldsymbol{b}_i \\ &= \sum_{i=1}^{n}\int \Bigg\{D_i\log\theta + \log\frac{\Gamma(D_i+\frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} + \sum_{k=1}^{2}\Delta_{ik}\Big[\log(\rho T_{ik}^{\rho-1}) + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k\{\mathbf{x}_{ik}(T_{ik})\boldsymbol{\beta}_k + \mathbf{z}_{ik}(T_{ik})\boldsymbol{b}_{ik}\}\Big] \\ &\quad - (D_i+\frac{1}{\theta})\log\Bigg\{1+\theta\sum_{k=1}^{2}\exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k)\int_0^{T_{ik}}\rho s^{\rho-1}\exp\Big[\alpha_k\{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\Big]\mathrm{d}s\Bigg\} \\ &\quad + \sum_{k=1}^{2}\Bigg[-\frac{n_{ik}}{2}\Big(\log 2\pi + \log|\mathbf{R}|\Big)\Bigg] - \frac{\Big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\boldsymbol{b}_i\Big)^T\mathbf{R}^{*-1}\Big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\boldsymbol{b}_i\Big)}{2} \\ &\quad - \frac{q_b}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{D}| - \frac{\boldsymbol{b}_i^T\mathbf{D}^{-1}\boldsymbol{b}_i}{2}\Bigg\}p(\boldsymbol{b}_i|\boldsymbol{T}_i,\boldsymbol{\Delta}_i,\boldsymbol{y}_i;\boldsymbol{\phi}^{(it)})\mathrm{d}\boldsymbol{b}_i. \end{aligned} \tag{4.18}$$

A key computational difficulty in the calculation of the log-likelihood above is that the integral with respect to time in the survival function, as well as the integral with respect to random effects, do not have closed-form solutions. Thus, it requires numerical approaches to approximate these integrals. We employed the 7-point or 15-point Gauss-Kronrod quadrature rule (Press *et al.*, 2007 [120]) to approximate the one-dimensional integral with respect to time in the survival function. Under Gauss-Kronrod rule, the integrals of any function $f(s)$ can be approximated by a weighted sum of integrand evaluations at m Gaussian quadrature points as follows

$$\int_0^T f(s)\mathrm{d}\boldsymbol{s} \;=\; \int_{-1}^{1} f\left(\frac{T}{2}x+\frac{T}{2}\right)\frac{T}{2}\,\mathrm{d}\boldsymbol{x} \;\approx\; \frac{T}{2}\sum_{i=1}^{m} f\left(\frac{T}{2}t_i+\frac{T}{2}\right)\pi_i, \tag{4.19}$$

where π_i and t_i are the weights and points at which to evaluate the function $f(s)$. Using this approach, the integral with respect to time in the log-likelihood (4.18) was approximated by

$$\begin{aligned}&\int_0^{T_{ik}} \rho s^{\rho-1}\exp\left[\alpha_k\{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k+\mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right]\mathrm{d}s\\ &\approx \frac{T_{ik}}{2}\sum_{g=1}^{m}\pi_g\rho\, t_g^{\rho-1}\;\exp\left[\alpha_k\{\mathbf{x}_{ik}(t_g)\boldsymbol{\beta}_k+\mathbf{z}_{ik}(t_g)\boldsymbol{b}_{ik}\}\right],\end{aligned} \tag{4.20}$$

where π_g and t_g^* are the weights and points for a m-point Gauss-Kronrod quadrature rule and $t_g=\frac{T_{ik}}{2}t_g^*+\frac{T_{ik}}{2}$.

However, the integral with respect to random effects is computationally demanding to approximate as its dimensionality increases. In order to decrease the computational burden to some degree, we extended the pseudo-adaptive Gauss-Hermite rules proposed by Rizopoulos (2012) [121] to approximate the integral with respect to random effects. Rizopoulos' approach was used to approximate the integral of subject-specific random effects across individuals. However, in this dissertation 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 idea behind this approach is to first fit the bivariate mixed-effects model for the longitudinal outcome and extract information regarding the location and scale of the posterior distribution of the random effects given the bivariate longitudinal responses for each pair. This information is

then used to approximately rescale the subject-specific integrals in the log-likelihood of joint models. We first fitted the bivariate linear mixed model, then extracted the Bayes estimates of the random effects, $\widetilde{\boldsymbol{b}}_i = \arg\max_{b_i}\{\log p(\boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\phi}_y)\}$, and their covariance matrix $\widetilde{\boldsymbol{H}}_i^{-1}$ with $\widetilde{\boldsymbol{H}}_i$ given by

$$
\begin{aligned}
\widetilde{\boldsymbol{H}}_i &= -\frac{\partial^2}{\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^T} \log p(\boldsymbol{y}_i, \boldsymbol{b}_i; \widetilde{\boldsymbol{\phi}}_y)|_{\boldsymbol{b}_i = \widetilde{\boldsymbol{b}}_i} \\
&= -\frac{\partial^2}{\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^T} \log \Big[p(\boldsymbol{y}_i | \boldsymbol{b}_i, \widetilde{\boldsymbol{\phi}}_y) p(\boldsymbol{b}_i | \widetilde{\boldsymbol{\phi}}_b) \Big] \\
&= -\frac{\partial^2}{\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^T} \Bigg\{ \sum_{k=1}^{2} \Big[-\frac{n_{ik}}{2} \Big(\log 2\pi + \log |\widetilde{\mathbf{R}}| \Big) \Big] - \frac{\Big(\boldsymbol{y}_i - \mathbf{X}_i \widetilde{\boldsymbol{\beta}}_k - \mathbf{Z}_i \widetilde{\boldsymbol{b}}_i \Big)^T \widetilde{\mathbf{R}}^{*-1} \Big(\boldsymbol{y}_i - \mathbf{X}_i \widetilde{\boldsymbol{\beta}}_k - \mathbf{Z}_i \widetilde{\boldsymbol{b}}_i \Big)}{2} \\
&\quad - \frac{q_b}{2} \log 2\pi - \frac{1}{2} \log |\widetilde{\mathbf{D}}| - \frac{\widetilde{\boldsymbol{b}}_i^T \widetilde{\mathbf{D}}^{-1} \widetilde{\boldsymbol{b}}_i}{2} \Bigg\} \\
&= \frac{1}{2} \frac{\partial^2}{\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^T} \Big[(\mathbf{Z}_i \widetilde{\boldsymbol{b}}_i)^T \widetilde{\mathbf{R}}^{*-1} (\mathbf{Z}_i \widetilde{\boldsymbol{b}}_i) + \widetilde{\boldsymbol{b}}_i^T \widetilde{\mathbf{D}}^{-1} \widetilde{\boldsymbol{b}}_i \Big] \\
&= \mathbf{Z}_i^T \widetilde{\mathbf{R}}^{*-1} \mathbf{Z}_i + \widetilde{\mathbf{D}}^{-1},
\end{aligned}
\tag{4.21}
$$

where $\widetilde{\boldsymbol{\phi}}_y$ and $\widetilde{\boldsymbol{\phi}}_b$ are the maximum likelihood estimates from the bivariate linear mixed-effects model. Under the pseudo-adaptive Gauss-Hermite rules, and for any form $A(\cdot)$ function of random effects, the integral in the definition of the log-likelihood was approximated by a weighted sum of integrand evaluations at pre-specified points as follows:

$$
\begin{aligned}
E\{A(\boldsymbol{\phi}, \boldsymbol{b}_i) | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}\} &= \int A(\boldsymbol{\phi}, \boldsymbol{b}_i) p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i \\
&\approx 2^{q_b/2} |\widetilde{\boldsymbol{B}}_i|^{-1} \sum_{t_1 \ldots t_{q_b}} \boldsymbol{\pi}_t A(\boldsymbol{\phi}, \widetilde{\boldsymbol{r}}_t) p(\widetilde{\boldsymbol{r}}_t | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \exp\left(\|\boldsymbol{b}_t\|^2 \right),
\end{aligned}
\tag{4.22}
$$

where q_b denotes the dimension of the random-effects vector; $\sum_{t_1 \cdots t_{q_b}}$ is used as shorthand for $\sum_{t_1=1}^{K} \ldots \sum_{t_{q_b}=1}^{K}$ with K denoting the number of Gauss-Hermite quadrature points; $\widetilde{\boldsymbol{r}}_t = \widetilde{\boldsymbol{b}}_i + \sqrt{2} \widetilde{\boldsymbol{B}}_i^{-1} \boldsymbol{b}_t$ with $\widetilde{\boldsymbol{B}}_i$ denoting the Choleski factor of $\widetilde{\boldsymbol{H}}_i$ and $\boldsymbol{b}_t^T = (b_{t_1}, \ldots, b_{t_{q_b}})$ the Gauss-Hermite quadrature points with corresponding weights $\boldsymbol{\pi}_t$; $\|\boldsymbol{b}_t\| = \{ \sum_{t_1 \ldots t_{q_b}} b_{t_1} \cdots b_{t_{q_b}} \}^{1/2}$. However, we implemented this procedure only once, at the beginning of the optimization, and we did not further update the quadrature points forwards.

Using both the Gauss-Kronrod quadrature rule and the pseudo-adaptive Gauss-Hermite rules, the complete data log-likelihood (4.18) was computed as

$$
\begin{aligned}
&\varrho(\boldsymbol{\phi}|\boldsymbol{\phi}^{(it)}) \\
&= \sum_{i=1}^{n} \int \Big[\log p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i|\boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) + \log p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\phi}_y) + \log p(\boldsymbol{b}_i|\boldsymbol{\phi}_b) \Big] p(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i \\
&\approx \sum_{i=1}^{n} \Bigg\{ 2^{q_b/2} |\widetilde{\boldsymbol{B}}_i|^{-1} \sum_{t_1 \cdots t_{q_b}} \Bigg(D_i \log\theta + \log \frac{\Gamma(D_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})} \\
&\quad + \sum_{k=1}^{2} \Delta_{ik} \Big[\log\left(\rho T_{ik}^{\rho-1}\right) + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k \{\mathbf{x}_{ik}(T_{ik})\boldsymbol{\beta}_k + \mathbf{z}_{ik}(T_{ik})\widetilde{\boldsymbol{r}}_t\} \Big] \\
&\quad - (D_i + \frac{1}{\theta}) \log \Bigg[1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \frac{T_{ik}}{2} \Big(\sum_{g=1}^{m} \pi_g \rho\, t_g^{\rho-1} \exp\Big[\alpha_k \{\mathbf{x}_{ik}(t_g)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(t_g)\widetilde{\boldsymbol{r}}_t\} \Big] \Big) \Bigg] \\
&\quad + \sum_{k=1}^{2} \Big[-\frac{n_{ik}}{2} \Big(\log 2\pi + \log|\mathbf{R}| \Big) \Big] - \frac{\Big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\widetilde{\boldsymbol{r}}_t\Big)^T \mathbf{R}^{*-1} \Big(\boldsymbol{y}_i - \mathbf{X}_i\boldsymbol{\beta}_k - \mathbf{Z}_i\widetilde{\boldsymbol{r}}_t\Big)}{2} \\
&\quad - \frac{q_b}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{D}| - \frac{\widetilde{\boldsymbol{r}}_t^T \mathbf{D}^{-1} \widetilde{\boldsymbol{r}}_t}{2} \Bigg) p(\widetilde{\boldsymbol{r}}_t|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \pi_t \exp\left(\|\boldsymbol{b}_t\|^2\right) \Bigg\},
\end{aligned}
\tag{4.23}
$$

where $\widetilde{\boldsymbol{r}}_t$ is pair-specific and thus all subjects in the same pair had an identical $\widetilde{\boldsymbol{r}}_t$.

The posterior distribution of random effects is written as

$$
p(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) = \frac{p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i|\boldsymbol{b}_i, \boldsymbol{\phi}_t, \boldsymbol{\beta}) p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\phi}_y) p(\boldsymbol{b}_i|\boldsymbol{\phi}_b)}{p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi})} \tag{4.24}
$$

with mean, $\widetilde{\boldsymbol{b}}_i = E(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) = \int \boldsymbol{b}_i\, p(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i$, and variance, $\widetilde{v\boldsymbol{b}}_i = Var(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) = \int (\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i)^T (\boldsymbol{b}_i - \widetilde{\boldsymbol{b}}_i)\, p(\boldsymbol{b}_i|\boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i$.

The M-step for joint models

In the M-step, we updated the parameters by

$$
\phi^{(it+1)} = \arg\max_{\phi} \varrho(\boldsymbol{\phi}|\boldsymbol{\phi}^{(it)}) \tag{4.25}
$$

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 $\varrho(\boldsymbol{\phi}|\boldsymbol{\phi}^{(it)})$ 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 solution and were updated using the expressions

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{\sigma_1}^2 & 0 \\ 0 & \hat{\sigma_2}^2 \end{bmatrix} \quad \text{with}$$
$$\hat{\sigma_k}^2 = N_k^{-1} \sum_{i=1}^{n} \Big(\boldsymbol{y}_{ik} - \mathbf{X}_{ik}\boldsymbol{\beta}_k)^T \Big(\boldsymbol{y}_{ik} - \mathbf{X}_{ik}\boldsymbol{\beta}_k - 2\mathbf{Z}_{ik}\widetilde{\boldsymbol{b}}_{ik}\Big) + tr(\mathbf{Z}_{ik}^T\mathbf{Z}_{ik}\widetilde{v\boldsymbol{b}}_{ik}) + \widetilde{\boldsymbol{b}}_{ik}^T\mathbf{Z}_{ik}^T\mathbf{Z}_{ik}\widetilde{\boldsymbol{b}}_{ik}, \tag{4.26}$$

where $N_k = \sum_i n_{ik}$. Also,

$$\begin{aligned} \hat{\mathbf{D}} &= n^{-1} \sum_{i=1}^{n} \int (\boldsymbol{b}_i - \bar{\boldsymbol{b}})^T (\boldsymbol{b}_i - \bar{\boldsymbol{b}}) p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_{ik}, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i \\ &= n^{-1} \sum_{i=1}^{n} \int \boldsymbol{b}_i^T \boldsymbol{b}_i p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_{ik}, \boldsymbol{y}_i; \boldsymbol{\phi}^{(it)}) \mathrm{d}\boldsymbol{b}_i \\ &= n^{-1} \sum_{i=1}^{n} \{\widetilde{v\boldsymbol{b}}_i - \widetilde{\boldsymbol{b}}_i^T \widetilde{\boldsymbol{b}}_i\} \end{aligned} \tag{4.27}$$

The fixed effects $\boldsymbol{\beta}$ and the parameters of the survival submodel $\boldsymbol{\phi}_t$ do not have a closed-form solution and thus, we implemented the Newton-Raphson approach to update these parameters in the M-step as follows:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{(it+1)} &= \hat{\boldsymbol{\beta}}^{(it)} - \Big\{\frac{\partial S(\hat{\boldsymbol{\beta}}^{(it)})}{\partial \boldsymbol{\beta}}\Big\}^{-1} S(\hat{\boldsymbol{\beta}}^{(it)}), \\ \hat{\boldsymbol{\phi}}_t^{(it+1)} &= \hat{\boldsymbol{\phi}}_t^{(it)} - \Big\{\frac{\partial S(\hat{\boldsymbol{\phi}}_t^{(it)})}{\partial \boldsymbol{\phi}_t}\Big\}^{-1} S(\hat{\boldsymbol{\phi}}_t^{(it)}), \end{aligned} \tag{4.28}$$

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, receptively; $\partial S(\hat{\boldsymbol{\beta}}^{(it)})/\partial\boldsymbol{\beta}$ 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 $\boldsymbol{\phi}_t$ have the form

$$
\begin{aligned}
&S(\boldsymbol{\beta}_k) \\
&= \sum_{i=1}^{n} \Bigg\{ \mathbf{X}_i^T \mathbf{R}^{*-1} (\boldsymbol{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k - \mathbf{Z}_i \widetilde{\boldsymbol{b}}_i) + \sum_{k=1}^{2} \boldsymbol{\Delta}_{ik} \alpha_k \mathbf{x}_{ik}(T_{ik}) \\
&\quad - \int \frac{(D_i\theta + 1) \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \alpha_k \mathbf{x}_{ik}(s) \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s}{1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s} \Bigg\} \\
&\quad p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i
\end{aligned}
\tag{4.29}
$$

$$
\begin{aligned}
&S(\theta) \\
&= \sum_{i=1}^{n} \int \Bigg\{ D_i \theta^{-1} - I(D_i > 0) \sum_{l=0}^{D_i - 1} (\theta + l\theta^2)^{-1} \\
&\quad + \theta^{-2} \log\left(1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s\right) \\
&\quad - \frac{(D_i + \theta^{-1}) \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s}{1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s} \Bigg\} p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i
\end{aligned}
\tag{4.30}
$$

$$
\begin{aligned}
&S(\rho) \\
&= \sum_{i=1}^{n} \int \Bigg\{ \sum_{k=1}^{2} \boldsymbol{\Delta}_{ik} (\rho^{-1} + \log T_{ik}) \\
&\quad - \frac{(D_i\theta + 1) \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} (1 + \rho \log s) s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s}{1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s} \Bigg\} \\
&\quad p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i
\end{aligned}
\tag{4.31}
$$

$$
\begin{aligned}
&S(\boldsymbol{\gamma}_k) \\
&= \sum_{i=1}^{n} \int \Bigg\{ \sum_{k=1}^{2} \boldsymbol{\Delta}_{ik} \mathbf{w}_{ik} \\
&\quad - \frac{(D_i\theta + 1) \sum_{k=1}^{2} \mathbf{w}_{ik} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s}{1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s} \Bigg\} p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i
\end{aligned}
\tag{4.32}
$$

$$
\begin{aligned}
&S(\boldsymbol{\alpha}_k) \\
&= \sum_{i=1}^{n} \int \Bigg\{ \boldsymbol{\Delta}_{ik} \{\mathbf{x}_{ik}(T_{ik})\boldsymbol{\beta}_k + \mathbf{z}_{ik}(T_{ik})\boldsymbol{b}_{ik}\} \\
&\quad - \frac{(D_i\theta + 1) \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s}{1 + \theta \sum_{k=1}^{2} \exp(\mathbf{w}_{ik}\boldsymbol{\gamma}_k) \int_0^{T_{ik}} \rho s^{\rho-1} \exp\left[\alpha_k \{\mathbf{x}_{ik}(s)\boldsymbol{\beta}_k + \mathbf{z}_{ik}(s)\boldsymbol{b}_{ik}\}\right] \mathrm{d}s} \Bigg\} \\
&\quad p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \mathrm{d}\boldsymbol{b}_i
\end{aligned}
\tag{4.33}
$$

Both the Gauss-Kronrod quadrature (4.19) and pseudo-adaptive Gauss-Hermite (4.22) rules were implemented to approximate the score functions above.

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

$$
S(\hat{\boldsymbol{\phi}}) = \sum_{i=1}^{n} \int \left\{ \frac{\partial}{\partial \boldsymbol{\phi}} \log \left[p(\boldsymbol{T}_i, \boldsymbol{\Delta}_i | \boldsymbol{b}_i, \boldsymbol{\phi}) p(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{\phi}) p(\boldsymbol{b}_i | \boldsymbol{\phi}) \right] \right\} p(\boldsymbol{b}_i | \boldsymbol{T}_i, \boldsymbol{\Delta}_i, \boldsymbol{y}_i; \boldsymbol{\phi}) \, \mathrm{d}\boldsymbol{b}_i \mid_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}} \tag{4.34}
$$

and the standard errors were calculated from

$$
\widehat{var}(\hat{\boldsymbol{\phi}}) = -\{H(\hat{\boldsymbol{\phi}})\}^{-1}, \quad \text{with} \quad H(\hat{\boldsymbol{\phi}}) = \sum_{i=1}^{n} \left. \frac{\partial S_i(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} \right|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}} \tag{4.35}
$$

5.0 SIMULATION STUDIES

5.1 DESCRIPTION OF SIMULATION DATA

In this section, we present a simulation study to evaluate the performance of our proposed model. We consider three sets of simulations 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 considered several sets of association parameters to assess how well our model estimates the effect of longitudinal measurement on the risk of event. We generated data with 600 female-male pairs as follows.

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 model with random intercept

$$\begin{aligned} y_{ikj} &= \beta_{0k} + \beta_{1k} t_{ikj} + b_{0ik} + \varepsilon_{ikj}, \\ \boldsymbol{b_{0i}} &= \begin{pmatrix} b_{0i1} \\ b_{0i2} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{D}),\ \boldsymbol{\varepsilon_i} = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{R}), \end{aligned} \tag{5.1}$$

where the fixed intercept and slope, $\begin{pmatrix} \beta_{01} \\ \beta_{11} \end{pmatrix} = \begin{pmatrix} 2 \\ 0.2 \end{pmatrix}$ and $\begin{pmatrix} \beta_{02} \\ \beta_{12} \end{pmatrix} = \begin{pmatrix} 1 \\ 0.1 \end{pmatrix}$, represent the average trend of longitudinal measurements over time in females and males, respectively; $\mathbf{D} = \begin{pmatrix} \sigma^2_{b_1} & \sigma_{a_1 b_1} \\ \sigma_{a_1 b_1} & \sigma^2_{b_2} \end{pmatrix} = \begin{pmatrix} 0.7 & 0.2 \\ 0.2 & 0.6 \end{pmatrix}$; we set $\sigma_{a_1 b_1} = 0.2$ to simulate a low correlation between the two longitudinal measurements (i.e, $r = \frac{0.2}{\sqrt{0.6 \times 0.7}} = 0.3$); $\mathbf{R} = \begin{pmatrix} \sigma^2_1 & 0 \\ 0 & \sigma^2_2 \end{pmatrix} = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.6 \end{pmatrix}$, assuming the two measurements were independent conditional on the random effects.

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 gamma frailty given by

$$\begin{aligned} h_{ik}(t) &= \rho t^{\rho-1}\mu_i \exp\{\log\lambda_k + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k(\beta_{0k} + \beta_{1k}t + b_{0ik})\}, \\ \mu_i &\sim GAM\left(\frac{1}{\theta}, \theta\right), \end{aligned} \tag{5.2}$$

where $\lambda_k \rho t^{\rho-1}$ is the Weibull baseline hazard of subject k. We set $\rho = 6$ and $(\lambda_1,\ \lambda_2) = (0.1,\ 0.2)$ 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. The frailty μ_i was generated from a gamma distribution with mean 1 and variance $\theta = 0.5$. We set $\theta = 0.5$ to simulate a low overall dependence between the two survival times (i.e., Kendall's $\tau = \frac{0.5}{0.5+2} = 0.2$). Dichotomous covariate diseases $(w_{i1},\ w_{i2})$ were both generated from a binomial distribution with probability 0.5. The disease effect parameters $(\gamma_1,\ \gamma_2)$ were set to $(0.7,\ 0.6)$ with corresponding hazard ratios (HR) = (2.01, 1.82) of disease vs. no disease in females and males, respectively.

We considered four sets of association parameters: (1) $(\alpha_1,\ \alpha_2) = (0.1,\ 0.05)$ represents a small effect of longitudinal measurement on survival time for both genders (i.e., $(\mathrm{HR}_1,\ \mathrm{HR}_2)$ = (1.11, 1.05), per unit increase in longitudinal measurement); (2) $(\alpha_1,\ \alpha_2) = (0.5,\ 0.3)$ represents a moderate effect for both genders (i.e., $(\mathrm{HR}_1,\ \mathrm{HR}_2) = (1.65,\ 1.35)$); (3) $(\alpha_1,\ \alpha_2)$ = (1.0, 0.8) represents a large effect for both genders (i.e., $(\mathrm{HR}_1,\ \mathrm{HR}_2) = (2.72,\ 2.23)$); (4) $(\alpha_1,\ \alpha_2) = (1.0,\ 0.05)$ represents a large effect for females but a small effect for males.

The survival function is given by

$$\begin{aligned} S_{ik}(t) &= \exp\left\{-\int_0^t h_{ik}(s)\,\mathrm{d}s\right\} \\ &= \exp\left\{-\int_0^t \rho s^{\rho-1}\mu_i \exp\left[\log\lambda_k + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k(\beta_{0k} + \beta_{1k}s + b_{0ik})\right]\mathrm{d}s\right\} \end{aligned} \tag{5.3}$$

and it follows a uniform distribution on the interval from 0 to 1

$$S_{ik}(t) \sim U(0,1) = U \tag{5.4}$$

Using Equations (5.3) and (5.4), we generated event times by randomly generating $S_{ik}(t)$ then solving for t using the uniroot() and integrate() functions in R [122]:

$$\begin{aligned} U = S_{ik}(t) &= \exp\left\{-\int_0^t \rho s^{\rho-1}\mu_i \exp\left[\log\lambda_k + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k(\beta_{0k} + \beta_{1k}s + b_{0ik})\right] \mathrm{d}s\right\} \\ G(t) &= \log U + \int_0^t \rho s^{\rho-1}\mu_i \exp\left[\log\lambda_k + \mathbf{w}_{ik}\boldsymbol{\gamma}_k + \alpha_k(\beta_{0k} + \beta_{1k}s + b_{0ik})\right] \mathrm{d}s = 0 \end{aligned} \tag{5.5}$$

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 $\sigma_{a_1b_1} = 0.4$ 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 $\sigma_{a_1b_1} = 0.5$ ($r = 0.8$), $\theta = 6$ (Kendall's $\tau = 0.8$), $\rho = 40$, and $(\lambda_1,\ \lambda_2) = (1.2,\ 1.2)$. Three sets of the association parameters were considered: $(\alpha_1,\ \alpha_2) = (0.1,\ 0.05)$, $(0.5,\ 0.3)$, and $(1.0,\ 0.8)$.

True parameter values used in the simulation studies are presented in Table 5.1. For each scenario, 1000 replications were conducted. We calculated the mean bias in the estimates, the mean standard error of the estimates (SE), the mean squared error (MSE), and the coverage probability (CP) of the estimated 95% confidence intervals to evaluate the model performance.

Table 5.1: True parameter values used in the three simulations.

		Simulation 1				Simulation 2			Simulation 3		
		True				True			True		
Submodel	Parameter	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Longitudinal	β_{01}	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	β_{02}	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	β_{11}	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	β_{12}	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	σ_1^2	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	σ_2^2	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	$\sigma_{b_1}^2$	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	$\sigma_{b_1 b_2}$	0.2	0.2	0.2	0.2	0.4	0.4	0.4	0.5	0.5	0.5
	$\sigma_{b_2}^2$	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Survival	θ	0.5	0.5	0.5	0.5	2.0	2.0	2.0	6.0	6.0	6.0
	ρ	6.0	6.0	6.0	6.0	12.0	12.0	12.0	40	40	40
	λ_1	0.1	0.1	0.1	0.1	1.0	1.0	1.0	1.2	1.2	1.2
	λ_2	0.2	0.2	0.2	0.2	1.0	1.0	1.0	1.2	1.2	1.2
	γ_1	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	γ_2	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	α_1	0.1	0.5	1.0	1.0	0.1	0.5	1.0	0.1	0.5	1.0
	α_2	0.05	0.3	0.8	0.05	0.05	0.3	0.8	0.05	0.3	0.8

5.2 SIMULATION RESULTS

The results of three simulation studies are summarized in Tables 5.2, 5.3, and 5.4, respectively. The results indicate that overall our model performs well under all simulated circumstances from three perspectives. First, the biases of the estimates are all minimal. Secondly, the mean standard errors of the estimates are in good agreement with the mean squared errors. Finally, the coverage probabilities are close to or achieve the nominal level. However, comparing the association parameter estimates (α_1, α_2) across all scenarios, relatively larger biases of (α_1, α_2) (i.e., true vs. estimate(bias) = 1.0 vs. 0.911(-0.089) was observed in females and 0.8 vs. 0.758(-0.042) in males) in the scenario 10 (Table 5.4, S10) indicate that our model may slightly underestimate the association between the longitudinal and survival outcomes when the true association is high with high dependence in both bivariate longitudinal and bivariate survival outcomes.

Table 5.2: Results of Simulation 1. Low dependence on both bivariate longitudinal and bivariate survival outcomes with 20% censoring rate and 1000 replications.

	S1						S2					
Parameter	True	Est.	Bias	SE	MSE	CP	True	Est.	Bias	SE	MSE	CP
β_{01}	2.0	2.000	0.000	0.044	0.045	0.949	2.0	1.998	-0.002	0.045	0.045	0.948
β_{02}	1.0	0.999	-0.001	0.042	0.043	0.950	1.0	0.999	-0.001	0.043	0.042	0.949
β_{11}	0.2	0.199	-0.001	0.041	0.041	0.949	0.2	0.195	-0.005	0.049	0.051	0.943
β_{12}	0.1	0.099	-0.001	0.046	0.045	0.953	0.1	0.098	-0.002	0.048	0.047	0.959
σ_1	0.77	0.774	-0.001	0.016	0.016	0.948	0.77	0.774	0.000	0.018	0.018	0.948
σ_2	0.77	0.774	-0.001	0.017	0.017	0.949	0.77	0.773	-0.001	0.018	0.018	0.949
$\sigma^2_{b_1}$	0.7	0.701	0.001	0.046	0.053	0.911	0.7	0.699	-0.001	0.048	0.055	0.915
$\sigma_{b_1b_2}$	0.2	0.200	0.000	0.040	0.036	0.975	0.2	0.200	0.000	0.041	0.038	0.970
$\sigma^2_{b_2}$	0.6	0.599	-0.001	0.046	0.048	0.946	0.6	0.597	-0.003	0.046	0.049	0.935
θ	0.5	0.515	0.015	0.070	0.072	0.948	0.5	0.517	0.017	0.071	0.074	0.948
ρ	6.0	6.113	0.113	0.211	0.234	0.933	6.0	6.114	0.114	0.214	0.244	0.925
$\log\lambda_1$	-2.30	-2.393	-0.090	0.213	0.232	0.935	-2.30	-2.400	-0.097	0.226	0.244	0.941
$\log\lambda_2$	-1.61	-1.691	-0.081	0.140	0.164	0.915	-1.61	-1.691	-0.082	0.144	0.169	0.913
γ_1	0.7	0.711	0.011	0.122	0.122	0.953	0.7	0.709	0.009	0.124	0.120	0.954
γ_2	0.6	0.608	0.008	0.121	0.123	0.944	0.6	0.607	0.007	0.122	0.124	0.950
α_1	0.1	0.099	-0.001	0.082	0.084	0.943	0.5	0.512	0.012	0.089	0.088	0.950
α_2	0.05	0.052	0.002	0.091	0.091	0.946	0.3	0.308	0.008	0.094	0.095	0.950
	S3						S4					
Parameter	True	Est.	Bias	SE	MSE	CP	True	Est.	Bias	SE	MSE	CP
β_{01}	2.0	1.998	-0.002	0.045	0.046	0.948	2.0	1.999	-0.001	0.045	0.045	0.938
β_{02}	1.0	0.998	-0.002	0.043	0.043	0.952	1.0	0.998	-0.002	0.042	0.043	0.950
β_{11}	0.2	0.190	-0.010	0.061	0.064	0.939	0.2	0.188	-0.012	0.061	0.064	0.936
β_{12}	0.1	0.097	-0.003	0.053	0.054	0.947	0.1	0.101	0.001	0.046	0.045	0.954
σ_1	0.77	0.773	-0.002	0.020	0.020	0.953	0.77	0.773	-0.002	0.020	0.020	0.957
σ_2	0.77	0.774	-0.001	0.019	0.019	0.949	0.77	0.773	-0.001	0.017	0.018	0.944
$\sigma^2_{b_1}$	0.7	0.699	-0.001	0.050	0.059	0.914	0.7	0.701	0.001	0.050	0.058	0.917
$\sigma_{b_1b_2}$	0.2	0.202	0.002	0.041	0.038	0.972	0.2	0.201	0.001	0.041	0.037	0.975
$\sigma^2_{b_2}$	0.6	0.598	-0.002	0.048	0.049	0.929	0.6	0.600	-0.000	0.046	0.047	0.943
θ	0.5	0.515	0.015	0.075	0.077	0.949	0.5	0.509	0.009	0.073	0.075	0.947
ρ	6.0	6.125	0.125	0.230	0.255	0.934	6.0	6.100	0.100	0.222	0.228	0.951
$\log\lambda_1$	-2.30	-2.396	-0.093	0.254	0.254	0.963	-2.30	-2.351	-0.049	0.248	0.212	0.983
$\log\lambda_2$	-1.61	-1.698	-0.089	0.158	0.179	0.933	-1.61	-1.687	-0.077	0.142	0.161	0.936
γ_1	0.7	0.708	0.008	0.130	0.124	0.955	0.7	0.706	0.006	0.129	0.129	0.948
γ_2	0.6	0.610	0.010	0.127	0.129	0.950	0.6	0.608	0.008	0.122	0.121	0.950
α_1	1.0	1.021	0.021	0.109	0.103	0.962	1.0	1.002	0.002	0.106	0.084	0.986
α_2	0.8	0.825	0.025	0.108	0.109	0.953	0.01	0.054	0.004	0.091	0.090	0.951

Table 5.3: Results of Simulation 2. Moderate dependence on both bivariate longitudinal and bivariate survival outcomes with 20% censoring rate and 1000 replications.

	S5						S6						S7					
Parameter	True	Est.	Bias	SE	MSE	CP	True	Est.	Bias	SE	MSE	CP	True	Est.	Bias	SE	MSE	CP
β_{01}	2.0	2.000	-0.000	0.045	0.046	0.945	2.0	1.997	-0.003	0.045	0.045	0.949	2.0	1.999	-0.001	0.045	0.045	0.957
β_{02}	1.0	0.999	-0.001	0.043	0.042	0.954	1.0	0.998	-0.002	0.043	0.043	0.949	1.0	0.997	-0.003	0.043	0.043	0.955
β_{11}	0.2	0.201	0.001	0.056	0.056	0.955	0.2	0.200	0.000	0.062	0.063	0.952	0.2	0.194	-0.006	0.069	0.071	0.948
β_{12}	0.1	0.100	0.000	0.054	0.054	0.951	0.1	0.100	0.000	0.056	0.056	0.947	0.1	0.098	-0.002	0.059	0.059	0.955
σ_1	0.77	0.774	-0.001	0.019	0.019	0.948	0.77	0.774	0.000	0.020	0.020	0.958	0.77	0.773	-0.001	0.021	0.021	0.958
σ_2	0.77	0.774	0.000	0.019	0.019	0.949	0.77	0.775	0.000	0.019	0.018	0.960	0.77	0.773	-0.001	0.020	0.019	0.957
$\sigma_{b_1}^2$	0.7	0.698	-0.002	0.049	0.058	0.886	0.7	0.699	-0.001	0.050	0.059	0.903	0.7	0.701	0.001	0.051	0.059	0.910
$\sigma_{b_1 b_2}$	0.4	0.398	-0.002	0.045	0.041	0.968	0.4	0.398	-0.002	0.045	0.041	0.967	0.4	0.399	-0.001	0.045	0.042	0.967
$\sigma_{b_2}^2$	0.6	0.597	-0.003	0.050	0.050	0.950	0.6	0.597	-0.003	0.051	0.049	0.961	0.6	0.596	-0.004	0.051	0.052	0.952
θ	2.0	2.032	0.032	0.161	0.157	0.958	2.0	2.045	0.045	0.163	0.168	0.949	2.0	2.035	0.035	0.167	0.168	0.951
ρ	12.0	12.121	0.121	0.450	0.451	0.943	12.0	12.145	0.145	0.458	0.466	0.950	12.0	12.114	0.114	0.484	0.470	0.961
$\log \lambda_1$	0.0	-0.041	-0.041	0.279	0.284	0.950	0.0	-0.040	-0.040	0.286	0.302	0.950	0.0	0.003	0.003	0.303	0.281	0.972
$\log \lambda_2$	0.0	-0.036	-0.036	0.186	0.188	0.954	0.0	-0.035	-0.035	0.188	0.195	0.951	0.0	-0.042	-0.042	0.198	0.200	0.946
γ_1	0.7	0.713	0.013	0.149	0.150	0.950	0.7	0.708	0.008	0.150	0.152	0.952	0.7	0.702	0.002	0.155	0.153	0.956
γ_2	0.6	0.606	0.006	0.148	0.150	0.942	0.6	0.613	0.013	0.149	0.150	0.949	0.6	0.611	0.011	0.153	0.155	0.944
α_1	0.1	0.102	0.002	0.114	0.115	0.945	0.5	0.509	0.009	0.120	0.124	0.942	1.0	0.991	-0.009	0.138	0.122	0.963
α_2	0.05	0.048	-0.002	0.125	0.125	0.946	0.3	0.302	0.002	0.128	0.129	0.950	0.8	0.807	0.007	0.142	0.139	0.951

Table 5.4: Results of Simulation 3. High dependence on both bivariate longitudinal and bivariate survival outcomes with 20% censoring rate and 1000 replications.

	S5						S6						S7					
Parameter	True	Est.	Bias	SE	MSE	CP	True	Est.	Bias	SE	MSE	CP	True	Est.	Bias	SE	MSE	CP
β_{01}	2.0	2.001	0.001	0.045	0.044	0.962	2.0	2.000	0.000	0.045	0.045	0.950	2.0	2.000	0.000	0.045	0.045	0.946
β_{02}	1.0	1.002	0.002	0.043	0.044	0.938	1.0	1.000	0.000	0.043	0.045	0.934	1.0	1.000	0.000	0.043	0.044	0.948
β_{11}	0.2	0.200	0.000	0.052	0.051	0.953	0.2	0.199	-0.001	0.054	0.056	0.955	0.2	0.200	0.000	0.056	0.055	0.950
β_{12}	0.1	0.098	-0.002	0.051	0.052	0.948	0.1	0.099	-0.001	0.051	0.053	0.939	0.1	0.099	-0.001	0.053	0.053	0.952
σ_1	0.77	0.773	-0.001	0.019	0.018	0.959	0.77	0.773	-0.001	0.019	0.019	0.956	0.77	0.771	-0.004	0.020	0.020	0.950
σ_2	0.77	0.774	0.000	0.018	0.018	0.948	0.77	0.774	-0.001	0.019	0.018	0.957	0.77	0.773	-0.002	0.019	0.019	0.952
$\sigma^2_{b_1}$	0.7	0.701	0.001	0.048	0.056	0.903	0.7	0.699	-0.001	0.049	0.055	0.913	0.7	0.708	0.008	0.049	0.053	0.932
$\sigma_{b_1 b_2}$	0.5	0.499	-0.001	0.047	0.042	0.968	0.5	0.499	-0.001	0.048	0.042	0.972	0.5	0.501	0.001	0.048	0.042	0.971
$\sigma^2_{b_2}$	0.6	0.599	-0.001	0.053	0.050	0.971	0.6	0.599	-0.001	0.053	0.050	0.967	0.6	0.600	0.000	0.053	0.048	0.974
θ	6.0	6.091	0.091	0.403	0.404	0.959	6.0	6.099	0.099	0.406	0.409	0.959	6.0	6.042	0.042	0.410	0.410	0.950
ρ	40.0	40.262	0.262	1.609	1.634	0.953	40.0	40.337	0.337	1.633	1.626	0.957	40.0	40.051	0.051	1.685	1.649	0.959
$\log \lambda_1$	0.18	0.175	-0.008	0.403	0.405	0.950	0.18	0.168	-0.015	0.407	0.408	0.957	0.18	0.357	0.175	0.415	0.392	0.960
$\log \lambda_2$	0.18	0.166	-0.016	0.267	0.273	0.946	0.18	0.163	-0.020	0.268	0.267	0.953	0.18	0.203	0.021	0.275	0.276	0.951
γ_1	0.7	0.704	0.004	0.167	0.163	0.954	0.7	0.696	-0.004	0.168	0.166	0.945	0.7	0.700	0.000	0.171	0.169	0.951
γ_2	0.6	0.607	0.007	0.166	0.169	0.946	0.6	0.608	0.008	0.167	0.166	0.945	0.6	0.598	-0.002	0.170	0.173	0.952
α_1	0.1	0.097	-0.003	0.163	0.166	0.950	0.5	0.508	0.008	0.170	0.176	0.954	1.0	0.911	-0.089	0.183	0.163	0.957
α_2	0.05	0.047	-0.003	0.178	0.179	0.950	0.3	0.305	0.005	0.182	0.182	0.952	0.8	0.758	-0.042	0.195	0.190	0.952

6.0 APPLICATION: SPOUSE PAIR DATA FROM THE CARDIOVASCULAR HEALTH STUDY (CHS)

In this chapter, 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.

6.1 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. Potential participants were excluded if they were wheelchair bound in the home, unable to participate in the examination at the field centers, or under active treatment for cancer. 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) [25] and Tell *et al.* (1993) [123]. 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.

6.2 MEASURES

Extensive demographic and health information was collected by trained interviewers or by clinical examinations.

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. Level of depressive symptoms was evaluated using the previously validated 10-item Center for Epidemiological Studies Depression Scale (CES-D) (Radloff, 1977 [124]; Andresen *et al.*, 1994 [125]) 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.

Other measures — *Sociodemographic* variables included: (a) age at entry into the CHS cohort; (b) race, coded as white or non-white (primarily African American due to the small number identifying as other racial groups); (c) education, coded as the highest grade or year of school ever completed; (d) stressful life events (total of 10 possible stressful life events in past 6 months); and (e) annual income. *Health behavior* included smoking status (never, former, or current), alcohol consumption (drinks per week), and body mass index (BMI), calculated as weight in kilograms divided by height in meters squared. *Difficulty with Activities of Daily Living (ADL)/Instrumental Activities of Daily (IADL)*: This variable is used to assess functional disability. It was coded 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". It was coded no if the participants reported no difficulty across the 12 activities. *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 [126]). *Caregiving status*: Caregiving status is a known risk factor for depression (Schulz *et al.*, 2005 [127]). Participants were asked a single yes/no question at baseline concerning whether they provided help to anyone with things like shopping, filling out forms, doing repairs, providing child care, etc. *CVD* was evaluated at CHS entry by clinical and laboratory examination. 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. The methods used to determine clinical and subclinical diseases in the CHS have been previously published (Psaty *et al.*, 1995 [128], Kuller *et al.*, 1995 [129]). *Anti-depressant medication use*: As part of an annual clinical assessment, participants brought their prescription medication containers to the clinic, where interviewers transcribed the drug name, strength, and dosing instructions from the medication labels. The participants were then asked how many doses of each medication they actually took within the past 2 weeks. 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).

6.3 STATISTICAL ANALYSIS

Longitudinal submodel

To satisfy the normality assumption for longitudinal CES-D score, we took the square root transformation of CES-D scores. We conducted bivariate linear mixed-effects models to identify baseline factors related to longitudinal CES-D score. 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 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 ($P<.05$). Race was included in the multivariable models to account for two cohorts combined where cohort 1 was almost white (97%) and cohort 2 was black. To model the trajectory of CES-D scores, linear and quadratic functions were considered depending on which function fits the trajectory better. Different random-effects structures (i.e., random intercept, random slope, random intercept and random slope) were also tested. The best structure was chosen using information criteria (e.g., AIC and BIC).

Survival submodel

Similarly, we fitted 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. 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 ($P<.05$).

Joint models

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 gamma frailty. Then the adjusted joint models were built, controlling for the covariates independently associated with longitudinal CES-D score in the bivariate linear mixed-effects model as well as the covariates independently associated with mortality in the Weibull proportional hazards model with gamma frailty. The marginal correlation between husbands' and wives' CES-D score was calculated and the dependence of husbands' and wives' mortality was measured by Kendall's τ.

Because the CHS includes two cohorts with different follow-up lengths and the CES-D score in each cohort was assessed up to half of its entire follow-up, we truncated the longitudinal CES-D score and the mortality at 6 years in both cohorts. In addition, we repeated the analyses on two cohorts combined with different follow-up lengths (18-year mortality and 10-year CES-D score in cohort 1; 15-year mortality and 6-year CES-D score in cohort 2) and on each cohort as supplementary analyses.

To understand the impact of ignoring the correlation of CES-D score and the dependence of mortality between husbands and wives on the model parameter estimates, we also fitted data using a joint model where the correlations were not considered. For example, linear mixed-effects models and Weibull proportional hazards models were used in the joint models.

6.4 RESULTS

Participant characteristics

Baseline characteristics of the analysis sample stratified by sex are presented in Table 6.1. Comparing to husbands, wives were younger (median age 70 vs. 73 years) and had higher 3MS score (29 vs. 28) and CES-D score (4 vs. 3) , and a higher proportion were non-smokers (59.4% vs. 32.3%), with ADL/IADL difficulty (30.2% vs. 21.1%), and on antidepressant medication (5.1% vs. 2.5%), while a lower proportion had prevalent clinical CVD (17.4% vs. 32.1%) and subclinical CVD (56.5% vs. 70.7%) (P for all$<$0.05).

Table 6.1: Baseline characteristics of husbands and wives (n=1330 spousal pairs).

	Wives(N=1330)[a]	Husbands (N=1330)[a]	*P*
Age, y			<.0001
Median (25th, 75th percentile)	70 (68,74)	73 (70,77)	
Range	65-91	65-93	
White, No. (%)	1229 (92.4)	1225 (92.1)	0.77
Education, y			0.10
Median (25th, 75th percentile)	12 (12,18)	12 (11,20)	
Range	1-21	0-21	
Income, No. (%)			0.99
<$5,000	13 (1.1)	13 (1.0)	
$5,000 - $7,999	42 (3.5)	36 (2.9)	
$8,000 - $11,999	91 (7.5)	96 (7.6)	
$12,000 - $15,999	186 (15.3)	193 (15.3)	
$16,000 - $24,999	270 (22.2)	295 (23.4)	
$25,000 - $34,999	244 (20.1)	248 (19.7)	
$35,000 - $49,999	165 (13.6)	168 (13.3)	
>$50,000	203 (16.7)	213 (16.9)	
Stressful life events			0.27
Median (25th, 75th percentile)	1 (0,2)	1 (0,2)	
Range	0-5	0-6	
Body Mass Index, kg/m^2			0.09
Median (25th, 75th percentile)	25.9 (23.2,29.2)	26.2 (24.2,28.5)	
Range	14.7-48.0	16.9-46.2	
Smoking status, No. (%)			<.0001
Never smoked	790 (59.4)	429 (32.3)	
Former smoker	411 (30.9)	782 (58.9)	
Current smoker	129 (9.7)	117 (8.8)	
Alcohol consumption (drink/week), No. (%)			<.0001
0	695 (52.4)	542 (41.0)	
1-7	505 (38.1)	552 (41.7)	
>7	126 (9.5)	229 (17.3)	
Provide help with IADL, No. (%)	507 (44.6)	505 (43.8)	0.72
Any ADL/IADL difficulty, No. (%)	401 (30.2)	280 (21.1)	<.0001
3MS score			<.0001
Median (25th, 75th percentile)	29 (27,30)	28 (27,29)	
Range	17-30	11-30	
CES-D score			<.0001
Median (25th, 75th percentile)	4 (1,7)	3 (1,5)	
Range	0-26	0-24	
Prevalent clinical CVD, No. (%)	232 (17.4)	427 (32.1)	<.0001
Subclinical CVD, No. (%)	738 (56.5)	930 (70.7)	<.0001
Antidepressant medication use, No. (%)	68 (5.1)	33 (2.5)	<.001

Abbreviation: ADL, Activities of Daily Living; IADL, Instrumental Activities of Daily; CES-D, Center for Epidemiological Studies Depression Scale; CVD, cardiovascular disease.

[a]The number of participants across categories may not sum to the total number of participants because of missing data.

Mortality

Out of 1330 spouse pairs, there are 588 (44%) pairs in which both died, 538 (41%) pairs in which one spouse died (419 husbands and 119 wives), and 204 (15%) pairs where both were censored or alive at the end of study. In cohort 1, 682(54.1%) wives and 962(76.4%) husbands died after 18 years of follow-up. In cohort 2, 25(35.7%) wives and 45(64.3%) husbands died after 6 years of follow-up. After truncating survival times at 6 years, 117(8.8%) wives and 334 (25.1%) husbands died at 6 years.

Joint modeling of bivariate longitudinal depressive symptoms and bivariate mortality

The quadratic term of time was not significant in the bivariate linear mixed-effects models and the model with random intercept fitted the data better with the smallest AIC and BIC values. Thus, the bivariate linear mixed-effects models with random intercept were used to model the longitudinal CES-D score in all analyses.

Table 6.2 shows the results of unadjusted and adjusted joint models for the two cohorts combined with data truncated at 6 years. Without controlling for any covariate, one increase of square root of 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. The correlation of CES-D scores and the dependence of mortality between husbands and wives were both low (r=0.36; Kendall's τ=0.21). 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 increase of square root of CES-D score. CES-D score increased with time in both genders. Older age, less educated, and having stressful life events, ADL/IADL difficulty, and prevalent clinical CVD, and on antidepressant medication were independently related to longitudinal CES-D score in both genders. Non-white race was associated to longitudinal CES-D score 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 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).

Table 6.2: Results of joint modeling of bivariate longitudinal CES-D score and bivariate mortality among spouse pairs in two cohorts combined with data truncated at 6 years.

	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	-	-
Correlation	0.36	-	-	-	0.30	-	-	-
Survival model:								
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	-	-
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
CES-D score[a], 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
CES-D score[a], 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
Kendall's τ	0.21	-	-	-	0.13	-	-	-

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.

[a]The square root of the CES-D score.

Table 6.3 shows the results of unadjusted and adjusted joint models for the two cohorts combined. Without controlling for any covariate, both husbands and wives with higher CES-D score had higher risks of mortality (HR[95%CI], 1.43[1.28-1.59] in wives and 1.43[1.30-1.58] in husbands, per 1 square root of CES-D score). The correlation of CES-D scores and the dependence of mortality between husbands and wives were both low (r=0.33; Kendall's τ=0.13). Age, race, education, stressful life events, ADL/IADL difficulty, prevalent clinical CVD, and antidepressant medication use were independently related to CES-D score in the longitudinal submodel. However, the adjusted joint models did not converge until the longitudinal submodel was reduced to only include age. In the adjusted joint models, mortality increased 23% (95%CI, 1.11-1.37) in wives and 19% (95%CI, 1.08-1.30) in husbands with one increase of square root of CES-D score. CES-D score increased with time and age. Older age and having ADL/IADL difficulty, prevalent clinical CVD, and subclinical CVD were independently associated with mortality. The dependence of mortality between husbands and wives became minimal after adjusting for the covariates (Kendall's τ=0.03).

The results of joint models for cohort 1 only (Table 6.4) were similar to the results of two cohorts combined except that more covariates were able to be adjusted in the longitudinal submodel of CES-D score (i.e., age, race, education, stressful life events, ADL/IADL difficulty, prevalent clinical CVD, and antidepressant medication use).

Due to a small sample size in cohort 2 (n=70 spouse pairs), the joint models did not converge after including any covariate in either longitudinal or survival submodels. Table 6.5 shows the results of unadjusted joint models for cohort 2. Without controlling for any covariates, the square root of the CES-D score was associated with mortality in husbands (HR, 1.79; 95%CI, 1.08-2.95) but not in wives (HR, 1.55; 95%CI, 0.82-2.94). However, due to a relatively small sample size, the model might be underpowered to detect these associations. The correlation of CES-D score and the dependence of mortality between husbands and wives in cohort 2 were both relative higher comparing to the other three situations (r=0.39; Kendall's τ=0.29).

Table 6.3: Results of joint modeling of bivariate longitudinal CES-D score and bivariate mortality among spouse pairs in two cohorts combined.

	Unadjusted (N=1330 spouse pairs)				Adjusted for covariates (N=1292 spouse pairs)			
	Estimate	95% CI	HR	95% CI	Estimate	95% CI	AHR	95% CI
Longitudinal model:								
Wives								
Intercept	1.89	1.84-1.95	-	-	0.38	-0.33-1.08	-	-
Time, per 1 y	0.05	0.05-0.06	-	-	0.05	0.05-0.06	-	-
Age, per 1 y	-	-	-	-	0.02	0.01-0.03		
σ_1	0.77	0.76-0.78	-	-	0.77	0.76-0.78	-	-
Husbands								
Intercept	1.59	1.54-1.64	-	-	0.00	-0.66-0.66	-	-
Time, per 1 y	0.06	0.06-0.07	-	-	0.06	0.05-0.07	-	-
Age, per 1 y	-	-	-	-	0.02	0.01-0.03	-	-
σ_2	0.77	0.75-0.78	-	-	0.77	0.76-0.78	-	-
$\sigma^2_{b_1}$	0.66	0.61-0.71	-	-	0.65	0.60-0.70	-	-
$\sigma_{b_1 b_2}$	0.22	0.18-0.27	-	-	0.21	0.16-0.25	-	-
$\sigma^2_{b_2}$	0.67	0.61-0.72	-	-	0.64	0.59-0.69	-	-
Correlation	0.33	-	-	-	0.33	-	-	-
Survival model:								
θ	0.29	0.19-0.40	-	-	0.06	-0.03-0.14	-	-
ρ	1.73	1.64-1.81	-	-	1.89	1.79-1.98	-	-
Wives								
$\log \lambda_1$	-6.04	-6.38--5.71	-	-	-14.97	-16.25--13.69	-	-
Age, per 1 y	-	-	-	-	0.11	0.10-0.13	1.12	1.10-1.14
White	-	-	-	-	0.28	-0.05-0.61	1.33	0.96-1.84
Any ADL/IADL difficulty	-	-	-	-	0.21	0.04-0.37	1.23	1.04-1.45
Prevalent clinical CVD	-	-	-	-	0.34	0.15-0.53	1.40	1.16-1.69
Subclinical CVD	-	-	-	-	0.38	0.22-0.55	1.47	1.25-1.73
CES-D score[a], per 1	0.36	0.25-0.46	1.43	1.28-1.59	0.21	0.10-0.31	1.23	1.11-1.37
Husbands								
$\log \lambda_2$	-5.17	-5.45--4.89	-	-	-12.26	-13.35--11.18	-	-
Age, per 1 y	-	-	-	-	0.09	0.07-0.10	1.09	1.08-1.11
White	-	-	-	-	-0.04	-0.29-0.21	0.96	0.75-1.23
Any ADL/IADL difficulty	-	-	-	-	0.47	0.30-0.63	1.59	1.35-1.87
Prevalent clinical CVD	-	-	-	-	0.33	0.19-0.47	1.40	1.21-1.61
Subclinical CVD	-	-	-	-	0.50	0.34-0.66	1.65	1.40-1.94
CES-D score[a], per 1	0.36	0.26-0.46	1.43	1.30-1.58	0.17	0.08-0.26	1.19	1.08-1.30
Kendall's τ	0.13	-	-	-	0.03	-	-	-

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.

[a]The square root of the CES-D score.

Table 6.4: Results of joint modeling of bivariate longitudinal CES-D score and bivariate mortality among spouse pairs in cohort 1.

	Unadjusted (N=1260 spouse pairs)				Adjusted for covariates (N=1210 spouse pairs)			
	Estimate	95% CI	HR	95% CI	Estimate	95% CI	AHR	95% CI
Longitudinal model:								
Wives								
Intercept	1.89	1.83-1.94	-	-	1.19	0.42-1.97	-	-
Time, per 1 y	0.06	0.05-0.06	-	-	0.06	0.05-0.06	-	-
Age, per 1 y	-	-	-	-	0.01	0.00-0.02	-	-
White	-	-	-	-	-0.13	-0.43-0.17	-	-
Education, per 1 y	-	-	-	-	-0.02	-0.03--0.01	-	-
Stressful life event, per 1	-	-	-	-	0.11	0.07-0.15	-	-
Any ADL/IADL difficulty	-	-	-	-	0.42	0.32-0.52	-	-
Prevalent clinical CVD	-	-	-	-	0.15	0.03-0.27	-	-
Antidepressant medication use	-	-	-	-	0.39	0.18-0.60	-	-
σ_1	0.77	0.76-0.78	-	-	0.77	0.76-0.78	-	-
Husbands								
Intercept	1.57	1.52-1.62	-	-	0.91	0.19-1.64	-	-
Time, per 1 y	0.06	0.06-0.07	-	-	0.06	0.06-0.07	-	-
Age, per 1 y	-	-	-	-	0.01	0.00-0.02	-	-
White	-	-	-	-	-0.29	-0.58--0.01	-	-
Education, per 1 y	-	-	-	-	-0.02	-0.03--0.01	-	-
Stressful life event, per 1	-	-	-	-	0.11	0.07-0.15	-	-
Any ADL/IADL difficulty	-	-	-	-	0.41	0.30-0.52	-	-
Prevalent clinical CVD	-	-	-	-	0.22	0.12-0.32	-	-
Antidepressant medication use	-	-	-	-	0.67	0.36-0.99	-	-
σ_2	0.77	0.76-0.78	-	-	0.77	0.76-0.78	-	-
$\sigma^2_{b_1}$	0.66	0.61-0.71	-	-	0.57	0.52-0.61	-	-
$\sigma_{b_1 b_2}$	0.22	0.17-0.27	-	-	0.15	0.11-0.20	-	-
$\sigma^2_{b_2}$	0.65	0.60-0.70	-	-	0.54	0.49-0.58	-	-
Correlation	0.34	-	-	-	0.27	-	-	-
Survival model:								
θ	0.29	0.19-0.40	-	-	0.07	-0.02-0.16	-	-
ρ	1.74	1.65-1.83	-	-	1.90	1.80-2.00	-	-
Wives								
$\log \lambda_1$	-6.07	-6.41--5.73	-	-	-15.28	-16.67--13.89	-	-
Age, per 1 y	-	-	-	-	0.12	0.10-0.14	1.13	1.11-1.15
White	-	-	-	-	0.20	-0.33-0.74	1.23	0.72-2.09
Any ADL/IADL difficulty	-	-	-	-	0.20	0.03-0.38	1.23	1.03-1.46
Prevalent clinical CVD	-	-	-	-	0.32	0.12-0.51	1.37	1.13-1.67
Subclinical CVD	-	-	-	-	0.39	0.22-0.56	1.48	1.25-1.74
CES-D score[a], per 1	0.35	0.25-0.46	1.42	1.28-1.59	0.22	0.11-0.33	1.25	1.12-1.39
Husbands								
$\log \lambda_2$	-5.19	-5.48--4.91	-	-	-12.48	-13.66--11.30	-	-
Age, per 1 y	-	-	-	-	0.09	0.08-0.10	1.09	1.08-1.11
White	-	-	-	-	0.03	-0.38-0.44	1.03	0.68-1.55
Any ADL/IADL difficulty	-	-	-	-	0.40	0.24-0.57	1.50	1.26-1.78
Prevalent clinical CVD	-	-	-	-	0.37	0.22-0.52	1.45	1.25-1.68
Subclinical CVD	-	-	-	-	0.47	0.31-0.64	1.61	1.36-1.90
CES-D score[a], per 1	0.35	0.25-0.45	1.42	1.29-1.57	0.19	0.10-0.29	1.21	1.10-1.34
Kendall's τ	0.13	-	-	-	0.03	-	-	-

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.

[a]The square root of the CES-D score.

Table 6.5: Results of joint modeling of bivariate longitudinal CES-D score and bivariate mortality among spouse pairs in cohort 2.

	Unadjusted (N=70 spouse pairs)			
	Estimate	95% CI	HR	95% CI
Longitudinal model:				
Wives				
Intercept	2.08	1.83-2.33	-	-
Time, per 1 y	0.02	-0.02-0.06	-	-
σ_1	0.75	0.69-0.82	-	-
Husbands				
Intercept	1.80	1.52-2.09	-	-
Time, per 1 y	0.04	0.00-0.09	-	-
σ_2	0.74	0.67-0.80	-	-
$\sigma^2_{b_1}$	0.67	0.41-0.92	-	-
$\sigma_{b_1 b_2}$	0.30	0.01-0.58	-	-
$\sigma^2_{b_2}$	0.96	0.66-1.26	-	-
Correlation	0.37	-	-	-
Survival model:				
θ	0.82	-0.09-1.72	-	-
ρ	1.65	1.20-2.09	-	-
Wives				
$\log \lambda_1$	-6.22	-8.34--4.10	-	-
CES-D score[a], per 1	0.47	-0.25-1.19	1.60	0.78-3.28
Husbands				
$\log \lambda_2$	-5.51	-8.34--4.10	-	-
CES-D score[a], per 1	0.58	0.08-1.08	1.79	1.08-2.95
Kendall's τ	0.29	-	-	-

Abbreviation: CI, confidence interval; HR, hazard ratio; CES-D, Center for Epidemiological Studies Depression Scale.
[a]The square root of the CES-D score.
[*]The joint models did not converge after including any covariate in either longitudinal or survival submodels due to a small sample size (n=70 spouse pairs).

The impact of ignoring the correlation of CES-D score and the dependence of mortality between husbands and wives

The joint models with correlations ignored did not converge after including any covariate in either longitudinal or survival submodels. Thus, we refitted the unadjusted joint model for the two cohorts combined with data truncated at 6 years. In addition, we also refitted the unadjusted joint model for cohort 2 because the dependence of mortality between husbands and wives are higher comparing to other models.

Table 6.6 shows the results of unadjusted joint models ignoring the correlations for two cohorts combined with data truncated at 6 years and cohort 2 only. For two cohorts combined with data truncated at 6 years, the longitudinal parameter estimates were similar to the model taking into account the correlations in Table 6.2. The wives' association parameter estimate in the survival submodel was similar but the 95%CI was wider when ignoring the correlations (i.e., HR(95%CI) = 1.83(1.38-2.43) in Table 6.6 vs. 1.82(1.42-2.34) in Table 6.2). The husbands' association parameter estimate was smaller when ignoring the correlations (i.e., HR(95%CI) = 1.55(1.34-1.80) vs. 1.66(1.40-1.96)). For cohort 2, the longitudinal parameter estimates were similar to the model accounting for the correlations in Table 6.5. Both wives' and husbands' association parameter estimates were smaller when ignoring the correlations (i.e., wives, HR(95%CI) = 1.51(0.79-2.88) in Table 6.6 vs. 1.60(0.78-3.28) in Table 6.5; husbands, HR(95%CI) = 1.46(1.03-2.06) vs. 1.79(1.08-2.95)).

Table 6.6: Results of joint modeling of longitudinal CES-D score and mortality among spouse pairs when the correlations were ignored.

	Two cohorts combined truncated at 6 years (N=1330 spouse pairs)				Cohort 2 only (N=70 spouse pairs)			
	Estimate	95% CI	HR	95% CI	Estimate	95% CI	HR	95% CI
Longitudinal model[a]:								
Wives								
Intercept	1.87	1.82-1.92	-	-	2.10	1.86-2.34	-	-
Time, per 1 y	0.07	0.06-0.08	-	-	0.02	-0.02-0.06	-	-
Husbands								
Intercept	1.56	1.51-1.61	-	-	1.87	1.62-2.11	-	-
Time, per 1 y	0.08	0.07-0.08	-	-	0.03	-0.01-0.07	-	-
σ	0.76	0.75-0.77	-	-	0.75	0.69-0.83	-	-
σ_b^2	0.67	0.62-0.72	-	-	0.77	0.51-0.99	-	-
Survival model[b]:								
ρ	1.37	1.24-1.50	-	-	1.43	1.15-1.78	-	-
Wives								
log λ_1	-6.34	-7.06--5.62	-	-	-5.63	-7.28--3.98	-	-
CES-D score[c], per 1	0.61	0.32-0.89	1.83	1.38-2.43	0.41	-0.23-1.06	1.51	0.79-2.88
Husbands								
log λ_2	-4.65	-5.04--4.27	-	-	-4.55	-5.67--3.44	-	-
CES-D score[c], per 1	0.44	0.29-0.59	1.55	1.34-1.80	0.38	0.03-0.73	1.46	1.03-2.06

Abbreviation: CI, confidence interval; HR, hazard ratio; CES-D, Center for Epidemiological Studies Depression Scale.
[a]Linear mixed-effects models were used.
[b]Weibull proportional hazards models were used.
[c]The square root of the CES-D score.

6.5 DISCUSSION

We utilized a new joint modeling approach to simultaneously investigate the association of both longitudinal depressive symptoms and mortality between husbands and wives and to examine the effect of longitudinal depressive symptoms level on mortality among spouses in a large population-based sample of older adults. We show that longitudinal CES-D score was a significant independent risk factor for mortality in both husbands and wives after adjusting for sociodemographic factors (i.e. age, race), prevalent clinical CVD, subclinical CVD, and ADL/IADL difficulty. The inconsistent findings in the literature about the relationship between depression and mortality in older population might be due to methodologic limitations. The present joint modeling methodology attempted to overcome these problems. First, most studies only had a one-time assessment of depression. However, depression status is a dynamic process and the change in severity of depression may have temporal effects on mortality, which cannot be captured by a single assessment. In contrast, our study had 6-10 annual assessments of depressive symptoms, which gave a better picture of how an individual's depression status changed over time and reflected the chronic nature of depressive symptoms. Second, most studies had relatively short follow-up periods with a small number of deaths. Our study had 15-18 years of follow-up on mortality. With a longer follow-up and a large community sample (1330 spouse pairs), we were able to capture more events of mortality (53.2% in wives; 75.7% in husbands) and thus increase the statistical power to detect the association. Even after truncating survival times at 6 years, we still had 117 (8.8%) deaths in wives and 334 (25.1%) deaths in husbands. In addition, our study shows that the associations between longitudinal depressive symptoms and mortality were slightly attenuated after adjusting for age, race, prevalent clinical CVD, subclinical CVD, and ADL/IADL difficulty. 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 underestimating the true association between the longitudinal depressive symptoms and mortality.

Consistent with previous work, the longitudinal parameter estimates obtained from the joint models indicated 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 CES-D score does not mean that depressive symptoms were worse among those taking medication. This association is not causal, and thus it is more possible that not antidepressant medication caused symptoms but rather antidepressant medication use is a proxy for having clinical depression (i.e., severe depressive symptoms).

From a clinical perspective, the current findings suggested that levels 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.

In spite of the advantages of the current study, there are at least two limitations in this work. First, we did not have clinical diagnoses of depression or the psychiatric history of the participants. The association between clinical depression and mortality may be stronger than the association of depressive symptoms and mortality we observed in this study. Second, there are increasing numbers of studies indicating a possible link between depression and cardiovascular mortality. A further analysis linking depression to specific causes of death will help us understand better these possible mechanisms.

7.0 CONCLUSIONS

In this dissertation, 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/physical conditions to the time to mortality in paired subjects and simultaneously investigate the association of both longitudinal psychiatric/physical 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 a lot of possible future work related to this research. We assume a parametric Weibull function for the baseline hazard. The advantage of choosing a parametric proportional hazards model with gamma frailty is that the marginal likelihood is fully parametric and we can rely on classical maximum likelihood technique to estimate the parameters. However, in many cases we may not know what the appropriate baseline hazard distribution is and it is preferred not to make any assumption on its distribution. We will extend the Weibull proportional hazards model with gamma frailty to have a unspecified baseline hazard function. In addition, besides gamma frailty there are other distributions proposed in the literature for frailty, such as the positive stable and the inverse Gaussian distributions. The gamma distribution has been extensively used due to its simple interpretation and mathematical tractability but there is a restriction that the gamma frailty model results in hazard ratios (the ratio of the hazard for a spouse given the other spouse dead at time t and the hazard for a spouse given the other spouse still alive at time t) that are time invariant. It may not be suitable in married couples because an increased risk of mortality in the widowed spouse usually occurs in the early years after loss. A positive stable frailty model has hazard ratios decreasing to one over time and we will consider a positive stable frailty in the future and then compare the preference of the two frailty models.

APPENDIX A

R PROGRAM FOR THE BIVARIATE LINEAR MIXED MODEL

The R function "lme()" is used to conduct the bivariate linear mixed model, i.e., the model fitted in Table 6.2.

```
fitLME <- lme(sqrt(depscr05) ~ -1 + female + male +
                I(female*stdytime_yr) + I(male*stdytime_yr) +
                I(female*agebl) + I(male*agebl) +
                I(female*white) + I(male*white) +
                I(female*grade01) + I(male*grade01) +
                I(female*lescr05) + I(male*lescr05) +
                I(female*anyADLb) + I(male*anyADLb) +
                I(female*prevalent) + I(male*prevalent) +
                I(female*depmedb) + I(male*depmedb),
              random = ~ -1 + female + male | spouse,
              weight = varIdent(form = ~ 1 | male),
              corCompSymm(value = 0, form = ~ 1|spouse/male, fixed = TRUE),
              data = CHS, control = list(apVar = TRUE))
```

APPENDIX B

R PROGRAM FOR THE WEIBULL PROPORTIONAL HAZARDS MODEL WITH GAMMA FRAILTY

The R function "WB.GMfrailty()" is created to conduct the Weibull proportional hazards model with gamma frailty, i.e., the model fitted in Table 6.2.

```
fitPARFM <- WB.GMfrailty(formula = Surv(ttodth_yr2, death2) ~ -1 + female + male +
                                    I(female*agebl) + I(male*agebl) +
                                    I(female*white) + I(male*white) +
                                    I(female*anyADLb) + I(male*anyADLb) +
                                    I(female*prevalent) + I(male*prevalent) +
                                    I(female*subclinical) + I(male*subclinical),
                                  formula.cox = Surv(ttodth_yr2, death2) ~ female +
                                    agebl + white + anyADLb +  prevalent + subclinical +
                                    female:agebl + female:white + female:anyADLb +
                                    female:prevalent + female:subclinical,
                                  cluster = "spouse", method = "BFGS", data = CHS.id)

WB.GMfrailty <-
function (formula, formula.cox, cluster = NULL, data, inip = NULL, iniFpar = NULL,
          method = "BFGS", maxit = 500, Fparscale = 1) {
#------------------  Survival Process -----------------------------------------
obsdata <- NULL
if (length(formula[[2]]) == 3) {
  obsdata$time <- eval(formula[[2]][[2]], envir = data)
  obsdata$event <- eval(formula[[2]][[3]], envir = data)
}
obsdata$x <- as.data.frame(model.matrix(formula, data = data))
obsdata$cluster <- eval(as.name("spouse"), envir = data)
  obsdata$ncl <- length(levels(as.factor(obsdata$cluster)))
  obsdata$di <- aggregate(obsdata$event, by = list(obsdata$cluster), FUN = sum)
              [, , drop = FALSE]
  cnames <- obsdata$di[, 1]
  obsdata$di <- as.vector(obsdata$di[, 2])
  names(obsdata$di) <- cnames
```

```
#----- Dimensions -----------------------------------------------------------
nFpar <- 1
obsdata$nFpar <- nFpar
nBpar <- 1
obsdata$nBpar <- nBpar
nRpar <- ncol(obsdata$x)
obsdata$nRpar <- nRpar

#----- Initial parameters ---------------------------------------------------
coxMod <- phreg(formula = formula.cox, data = data, dist = "weibull", shape = 0,
          control = list(maxiter = maxit))
  logshape <- as.numeric(coxMod$coef[substr(names(coxMod$coef), 5, 9) == "shape"])
  logscale <- as.numeric(coxMod$coef[substr(names(coxMod$coef), 5, 9) == "scale"])

  if (nRpar==2) {
    p.init <- numeric(nRpar)
    p.init [1] <- -exp(logshape)*logscale + coxMod$coef[1]
    p.init [2] <- -exp(logshape)*logscale
  } else if (nRpar>2) {
    p.init1 <- numeric(2)
    p.init1 [1] <- -exp(logshape)*logscale + coxMod$coef[1]
    p.init1 [2] <- -exp(logshape)*logscale

    k <- (nRpar)/2-1
    coef2 <- coxMod$coef[2:(2*k+1)]
    p.init2 <- numeric(k*2)
    for (i in 1:(k*2)) {
       if (i %% 2 == 1)  {
          p.init2[i] <- coef2[(i+1)/2] + coef2[(i+1)/2+k]
       }
       else if (i %% 2 == 0) {
          p.init2[i] <- coef2[i/2]
       }
    }
    p.init <- c(p.init1, p.init2)
  }

p.init <- c(logshape, p.init)
iniFpar <- 1
pars <- log(iniFpar)
pars <- c(pars, p.init)

# ---- Mloglikelihood: Minus the log-likelihood  ----------------------------------
Mloglikelihood <- function(p) {
  theta <- exp(p[1])
  nFpar <- 1
  rho   <- exp(p[nFpar+1])
  beta  <- p[-(1:(nFpar+1))]
  obs<- obsdata

  cumhaz <- NULL
  cumhaz <- aggregate(obs$time^(rho) * exp(as.matrix(obs$x) %*% c(beta))[,1],
                    by=list(obs$cluster),FUN=sum)[,2]
  loghaz <- NULL
```

```
  loghaz <- aggregate(obs$event*(log(rho * obs$time^(rho-1)) +
                                   as.matrix(obs$x) %*% c(beta)),
                      by=list(obs$cluster), FUN=sum)[, 2]
  Mloglik <-sum(obs$di)*log(theta)-
            sum((obs$di+1/theta)*log(1+cumhaz*theta)) + sum(loghaz) +
            sum(sapply(obs$di,function(x)
                       ifelse(x==0,0,log(prod(x+1/theta-seq(1,x))))))
  Mloglik<- -Mloglik
  attributes(Mloglik)$cumhaz <- cumhaz
  return(Mloglik)
}

# ---- Estimate the MLE   ------------------------------------------------------
Fparscale <- 1
res <- NULL
res <- optim(par = pars, fn = Mloglikelihood, method = method, hessian = TRUE,
             control = list(maxit = maxit, parscale = c(rep(Fparscale, nFpar),
 rep(1, nBpar + nRpar))))
res
attributes(res)
it <- res$counts[1]
lL <- -res$value    loglikelihood

#----- Recover the estimates ---------------------------------------------------
theta <- exp(res$par[1:nFpar])
rho <- exp(res$par[nFpar+1])
beta <- res$par[-(1:(nFpar + nBpar))]
names(beta) <- paste(names(obsdata$x), sep = ".")
ESTIMATE <- c(theta = theta, rho = rho, beta = beta)

#----- Output ------------------------------------------------------------------
resmodel<- ESTIMATE
Terms <- terms(formula, data = data)
y<-cbind(obsdata$time,obsdata$event)
rownames(y)<- rep(1:nrow(y))
colnames(y)<- c("time","event")
obsdata$x.cox <- as.data.frame(model.matrix(formula.cox, data = data))
output<-list(model=resmodel,
             x = obsdata$x,
             x.cox = obsdata$x.cox,
             y = y,
             formula = formula,
             formula.cox = formula.cox,
             terms = attr(Terms, "term.labels"))
return(output)
}
```

APPENDIX C

R PROGRAM FOR THE JOINT MODELS

The R function "BivJM()" is created to conduct the joint models, i.e., the model fitted in Table 6.2.

```
lmeObject  <- fitLME        # longitudinal submodel
survObject <- fitPARFM      # survival submodel
timeVar <- "stdytime_yr"    # time variable used in the bivariate linear mixed model
strata  <- "female"
main.effect<- c("female", "agebl", "white", "anyADLb", "prevalent", "subclinical")
control <- list()

out<- BivJM(lmeObject = fitLME, survObject = fitPARFM, timeVar = "stdytime_yr",
            strata = "female", main.effect = c("female", "agebl", "white", "anyADLb",
"prevalent", "subclinical"))
BivJM <-
  function (lmeObject, survObject, timeVar, strata, main.effect, control=list()) {
#------------------  Survival Process -----------------------------------------
formT <- formula(survObject)
W     <- survObject$x
W.cox <- survObject$x.cox
Time  <- survObject$y[, 1]
d     <- survObject$y[, 2]
idT   <- seq_along(Time)
spouseT<- rep(1:(length(idT)/2), each=2)
nRisks <- 1
nT     <- length(unique(idT))

#------------------  Longitudinal Process -------------------------------------
id <- rep(1:(2*max(lmeObject$data[,"spouse"])), rle(lmeObject$data[,"id"])$lengths)
spouse   <- lmeObject$data$spouse
n.spouse <- max(lmeObject$data$spouse)
b <- cbind(rep(ranef(lmeObject)$female,each=2),rep(ranef(lmeObject)$male,each=2))
nY <- nrow(b)
if (nY != nT)
  stop("sample sizes in the longitudinal and event processes differ; ",
       "maybe you forgot the cluster() argument.\n")
```

```
TermsX <- lmeObject$terms
data   <- lmeObject$data[all.vars(TermsX)]
data   <- data[complete.cases(data), ]
formYx <- formula(lmeObject)
mfX <- model.frame(TermsX, data = data)
X   <- model.matrix(formYx, mfX)
formYz <- formula(lmeObject$modelStruct$reStruct[[1]])
mfZ    <- model.frame(terms(formYz), data = data)
TermsZ <- attr(mfZ, "terms")
Z      <- model.matrix(formYz, mfZ)
y.long <- model.response(mfX, "numeric")

#---------- check if there are any longitudinal measurements after the event times -----
data.id <- data[!duplicated(id), ]
data.id <- data.id[idT, ]
if (!timeVar %in% names(data))
  stop("\n'timeVar' does not correspond to one of the columns in the
model.frame of 'lmeObject'.")
max.timeY <- tapply(data[[timeVar]], id, max)
max.timeT <- tapply(Time, idT, max)
if (!all(max.timeT >= max.timeY)) {
  idnams <- factor(lmeObject$groups[[1]])
  stop("\nit seems that there are longitudinal measurements taken after the event times
for some subjects", "(i.e., check subject(s): ",
paste(levels(idnams)[(max.timeT < max.timeY)], collapse = ", "), ").")
}
data.id[[timeVar]] <- pmax(Time, 0)

#---------- Longitudinal outcome in the survival submodel ------------------------------
mfX.id <- model.frame(TermsX, data = data.id)
mfZ.id <- model.frame(TermsZ, data = data.id)
Xtime <- model.matrix(formYx, mfX.id)
Ztime <- model.matrix(formYz, mfZ.id)

#--------------------  Estimated longitudinal outcome  --------------------------------
long <- c(X %*% fixef(lmeObject)) + rowSums(Z * b[id, ])

#------------ response vectors and design matrices ---------------------------------------
y <- list(y = y.long, logT = log(Time), Time = Time, d = d)
x <- list(X = X, Z = Z, W = W, idT = idT, spouse = spouse, spouseT = spouseT,
          n.spouse = n.spouse, nRisks = nRisks,
          Xtime = Xtime, Ztime = Ztime)

#----------------- control values ------------------------------------------------------
con <- list(only.EM = FALSE, iter.EM = 50, iter.qN = 150, optimizer = "optim",
            tol1 = 0.001, tol2 = 1e-04, tol3 = sqrt(.Machine$double.eps),
            numeriDeriv = "fd", eps.Hes = 1e-06, parscale = NULL,
            step.max = 0.1, backtrackSteps = 2, knots = NULL, ObsTimes.knots = TRUE,
            lng.in.kn = 5, ord = 4, equal.strata.knots = TRUE,
            GHk = if (ncol(Z) <= 3 && nrow(Z) < 2000) 5 else 3, GKk = 15,
            verbose = TRUE)
control <- c(control, list())
namC <- names(con)
```

```
con[(namc <- names(control))] <- control

#-------- extra design matrices for Weibull-PH-GH -------------------------------------
# Gauss-Kronrod
wk <- gaussKronrod(con$GKk)$wk
sk <- gaussKronrod(con$GKk)$sk
P <- as.vector(Time)/2
st <- outer(P, sk + 1)
dimnames(st) <- names(P) <- NULL
id.GK <- rep(seq_along(Time), each = con$GKk)
data.id2 <- data.id[id.GK, , drop = FALSE]
data.id2[[timeVar]] <- pmax(c(t(st)) , 0)
mfX <- model.frame(TermsX, data = data.id2)
mfZ <- model.frame(TermsZ, data = data.id2)
Xs <- model.matrix(formYx, mfX)
Zs <- model.matrix(formYz, mfZ)
x <- c(x, list(P = P, st = c(t(st)), wk = wk,  Xs = Xs, Zs = Zs))

#------------------ initial values -------------------------------------------------------
VC <- lapply(pdMatrix(lmeObject$modelStruct$reStruct), "*", lmeObject$sigma^2)[[1]]
R.element<- (1/unique(attributes(lmeObject$modelStruct$varStruct)$weights)
             *lmeObject$sigma)^2
R.female <- R.element[1]
R.male   <- R.element[2]
Vs <- vector("list", n.spouse)
inv.VC <- solve(VC)
ni  <- as.vector(tapply(spouse, spouse, length))
diag.R  <- ifelse(id %% 2 == 0, R.male, R.female)
diag.R.s<- split(diag.R, spouse)
names(diag.R.s) <- NULL
for (i in 1:n.spouse) {
  Z.i <- Z[spouse == i, , drop = FALSE]
  R.i <- diag(diag.R.s[[i]],ni[i],ni[i])
  Vs[[i]] <- solve(t(Z.i) %*% solve(R.i) %*% Z.i + inv.VC)
}
con$inv.chol.VCs <- lapply(Vs, function(x) solve(chol(solve(x))))
con$det.inv.chol.VCs <- sapply(con$inv.chol.VCs, det)
con$ranef <- b
con$ranef.spouse<- b[which(idT %% 2 == 1),]
init.parm <- initial.parm(Time, d, W, W.cox, id, idT, long = long,
             cluster = "spouse")
initial.values <- c(list(betas = fixef(lmeObject), sigma = sqrt(R.element), D = VC),
                  init.parm)

#------------------ remove objects -------------------------------------------------------
rmObjs <- c(names(x), "y.long", "mfX", "mfZ")
rm(list = rmObjs)
gc()

#------------------ joint model fit ------------------------------------------------------
control <- con

#---- response vectors ------------------------------------------------------------------
spouse <-  x$spouse
```

```
spouseT<-  x$spouseT
idT <- x$idT
Time <- as.vector(y$Time)
d <- as.vector(y$d)
d.spouse <- as.vector(rowsum(d, spouseT, reorder=FALSE))
y <- as.vector(y$y)

#---- design matrices  ------------------------------------------------------
X <- x$X
Xtime <- x$Xtime
Xs <- x$Xs
Z <- x$Z
Ztime <- x$Ztime
Zs <- x$Zs
WW <- as.matrix(x$W)
X <- dropAttr(X); Z <- dropAttr(Z); WW <- dropAttr(WW)
Xtime <- dropAttr(Xtime); Ztime <- dropAttr(Ztime)
Xs <- dropAttr(Xs); Zs <- dropAttr(Zs)

#---- sample size settings --------------------------------------------------
ncx <- ncol(X)
ncz <- ncol(Z)
ncww<- ncol(WW)
n <- length(Time)
N <- length(y)
n.spouse <- x$n.spouse
nik<- as.vector(tapply(id, id, length))
ni <- as.vector(tapply(spouse, spouse, length))

#---- crossproducts and others ----------------------------------------------
XtX <- crossprod(X)
ZtZ <- lapply(split(Z, id), function (x) crossprod(matrix(x, ncol = ncz)))
names(ZtZ) <- NULL
ZtZ <- matrix(unlist(ZtZ), n, ncz * ncz, byrow=TRUE)
outer.Ztime <- lapply(1:n, function (x) Ztime[x, ] %o% Ztime[x, ])

#---- Gauss-Kronrod rule ----------------------------------------------------
st <- x$st
log.st <- log(st)
wk <- rep(x$wk, length(Time))
P <- as.vector(x$P)
id.GK <- rep(seq_along(Time), each = control$GKk)
id.GK.spouse <- rep(1:n.spouse, each = control$GKk*2)

#---- Pseudo-adaptive Gauss-Hermite quadrature rule components --------------
GH <- gauher(control$GHk)
b <- as.matrix(expand.grid(rep(list(GH$x), ncz)))
k <- nrow(b)
wGH <- as.matrix(expand.grid(rep(list(GH$w), ncz)))
wGH <- 2^(ncz/2) * apply(wGH, 1, prod) * exp(rowSums(b * b))
b <- sqrt(2) * b
dimnames(b) <- NULL
b2 <- if (ncz == 1) b * b else t(apply(b, 1, function (x) x %o% x))
VCdets <- control$det.inv.chol.VCs
```

```
lis.b <- vector("list", n.spouse)
for (i in 1:n.spouse) {
  lis.b[[i]] <- t(control$inv.chol.VCs[[i]] %*% t(b)) + rep(control$ranef.spouse[i, ],
              each = k)
}
lis.b2 <- lapply(lis.b, function (b) if (ncz == 1) b * b else
  t(apply(b, 1, function (x) x %o% x)))
Ztb <- matrix(NA,length(spouse),k)
Ztime.b <- matrix(NA,length(spouseT),k)
Zsb <- matrix(NA,length(id.GK.spouse),k)
for (i in 1:n.spouse) {
  Ztb[spouse == i, ] <- Z[spouse == i, , drop = FALSE] %*% t(lis.b[[i]])
  Ztime.b[spouseT == i, ] <- Ztime[spouseT == i, , drop = FALSE] %*% t(lis.b[[i]])
  Zsb[id.GK.spouse == i, ] <- Zs[id.GK.spouse == i, ] %*% t(lis.b[[i]])
}

#---- initial values -------------------------------------------------------------
betas  <- as.vector(initial.values$betas)
sigma  <- initial.values$sigma
gammas <- as.vector(initial.values$gammas)
alpha  <- as.vector(initial.values$alpha)
theta  <- initial.values$theta
rho    <- initial.values$rho
D <- initial.values$D
diag.D <- !is.matrix(D)
if (!diag.D) dimnames(D) <- NULL else names(D) <- NULL

#---- fix environments for functions ---------------------------------------------
environment(opt.survWBGM) <- environment(gr.survWBGM) <- environment()
environment(gr.longWBGM)  <- environment(H.longWBGM)  <- environment()
environment(LogLik.WBPH.GMfrailty) <- environment(Score.WBPH.GMfrailty) <- environment()
old <- options(warn = (-1))
on.exit(options(old))

#---- EM iterations --------------------------------------------------------------
iter <- control$iter.EM
Y.mat <- matrix(0, iter + 1, ncx + 2)
T.mat <- matrix(0, iter + 1, ncww + 2 + 2)
B.mat <- if (diag.D) matrix(0, iter + 1, ncz) else matrix(0, iter + 1, ncz * ncz)
lgLik <- numeric(iter + 1)
conv <- TRUE

for (it in 1:iter) {
  Y.mat[it, ] <- c(betas, sigma^2)
  T.mat[it, ] <- c(theta, rho, gammas, alpha)
  B.mat[it,]  <- D
  eta.yx <- as.vector(X %*% betas)
  eta.tw <- as.vector(WW %*% gammas)
  exp.eta.tw <-exp(eta.tw)
  Y <- as.vector(Xtime %*% betas) + Ztime.b
  Ys <- as.vector(Xs %*% betas) + Zsb
  alpha.id <- rep(alpha, times=n.spouse)
  alpha.GK <- rep(alpha.id, each = con$GKk)
  eta.t <- eta.tw + alpha.id * Y
```

```
eta.s <- alpha.GK * Ys

# E-step
mu.y <- eta.yx + Ztb
sigma.y  <- ifelse(id %% 2 == 0, sigma[2], sigma[1])
logNorm <- dnorm(y, mu.y, sigma.y, log = TRUE)
log.p.yb <- rowsum(logNorm, spouse, reorder = FALSE)
dimnames(log.p.yb) <- NULL
cumhaz1<- NULL; cumhaz<- NULL
cumhaz1 <-rowsum(wk*exp(log(rho)+(rho-1)*log.st+eta.s), id.GK, reorder=FALSE)
cumhaz <- rowsum(exp.eta.tw * P * cumhaz1, spouseT, reorder=FALSE)
loghaz <- NULL
loghaz <- rowsum(d * (log(rho * Time^(rho-1)) + eta.t), spouseT, reorder=FALSE)
log.p.tb <- d.spouse*log(theta) + loghaz - (d.spouse+1/theta)*log(1+theta*cumhaz) +
            log(gamma(d.spouse+1/theta)/gamma(1/theta))
log.p.b <- matrix(dmvnorm(do.call(rbind, lis.b), rep(0, ncz), D, log=TRUE),
         n.spouse, k, byrow = TRUE)
p.ytb <- exp(log.p.yb + log.p.tb + log.p.b) * VCdets
p.yt <- c(p.ytb %*% wGH)
p.byt <- p.ytb / p.yt
post.b <- sapply(seq_len(ncz), function (i)
  (p.byt * t(sapply(lis.b, "[", seq_len(k), i))) %*% wGH)
post.vb <- {
  dd <- sapply(seq_len(ncz^2), function (i)
    (p.byt * t(sapply(lis.b2, "[", seq_len(k), i))) %*% wGH)
  bb <- apply(post.b, 1, function (x) x %o% x)
  dd - if (ncz == 1) c(bb) else t(bb)
}
log.p.yt <- log(p.yt)
lgLik[it] <- sum(log.p.yt[is.finite(log.p.yt)])

if (control$verbose) {
  cat("\n\niter:", it, "\n")
  cat("log-likelihood:", lgLik[it], "\n")
  cat("betas:", round(betas, 4), "\n")
  cat("sigma squared:", round(sigma^2, 4), "\n")
  cat("theta:", round(theta, 4), "\n")
  cat("rho:", round(rho, 4), "\n")
  cat("lambda:", round(exp(gammas[1:2]), 4), "\n")
  cat("gammas:", round(gammas[-(1:2)], 4), "\n")
  cat("alpha:", round(alpha, 4), "\n")
  cat("D:", if (!diag.D) round(D[lower.tri(D, TRUE)], 4) else round(D, 4), "\n")
}

# check convergence
if (it > 5 && lgLik[it] > lgLik[it - 1]) {
  thets1 <- c(Y.mat[it - 1, ], T.mat[it - 1, ], B.mat[it - 1, ])
  thets2 <- c(Y.mat[it, ], T.mat[it, ], B.mat[it, ])
  check1 <- max(abs(thets2 - thets1) / (abs(thets1) + control$tol1)) < control$tol2
  check2 <- (lgLik[it] - lgLik[it - 1]) < control$tol3 * (abs(lgLik[it - 1])
        + control$tol3)
  if (check1 || check2) {
    conv <- FALSE
    if (control$verbose)
```

```
        cat("\n\nconverged!\ncalculating Hessian...\n")
      break
    }
  }
  if (iter == 0) break

  if (it == iter) {
    out <- list(flag=1)
    return(out)
  }

  # M-step
  Zb <- rowSums(Z * post.b[spouse, ], na.rm = TRUE)
  mu <- y - eta.yx
  tZZvarb <- ZtZ * post.vb[spouseT, ]
  tr.tZZvarb<- NULL
  tr.tZZvarb[1] <- sum(tZZvarb[which(idT %% 2 == 1),], na.rm = TRUE)
  tr.tZZvarb[2] <- sum(tZZvarb[which(idT %% 2 == 0),], na.rm = TRUE)
  N.f<- length(y[which(id %% 2 == 1)])
  N.m<- length(y[which(id %% 2 == 0)])
  sigman<-NULL
  sigman[1] <- sqrt(c(crossprod(mu[which(id %% 2 == 1)], (mu - 2 * Zb)[which(id %% 2 == 1)])
                      + tr.tZZvarb[1] + crossprod(Zb[which(id %% 2 == 1)])) / N.f)
  sigman[2] <- sqrt(c(crossprod(mu[which(id %% 2 == 0)], (mu - 2 * Zb)[which(id %% 2 == 0)])
                      + tr.tZZvarb[2] + crossprod(Zb[which(id %% 2 == 0)])) / N.m)
  Dn <- matrix(colMeans(dd, na.rm = TRUE), ncz, ncz)
  Dn <- if (diag.D) diag(Dn) else 0.5 * (Dn + t(Dn))
  scbetas <- gr.longWBGM(betas)
  Hbetas  <- H.longWBGM(betas)
  betasn <- betas - c(solve(Hbetas, scbetas))
  list.thetas <- list(logtheta=log(theta), logrho=log(rho), gammas=gammas, alpha=alpha)
  list.thetas <- list.thetas[!sapply(list.thetas, is.null)]
  thetas <- unlist(as.relistable(list.thetas))
  optz.surv <- optim(thetas, opt.survWBGM, gr.survWBGM, method = "BFGS",
                     control = list(maxit = if (it < 5) 20 else 5,
                                    parscale = if (it < 5) rep(0.01, length(thetas))
else rep(0.1, length(thetas))))
  thetasn <- relist(optz.surv$par, skeleton = list.thetas)

  if (is.nan(betasn[1]) | is.nan(betasn[2]) | is.nan(betasn[3]) | is.nan(betasn[4]) |
        is.nan(sigman[1]) | is.nan(sigman[2]) |
        is.nan(Dn[1,1]) | is.nan(Dn[1,2]) | is.nan(Dn[2,1]) | is.nan(Dn[2,2]) |
        is.nan(thetasn$logtheta) | is.infinite(thetasn$logrho) |
        is.infinite(thetasn$gammas[1]) | is.infinite(thetasn$gammas[2]) |
        is.infinite(thetasn$gammas[3]) | is.infinite(thetasn$gammas[4]) |
        is.infinite(thetasn$alpha[1]) | is.infinite(thetasn$alpha[2]) |
        is.infinite(betasn[1]) | is.infinite(betasn[2]) |
        is.infinite(betasn[3]) | is.infinite(betasn[4]) |
        is.infinite(sigman[1]) | is.infinite(sigman[2]) |
        is.infinite(Dn[1,1]) | is.infinite(Dn[1,2]) | is.infinite(Dn[2,1]) |
is.infinite(Dn[2,2]) |
        is.infinite(exp(thetasn$logtheta)) | is.infinite(exp(thetasn$logrho)) |
        is.infinite(exp(thetasn$gammas[1])) | is.infinite(thetasn$gammas[2]) |
        is.infinite(exp(thetasn$gammas[3])) | is.infinite(thetasn$gammas[4]) |
```

```
        is.infinite(thetasn$alpha[1]) | is.infinite(thetasn$alpha[2])){
    out <- list(flag=1)
    return(out)
  }

  # update parameter values
  betas  <- betasn
  sigma  <- sigman
  D      <- Dn
  theta  <- exp(thetasn$logtheta)
  rho    <- exp(thetasn$logrho)
  gammas <- thetasn$gammas
  alpha  <- thetasn$alpha
}

list.thetas <- list(betas = betas, sigma = sigma, D = if (diag.D) log(D)
                    else chol.transf(D),
                    theta = theta, rho = rho, gammas = gammas, alpha = alpha)
thetas <- unlist(as.relistable(list.thetas))
lgLik <- - LogLik.WBPH.GMfrailty(thetas)

#---- Calculate Score vector ------------------------------------------------------
Score <- Score.WBPH.GMfrailty(unlist(thetas))
if (any(Score==0)) {
  out <- list(flag=1)
  return(out)
}

#----  calculate Hessian matrix ---------------------------------------------------
Hessian <- if (control$numeriDeriv == "fd") {
  fd.vec(unlist(thetas), Score.WBPH.GMfrailty, eps = control$eps.Hes)
} else {
  cd.vec(unlist(thetas), Score.WBPH.GMfrailty, eps = 1e-04)
}
se.thetas <- sqrt(diag(solve(Hessian)))
if (any(is.na(se.thetas)) | any(se.thetas>=100)){
  out <- list(flag=1)
  return(out)
}

#----  Final data -----------------------------------------------------------------
names(betas) <- names(initial.values$betas)
names(sigma) <- c("sigma.f", "sigma.m")
if (!diag.D) dimnames(D) <- dimnames(initial.values$D) else
                            names(D) <- names(initial.values$D)
names(theta)  <- "theta"
names(rho)    <- "rho"
names(gammas) <- c("log.lambda.f","log.lambda.m")
names(alpha)  <- c("alpha.f","alpha.m")

nams <- c(paste("Y.", c(names(betas), names(sigma)), sep = ""),
          paste("B.", if (!diag.D) paste("D", seq(1, ncz * (ncz + 1) / 2), sep = "")
      else names(D), sep = ""),
          paste("T.", c(names(theta), names(rho), names(gammas), names(alpha)), sep = ""))
```

```
dimnames(Hessian) <- list(nams, nams)
names(se.thetas)  <- c(nams)
colnames(post.b)  <- colnames(x$Z)

D.vector <- as.vector(D)[c(1,2,4)]
names(D.vector) <- c("D1", "D2", "D3")

if (conv == FALSE) conv <- 0

out<- list(flag=0, coefficients = c(betas, sigma, D.vector, theta, rho, gammas, alpha),
           se = se.thetas, Score = Score, Hessian = Hessian, logLik = lgLik,
           EB = list(iters = it, convergence = conv, n.spouse = n.spouse,
                     n = n, N = N, d = d))

#------------ check for problems with the Hessian at convergence -----------------------
H <- out$Hessian
if (any(is.na(H) | !is.finite(H))) {
  warning("infinite or missing values in Hessian at convergence.\n")
} else {
  ev <- eigen(H, symmetric = TRUE, only.values = TRUE)$values
  if (!all(ev >= -1e-06 * abs(ev[1])))
    warning("Hessian matrix at convergence is not positive definite.\n")
}
return(out)
}
```

BIBLIOGRAPHY

[1] R Thiébaut, H Jacqmin-Gadda, A Babiker, D Commenges, and CASCADE Collaboration. 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, Jan 2005.

[2] AA Tsiatis, V DeGruttola, and MS Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90:27–37, 1995.

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

[4] X Song, M Davidian, and AA Tsiatis. An estimator for the proportional hazards model with multiple longitudinal covariates measured with errors. *Biostatistics*, 3:511–528, 2002.

[5] H Lin, CE McCulloch, and ST Mayne. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, 21(16):2369–2382, Aug 2002.

[6] X Guo and BP Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):16–24, Feb 2004.

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

[8] J Choi, SJ Anderson, TJ Richardsc, and WK Thompsond. 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, 2014.

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

[10] L Liu, X Huang, and J O'Quigley. 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, 2008.

[11] L Liu and X Huang. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C*, 58(1):65–81, 2009.

[12] A Ciocco. On the mortality in husbands and wives. *Human Biology*, 12:508–531, 1940.

[13] J Kaprio, M Koskenvuo, and Rita H. Mortality after bereavement: a prospective study of 95,647 widowed persons. *Am J Public Health*, 77(3):283–287, 1987.

[14] P Martikainen and T Valkonen. Mortality after the death of a spouse: rates and causes of death in a large finnish cohort. *Am J Public Health*, 86(8):1087–1093, 1996.

[15] O Manor and Z Eisenbach. Mortality after spousal loss: Are there socio-demographic differences? *Social Science and Medicine*, 56(2):405–413, 2003.

[16] CL Hart, DJ Hole, DA Lawlor, GD Smith, and TF Lever. Effect of conjugal bereavement on mortality of the bereaved spouse in participants of the renfrew/paisley study. *Journal of Epidemiology and Community Health*, 61(5):455–460, 2007.

[17] ST Stahl, J Chen, S Anderson, and R Schulz. Mortality after bereavement: The role of cardiovascular disease and depression. *Psychosomatic Medicine*, [Epub ahead of print], 2016.

[18] JC Coyne. Depression and the response of others. *Journal of Abnormal Psychology*, 85:186–193, 1976.

[19] JC Coyne, RC Kessler, M Tal, J lurnbull, CB Wortman, and J Greden. Living with a depressed person. *Journal of Clinical and Consulting Psychology*, 55:347–352, 1987.

[20] G Galbaud du Fort, V Kovess, and JF Boivin. Spouse similarity for psychological distress and well-being: a population study. *Psychological Medicine*, 24(2):431–447, 1984.

[21] K Tambs. Transmission of symptoms of anxiety and depression in nuclear families. *Journal of Affective Disorders*, 21:117–126, 1991.

[22] H Kelley. Marriage relationships and aging. In RW Fogel, E Hatfield, SB Kiesler, and Shanas E, editors, *Aging: Stability and change in the family*, page 275–300. New York: Academic Press, 1981.

[23] J Bookwala and R Schulz. Spousal similarity in subjective well-being: the cardiovascular health study. *Psychology and Aging*, 11(4):4582–4590, 1996.

[24] AL Townsend, B Miller, and S Guo. Depressive symptomatology in middle-aged and older married couples: a dyadic analysis. *J Gerontol B Psychol Sci Soc Sci*, 56(6): S352–364, 2001.

[25] LP Fried, NO Borhani, P Enright, CD Furberg, JM Gardin, RA Kronmal, LH Kuller, TA Manolio, MB Mittelmark, and A Newman. The cardiovascular health study: design and rationale. *Ann Epidemiol*, 1(3):263–276, Feb 1991.

[26] AR Ribe, TM Laursen, A Sandbaek, M Charles, M Nordentoft, and M Vestergaard. Long-term mortality of persons with severe mental illness and diabetes: a population-based cohort study in denmark. *Psychological Medicine*, 24:1–11, Mar 2014.

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

[28] S Murrell, S Himmelfarb, and K Wright. Prevalence of depression and its correlates in older adults. *American Journal of Epidemiology*, 111:173–185, 1983.

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

[30] LR Wulsin, GE Vaillant, and VE Wells. A systematic review of the mortality of depression. *Psychosomatic Medicine*, 61(1):6–17, 1999.

[31] R Schulz, SR Beach, DG Ives, LM Martire, AA Ariyo, and WJ Kop. Association between depression and mortality in older adults. *Archives of Internal Medicine*, 160(12): 1761–1768, Jun 2000.

[32] R Schulz, RA Drayer, and BL Rollman. Depression as a risk factor for non-suicide mortality in the elderly. *Biological Psychiatry*, 52(3):205–225, 2002.

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

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

[35] Reinsel G. Estimation and prediction in a multivariate random effect generalized linear model. *Journal of the American Statistical Association*, 79:406–414, 1984.

[36] A Shah, NM Laird, and D Schoenfeld. A random effects model with multiple characteristics with possibly missing data. *Journal of the American Statistical Association*, 92:775–779, 1997.

[37] JL Schafer. Imputation of missing covariates under a multivariate linear mixed model. Technical report, Dept. of Statistics, Penn State University, 1997.

[38] JL Schafer and RM Yucel. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*, 11:437–457, 2002.

[39] CH Morrell, LJ Brant, S Sheng, and S Najjar. Using multivariate mixed effect model to predict hypertension. *In Joint Statistical Meetings, Biometrics Section*, 2003.

[40] YY Chi and JG Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445, Jun 2006.

[41] D Ruppert, W Wand, and R Carroll. *Semiparametric regression*. Cambridge University Press, Cambridge, 2003.

[42] ER Brown, JG Ibrahim, and V DeGruttola. A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73, 2005.

[43] JP Sy, JMG Taylor, and WG Cumberland. A stochastic model for the analysis of bivariate longitudinal aids data. *Biometrics*, 53:542–555, 1997.

[44] BA Coull and J Staudenmayer. Self-modeling regression for multivariate curve data. *Statistica Sinica*, 14:695–711, 2004.

[45] WH Lawton and EA Sylvestre. Self modeling curve resolution. *Technometrics*, 13: 617–633, 1971.

[46] R Gueorguieva. A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*, 1:177–193, 2001.

[47] X Liu, MJ Daniels, and B Marcus. Joint models for the association of longitudinal binary and continuous processes with application to a smoking cessation trial. *Journal of the American Statistical Association*, 104:429–438, 2010.

[48] EJ Gumbel. Bivariate exponential distributions. *Journal of the American Statistical Association*, 55:698–707, 1960.

[49] JE Freund. A bivariate extension of the exponential distribution. *Journal of the American Statistical Association*, 56:971–977, 1961.

[50] AW Marshall and I Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62:30–44, 1967.

[51] HW Block and AP Basu. A continuous bivariate exponential extension. *Journal of the American Statistical Association*, 69(348):1031–1037, 1974.

[52] F Proschan and P Sullo. Estimating the parameters of a multivariate exponential distribution. *Journal of the American Statistical Association*, 71(354):465–472, Jun 1976.

[53] JP Klein, N Keiding, and C Kamby. Semiparametric marshall-olkin models applied to the occurrence of metastases at multiple sites after breast cancer. *Biometrics*, 45: 1073–1086, 1989.

[54] SK Ghosh and AE Gelfand. Latent waiting time models for bivariate event times with censoring. *Sankhya*, 60(1):34–47, 1998.

[55] DD Hanagal. A multivariate weibull distribution. *Economic Quality Control*, 11:193–200, 1996.

[56] DD Hanagal. Bivariate weibull regression model based on censored samples. *Statistical Papers*, 47:137–148, 2005.

[57] DD Hanagal. Parametric bivariate regression analysis based on censored samples: A weibull model. *Economic Quality Control*, 19(1):1–8, 2004.

[58] DD Hanagal. Modeling heterogeneity for bivariate survival data by power variance function distribution. *Journal of Reliability and Statistical Studies*, 2(1):14–27, 2009.

[59] DD Hanagal. Weibull extension of bivariate exponential regression model with different frailty distributions. *Statistical Papers*, 50:29–49, 2009.

[60] DD Hanagal. Modeling heterogeneity for bivariate survival data by the weibull distribution. *Statistical Papers*, 51:947–958, 2010.

[61] JW Vaupel, KG Manton, and E Stallard. The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography*, 16:439–454, 1979.

[62] T Lancaster. Econometric methods for the duration of unemployment. *Econometrica*, 47:939–956, 1979.

[63] SK Sahu and DK Dey. A comparison of frailty and other models for bivariate survival data. *Lifetime Data Analysis*, 6(3):207–228, Sep 2000.

[64] DG Clayton. 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, 1978.

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

[66] DG Clayton and J Cuzick. Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society: Series A*, 148:82–117, 1985.

[67] G Guo and G Rodriguez. 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, 1992.

[68] DD Hanagal. A gamma frailty regression model in bivariate survival data. *IAPQR Transactions*, 31:73–83, 2006.

[69] DD Hanagal. Gamma frailty regression models in mixture distributions. *Economic Quality Control*, 22(2):295–302, 2007.

[70] DD Hanagal and AD Dabade. Bayesian estimation of parameters and comparison of shared gamma frailty models. comm. statist. *Communications in Statistics - Simulation and Computation*, 42(4):910–931, 2013.

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

[72] JH Shih. A goodness-of-fit test for association in a bivariate survival model. *Biometrika*, 85:189–200, 1998.

[73] DV Glidden. Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika*, 86:381–394, 1999.

[74] JJ Fan, L Hsu, and RL Prentice. Dependence estimation over a finite bivariate failure time region. *Lifetime Data Analysis*, 6:343–355, 2000.

[75] P Hougaard. A class of multivariate failure time distributions. *Biometrika*, 73(3): 671–678, 1986.

[76] JP Fine, DV Glidden, and KE Lee. A simple estimator for a shared frailty regression model. *Journal of the Royal Statistical Society: Series B*, 65:317–329, 2003.

[77] P Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986.

[78] GA Whitmore and MT Lee. A multivariate survival distribution generated by an inverse gaussian mixture of exponentials. *Technometrics*, 33(1):39–50, 1991.

[79] N Laird. Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association*, 73:805–811, 1978.

[80] J Heckman and B Singer. A method of minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrical*, 52:231–241, 1984.

[81] S Manda. A nonparametric frailty model for clustered survival data. *Communications in Statistics - Theory and Methods*, 40(5):863–875, 2011.

[82] JH Shih and TA Louis. Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis*, 1:205–220, 1995.

[83] JH Shih and TA Louis. Inference on the association parameter in copula models for bivariate survival data. *Biometrics*, 51:138–499, 1995.

[84] B Nan, X Lin, LD Lisabeth, and S Harlow. Piecewise constant cross-ratio estimation for association of age at a marker event and age at menopause. *Journal of the American Statistical Association*, 101:65–77, 2006.

[85] Y Li, RL Prentice, and X Lin. Semiparametric maximum likelihood estimation in normal transformation models for bivariate survival data. *Biometrika*, 95:947–960, 2008.

[86] T Hu, B Nan, X Lin, and JM Robins. Time-dependent cross ratio estimation for bivariate failure times. *Biometrika*, 98(2):341–354, 2011.

[87] AI Yashin, JW Vaupel, and I A Iachine. Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, 5(2): 145–159, 1995.

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

[89] JW Hardin and JM Hilbe. *Generalized Estimating Equations*. London: Chapman and Hall/CRC, 2003.

[90] KY Liang and S Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[91] LJ Wei, D Lin, and L Weissfeld. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84:1065–1073, 1989.

[92] EW Anderson. Two-stage estimation in copula models used in family studies. *Lifetime Data Analysis*, 11:333–350, 2005.

[93] DR Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, 34(2):187–220, 1972.

[94] D Rizopoulos. Jm: an r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33, 2010.

[95] J. Kalbfleisch and R Prentice. *The Statistical Analysis of Failure Time Data*. Hoboken, New Jersey: JohnWiley and Sons, 2002.

[96] SG Self and Y Pawitan. Modeling a marker of disease progression and onset of disease. In *Statistical Methodology for Study of the AIDS Epidemic*. Boston: Birkhauser, 1992.

[97] MS Wulfsohn and AA Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339, 1997.

[98] P Philipson, I Sousa, P Diggle, P Williamson, R Kolamunnage-Dona, and R Henderson. joiner: Joint modelling of repeated measurements and time-to-event data, 2012. URL http://cran.r-project.org/web/packages/joineR/index.html.

[99] J Ding and JL Wang. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, 64(2):546–556, Jun 2008.

[100] D Rizopoulos, G Verbeke, and E Lesaffre. Fully exponential laplace approximations for the joint modeling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B*, 71:637–654, 2009.

[101] Y Wang and JMG Taylor. Jointly modeling longi-tudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96:895–905, 2001.

[102] SJ Ratcliffe, W Guo, and TRT Have. Joint modeling of longitudinal and survival data via a common frailty. *Biometrics*, 60(4):892–899, 2004.

[103] JG Ibrahim, H Chu, and LM Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796–2801, 2010.

[104] MJ Sweeting and SG Thompson. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763, 2011.

[105] PJ Diggle, I Sousa, and AG Chetwynd. Joint modelling of repeated measurements and time-to-event outcomes: the fourth armitage lecture. *Statistics in Medicine*, 27(16): 2981–2998, 2008.

[106] I Sousa. A review on joint modelling of longitudinal measurements and time-to-event. *REVSTAT*, 9(1):57–81, 2011.

[107] LM McCrink, AH Marshall, and KJ Cairns. Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. *International Statistical Review*, 81(2):249–269, 2013.

[108] MD Schluchter. Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*, 11(14–15):1861–1870, 1992.

[109] CL Faucett and DC Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–1685, 1996.

[110] AA Tsiatis and D Davidian. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 14:809–834, 2004.

[111] R Henderson, PJ Diggle, and A Dobson. Identification sand efficacy of longitudinal markers for survival. *Biostatistics*, 3:33–50, 2002.

[112] AP Dempster, NM Laird, and DB Rublin. Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. B Methodol*, 39(1):1–38, 1977.

[113] J Xu and SL Zeger. Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics*, 50:375–387, 2001.

[114] J Xu and SL Zeger. The evaluation of multiple surrogate endpoints. *Biometrics*, 57: 81–87, 2001.

[115] AA Tsiatis and M Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458, 2001.

[116] F Hsieh, Y Tseng, and J Wang. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 64(4):1037–1043, 2006.

[117] ER Brown and JG Ibrahim. A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2):221–228, 2003.

[118] P Slasor and N Laird. Joint models for efficient estimation in proportional hazards regression models. *Stat. Med*, 22(13):2137–2148, 2003.

[119] YK Tseng, F Hsieh, and JL Wang. Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92(3):587–603, 2005.

[120] WH Press, SA Teukolsky, WT Vetterling, and BP Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, New York, 3rd edition, 2007.

[121] D Rizopoulos. 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, 2012.

[122] R. Ihaka and R Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 1996.

[123] GS Tell, LP Fried, and B Hermanson. Recruitment of adults 65 years and older as participants in the cardiovascular health study. *Ann Epidemiol*, 3:358–366, 1993.

[124] LS Radloff. The ces-d scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1:385–401, 1977.

[125] EM Andresen, JA Malmgren, and WB Carter. 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, 1994.

[126] EL Teng and HC Chui. The modified mini-mental state (3ms) examination. *J Clin Psychiatry*, 48:314 –318, 1897.

[127] R Schulz, LM Martire, and J Klinger. Evidence-based caregiver interventions in geriatric psychiatry. *Psychiatr Clin N Am*, 28:1007–1038, 2005.

[128] BM Psaty, LH Kuller, D Bild, GL Burke, r SJ Kittne, M Mittelmark, TR Price, PM Rautaharju, and J Robbins. Methods of assessing prevalent cardiovascular disease in the cardiovascular health study. *Ann Epidemiol*, 5:270–277, 1995.

[129] J Kuller, L Shemanski, BM Psaty, NO Borhani, J Gardin, MN Haan, DH O'Leary, PJ Savage, GS Tell, and R Tracy. Subclinical disease as an independent risk factor for cardiovascular disease. *Circulation*, 92:720–726, 1995.