PREDICTION IN THE JOINT MODELING OF MIXED TYPES OF MULTIVARIATE LONGITUDINAL OUTCOMES AND A TIME-TO-EVENT OUTCOME

by

Jiin Choi

B.S., Ewha Womans University, South Korea, 2004

M.S., Ewha Womans University, South Korea, 2006

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Jiin Choi

It was defended on

August 8, 2012

and approved by

Dissertation Advisor: **Stewart Anderson, Ph.D.**, Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Jong-Hyeon Jeong, Ph.D., Associate Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Wesley K. Thompson, Ph.D., Assistant Professor In-Residence, Department of Psychiatry,

School of Medicine, University of California, San Diego

Ada O. Youk, Ph.D., Assistant Professor, Department of Biostatistics,

Graduate School of Public Health, University of Pittsburgh

Charles F. Reynolds, M.D., Professor, Department of Psychiatry,

School of Medicine, University of Pittsburgh

PREDICTION IN THE JOINT MODELING OF MIXED TYPES OF MULTIVARIATE LONGITUDINAL OUTCOMES AND A TIME-TO-EVENT OUTCOME

Jiin Choi, Ph.D.

University of Pittsburgh, 2012

A common goal of longitudinal studies is to relate a set of repeated observations to a time-to-event endpoint. One example of such a design is in the area of a late-life depression research where repeated measurement of cognitive and functional outcomes can contribute to one's ability to predict whether or not an individual will have a major depressive episode over a period of time. This research proposes a novel model for the relationship between multivariate longitudinal measurements and a time-to-event outcome. The goal of this model is to improve prediction for the time-to-event outcome by considering all longitudinal measurements simultaneously.

In this dissertation, we investigate a joint modeling approach for mixed types of multivariate longitudinal outcomes and a time-to-event outcome using a Bayesian paradigm. For the longitudinal model of continuous and binary outcomes, we formulate multivariate generalized linear mixed models with two types of random effects structures: shared random effects and correlated random effects. For the joint model, the longitudinal outcomes and the time-to-event outcome are assumed to be independent conditional on available covariates and the shared parameters, which are associated with the random effects of the longitudinal outcome processes. A Bayesian method using Markov chain Monte Carlo (MCMC) computed in OpenBUGS is implemented for parameter estimation.

We illustrate the prediction of future event probabilities within a fixed time interval for patients based on our joint model, utilizing baseline data, post-baseline longitudinal measurements, and the time-to-event outcome. Prediction of event or mortality probabilities allows one to intervene clinically when appropriate. Hence, such methods provide a useful public health tool at both the individual and the population levels.

The proposed joint model is applied to data sets on the maintenance therapies in a late-life depression study and the mortality in idiopathic pulmonary fibrosis. The performance of the method is also evaluated in extensive simulation studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I am thankful to the God Almighty. Without His purpose and grace, my doctoral journey would not have been conceivable. I would like to express deepest appreciation to my academic adviser, Dr. Stewart Anderson, for his guidance, great encouragement, and constant advice. I would sincerely like to thank my committee members: Dr. Jong-Hyeon Jeong, for his perceptive comments in survival analysis; Dr. Wesley Thompson, for his insightful discussion in prediction models; Dr. Ada Youk, for valuable proofreading; and Dr. Charles Reynolds, for the use of the MTLD-III data and for his clinical advices.

I am deeply grateful to my GSR work supervisor, Dr. Naftali Kaminski, for constant financial support through my doctoral study. Special thanks to go to my work supervisor and friend of many years, Dr. Thomas J. Richards, for his sincerity as a Biostatistics Director, his collaboration on the IPF manuscript and other projects, and his full support and encouragement. I would also like to thank all the members in Simmons Center for Interstitial Lung Diseases. It was a great pleasure to work with them. I also wish to thank all of my friends whose constant pray and encouragement have helped me a lot.

Finally, my parents who are always praying for me, four siblings, two sisters-in-law, and brother-in-law who are always standing with me, two nephews and two nieces I have now, and my boyfriend who is always getting along well, I love you... SHALOM…

1.0 INTRODUCTION

In recent years, a common goal of longitudinal studies is to relate a set of repeated observations to a time-to-event endpoint. In such studies, the longitudinal measurements are considered as an outcome variable or a time-dependent covariate measured with error in the time-to-event outcome process. For example, in a late-life depression research repeated measurement of cognitive and functional outcomes can contribute to one's ability to predict whether an individual will have a major depressive episode over a period of time. A further example in chronic lung disease research is a study of the mortality in idiopathic pulmonary fibrosis (IPF) where longitudinally measured pulmonary function tests (PFT) are considered as a predictor for the time-to-event.

Several authors have proposed joint models in which a single quantitative longitudinal outcome is related to a time-to-event outcome. However, longitudinal studies typically have more than one repeated response variable which can be related to a time-to-event outcome. In particular, there can be different types of longitudinal outcomes, both the quantitative and the dichotomous outcomes. We focus on the joint model with mixed types of the longitudinal continuous and binary outcomes and a time-to-event outcome.

This dissertation was motivated by a study of cognitive impairment in a late-life depression. This longitudinal study consisted of two phases of treatment. In the first phase, participants responded to an open antidepressant treatment and completed assessment for the pre-

randomized controlled trial to establish eligibility. The duration of the first phase was 12 to 16 weeks. In the second phase, 130 eligible patients were randomized to receive either to Donepezil hydrochloride or Placebo. The duration of second phase was two years. The primary outcome for each patient was recurrence of major depression. Also of importance were a global measure of neuropsychological functioning, and a composite measure of cognitive instrumental activities of daily living (C-IADL) each of which, were measured longitudinally. One interest in this research is to relate the longitudinal measurements to the depression recurrence outcome. We want to improve prediction for the time to recurrence of depression by considering all longitudinal measurements simultaneously.

We formulate multivariate correlated logistic models for a combination of longitudinal continuous and binary outcomes. For the longitudinal process, we consider random effects mixed models for both the longitudinal continuous and binary outcomes. We consider two types of structure for the random effects of longitudinal outcomes. One is shared random effects model. Another is correlated random effects model. This model falls within the general class of generalized linear mixed models. To build a joint model of mixed types of multivariate longitudinal outcomes and a time-to-event outcome, we assume shared parameters in the time-to-event outcome process, which are associated with the random effects of the longitudinal outcomes. The longitudinal outcomes and the time-to-event outcome are assumed independent conditional on the shared parameters and available covariates. The Bayesian method through Markov chain Monte Carlo (MCMC) approach computed in OpenBUGS software is used for parameter estimation and simulations. We illustrate the dynamic prediction of an event occurring probabilities within a fixed time when a subject still does not fail at just before time t. The

survival function is based on the joint modeling of baseline data, post-baseline longitudinal measurements, and a time-to-event outcome.

This dissertation is organized as follows. In chapter 2, we present a literature review of statistical models with key properties in the analysis of joint models for the longitudinal and the time-to-event outcomes and predictive modeling of an event probability. In chapter 3, we propose our joint models with the shared parameters assuming the shared random effects and the correlated random effects for both the longitudinal outcomes. We illustrate the dynamic prediction of event probabilities within a fixed time. In Chapters 4 and 5, we present the application of the proposed joint models to maintenance therapies in a late-life depression and mortality in idiopathic pulmonary fibrosis outcomes study, respectively. In Chapter 6, simulation studies are performed using the proposed method in the joint model with shared random effects and also with correlated random effects. Chapter 7 concludes with a discussion.

2.0 REVIEW OF LITERATURE

In this chapter, an extensive literature review is presented. In section 2.1, we review key literature on statistical models and methods in the analysis of joint models for multivariate longitudinal data with different types of outcomes. In section 2.2, we review models that include longitudinal and time-to-event data. Finally, in section 2.3, we review methods for predictive modeling of time-to-event probabilities.

2.1 MODELS FOR MULTIVARIATE LONGITUDINAL DATA WITH DIFFERENT TYPES OF OUTCOMES

Longitudinal studies typically involve following one or more cohorts of subjects or experimental units repeatedly over two or more time points. *Multivariate* longitudinal studies are comprised of repeated responses each of which consists of two or more elements. In a multivariate longitudinal model, there are two types of correlations. One, called serial correlation, is between observations at different time points within a subject and the other, called cross correlation, is between observations on different response variables at each time point. If different types of outcomes are measured at each time point, the correlation structure is more complicated and hence, more difficult for drawing inference. Separate analyses of the different types of outcomes can lead to biased inferences because of those correlations. Therefore, it is more desirable to

jointly model multivariate outcome variables of different types together. As many studies measure multiple response outcomes of different types for each subject repeatedly, there are many approaches to model the different outcomes jointly (Olkin and Tate, 1961; Zeger and Liang, 1986; Lauritzen and Wermuth, 1989; Liu, Daniels, and Marcus, 2010; Molenberghs et al., 2010).

There are two general approaches for modeling multivariate longitudinal observations with differing outcome types. One proposed method for formulating the joint distribution of different types of outcomes is to model the relationship between the different outcomes using random effects. In this approach, different mixed models for each outcome are joined by imposing a common distribution for their random effects. It allows their model-specific random effects to be correlated, and this model allows for flexible correlation patterns. This model has a disadvantage of the high-dimensionality of the vector of random effects as the number of outcome variables gets large.

Another approach is using the product of the marginal distribution of one of the responses and the conditional distribution of the remaining response given the other response, that is,

$$f(y_{\text{continuous}}, y_{\text{discrete}}) = f(y_{\text{continuous}}) f(y_{\text{discrete}} | y_{\text{continuous}})$$

$$= f(y_{\text{discrete}}) f(y_{\text{continuous}} | y_{\text{discrete}}).$$

Here, $f(\cdot)$ denotes the probability density functions associated with the outcomes. In the conditional model, one has to choose an outcome to condition on which plays the role of a time-varying covariate. Thus, two possible types of models can lead to very different results depending on whether the conditioning variable is a discrete or a continuous outcome. The main disadvantages with conditional modeling approach are that it is hard to get easy expressions for

the association between both continuous and discrete outcomes, and that it does not directly lead to marginal inference. Also, if we have more than two outcomes, there will be many more possible factorizations instead of only the two associated with two outcomes. Hence, a conditional model is often not the preferred choice for an analysis of high-dimensional multivariate longitudinal data.

Catalano and Ryan (1992) described a joint distribution for bivariate clustered binary and continuous outcomes by factorizing the marginal distribution of a continuous outcome and a conditional distribution of a binary outcome given the continuous outcome. They used the concept of a latent variable. The type of latent variable used by Catalano and Ryan supposed that an unobserved continuous variable underlies the observed binary variable. Hence, they assume that a binary outcome results from dichotomizing the continuous latent variable. Latent variable models are useful to derive the distribution of a discrete outcome using a known CDF (Roeder, Lynch, and Nagin, 1999). Accordingly, they used a linear link function for the marginal distribution of the continuous outcome and used a correlated probit model for the conditional distribution of the binary outcome. They considered the following bivariate model:

$$y_{1ij} = \alpha_0 + \alpha_1 x_i + \varepsilon_{1ij}$$

$$y^*_{2ij} = \beta_0 + \beta_1 x_i + \varepsilon_{2ij}$$

where y_{1ij} and y^*_{2ij} denote the continuous variable and the unobserved latent variable corresponding to a binary variable for the subject i at time point j, respectively; α_0 and β_0 are the intercepts, $\alpha_1$and β_1 are effects of the covariate x, and ε_{1ij} and ε_{2ij} are correlated error terms

assumed to be mean-zero normally distributed. Let y_{2ij} be the corresponding observed binary variable determined by the latent variable y_{2ij}^*, such that

$$y_{2ij} = \begin{cases} 1 & \text{if } y_{2ij}^* > 0 \\ 0 & \text{otherwise .} \end{cases}$$

Catalano and Ryan represented the joint distribution of y_{1ij} and y_{2ij} as a product of the marginal and conditional distributions,

$$f_{y_{1ij},y_{2ij}}(y_1, y_2) = f_{y_{1ij}}(y_1) f_{y_{2ij}|y_{1ij}}(y_2|y_1).$$

The continuous outcome, y_{1ij}, is normally distributed. For the conditional distribution of y_{2ij} given y_{1ij}, they used probit link function,

$$P(y_{2ij} = 1|y_{1ij}, x_i) = \Phi\left(\frac{\mu_1}{\sqrt{\sigma_2^2(1-\tau^2)}}\right),$$

where

$$\mu_1 = \beta_0 + \beta_1 x_i + \left(\frac{\sigma_2}{\sigma_1}\right)\tau e_{1ij},$$

$$e_{1ij} = y_{1ij} - (\alpha_0 + \alpha_1 x_i),$$

σ_1^2 and σ_2^2 are variances of ε_{1ij} and ε_{2ij}, respectively, and τ is the correlation between ε_{1ij} and ε_{2ij}. They reparameterized the probit model to the more parsimonious and fully estimable form:

$$P(y_{2ij} = 1|y_{1ij}, x_i) = \Phi(\beta_0^* + \beta_1^* x_i + \beta_2^* e_{1ij}).$$

These formulas are based on the assumption of independence for all subjects i and times j. They generalized the covariance to allow for separate within-subject correlations for each outcome variable (serial correlations) and the correlation between two outcome variables for different observations in the same subject (cross correlation). They used a generalized estimating equation (GEE) method to fit the marginal and conditional distributions. Catalano (1997) extended this joint model for ordered categorical and continuous outcomes.

Fitzmaurice and Laird (1995) also considered a similar approach for modeling the joint distribution for continuous and discrete outcomes but reversed the conditioning order so that conditioning was now on the discrete outcome. In their model, the marginal distribution of the binary outcome was related to covariates using a logit link function, while the conditional distribution of the continuous outcome was related to covariates using a linear link function. In contrast to the model of Catalano and Ryan (1992), the regression parameters have marginal interpretations of the Fitzmaurice and Laird model for both the binary and continuous outcomes. The regression parameters can have a marginal interpretation only if the conditional mean of the binary outcome is related to the covariates by a linear link function (Cox and Wermuth, 1992). Also, in this latter model, maximum likelihood estimates of the regression parameters are consistent regardless of whether the model for the association between the binary and continuous outcomes has been correctly specified. By reversing the conditioning variable they evade both the lack of a marginal interpretation and the lack of robustness to misspecification.

To develop the Fitzmaurice and Laird model, we first consider a simplified model without repeated measurements. Let y_{1i} and y_{2i} denote a continuous and a binary outcome, respectively. In addition, $\boldsymbol{x}_i$ will denote a $1 \times P$ covariate vector for each subject i. Fitzmaurice and Laird made the assumption that the marginal distribution of y_{2i} is Bernoulli, that is,

$$f(y_{2i}|\boldsymbol{x}_i) = \exp[y_{2i}\theta_i - \log\{1 + \exp(\theta_i)\}],$$

where

$$\theta_i = \log\left\{\frac{\mu_{1i}}{(1-\mu_{1i})}\right\} = \boldsymbol{x}_i\boldsymbol{\beta}_1$$

and $\mu_{1i} = E[y_{2i}] = \Pr(y_{2i} = 1|\boldsymbol{x}_i, \boldsymbol{\beta}_1)$is the probability of success and $\boldsymbol{\beta}_1$ is a $P \times 1$ vector of marginal parameters. They wrote the joint distribution of (y_{1i}, y_{2i}) as

$$f_{y_{1i},y_{2i}}(y_{1i}, y_{2i}) = f_{y_{2i}}(y_{2i})f_{y_{1i}|y_{2i}}(y_{1i}|y_{2i}).$$

They also assumed that the distribution of y_{1i} given y_{2i} is normal, i.e., the pdf is given by

$$f_{y_{1i}|y_{2i}}(y_{1i}|y_{2i}) = (2\pi\sigma^2)^{-1/2} \times \exp\left[-\frac{1}{2\sigma^2}\{y_{1i} - \mu_{2i} - \gamma(y_{2i} - \mu_{1i})\}^2\right],$$

where $\mu_{2i} = E[y_{1i}] = \boldsymbol{x}_i\boldsymbol{\beta}_2$ and γ is a parameter for the regression of y_{1i} on y_{2i}. Thus,

$$E[y_{1i}|y_{2i}] = \boldsymbol{x}_i\boldsymbol{\beta}_2 + \gamma(y_{2i} - \mu_{1i}).$$

Hence, both $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are regression parameter that have marginal interpretations.

Now, let us consider the model of interest, that is, one with repeated measurements. Let $\boldsymbol{y}_{1i} = (y_{1i1}, \cdots, y_{1in_i})'$ and $\boldsymbol{y}_{2i} = (y_{2i1}, \cdots, y_{2in_i})'$denote the vectors of longitudinal continuous and binary outcomes, respectively. Let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{in_i})'$ be a $n_i \times P$ matrix of covariates for each subject i assuming each element, $\boldsymbol{x}_{ik}$, also is a $P \times 1$ vector. The model for the mean is

$$\text{logit}(\text{E}[\boldsymbol{y}_{2i}]) = \boldsymbol{X}_i\boldsymbol{\beta}_1,$$

$$\text{E}[y_{1ik}|\boldsymbol{y}_{2i}] = \boldsymbol{x}_{ik}\boldsymbol{\beta}_2 + \gamma_1(y_{2ik} - \mu_{1_{ik}}) + \gamma_2 S_i,$$

where $S_i = \sum_{j=1}^{n_i} \left(y_{2ij} - \mu_{1_{ij}} \right)$. In this model, the parameters γ_1 and γ_2 induce a correlation between $\boldsymbol{y}_{2i}$ and $\boldsymbol{y}_{1i}$. Fitzmaurice and Laird used a generalized estimating equation (GEE) method for the estimation of $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma_1,$ and γ_2. Their particular focus was on the regression parameters for the marginal expectation of the outcomes rather than the association between the outcomes which was considered to be a nuisance characteristic of the model.

Both probit and logistic link functions have symmetric s-shaped cumulative distribution functions. However, the logistic places more probability in the tails of the distribution than does the normal, because the variance of the standard normal is equal to 1, the standard logistic has variance equal to $\pi^2/3$. Thus, the scale of the logistic is greater than the normal. The logistic link function is popular in many fields and interpretation is easier for logistic version.

Fieuws and Verbeke (2004) were interested in the questions of how the evolution of one outcome is related to the evolution of another outcome ('association of evolutions') and how the association between outcomes evolves over time ('evolution of the association') for longitudinal multivariate data. To get flexible solutions to such questions, they investigated a joint model using a random effects approach. In this approach, random effects were assumed for each outcome and by adopting a joint multivariate distribution for the random effects, the different outcomes were associated. They applied their linear mixed-effects model to the hearing thresholds of two frequencies, $y_{1i}(t)$ and $y_{2i}(t)$ measured over time in subject i. Hence, for subject i taken at time t, their model was

$$\begin{aligned} y_{1i}(t) &= \mu_1(t) + a_{1i} + b_{1i}t + \varepsilon_{1i}(t) \\ y_{2i}(t) &= \mu_2(t) + a_{2i} + b_{2i}t + \varepsilon_{2i}(t) \end{aligned} \tag{2.1.1}$$

where $\mu_1(t)$ and $\mu_2(t)$ indicate the mean responses. The a_i are random intercepts, the b_i are random slopes for time. A joint distribution for the random effects can tie both outcome trajectories together assuming that a_{1i}, a_{2i}, b_{1i}, and b_{2i} follow a 4-dimensional multivariate normal distribution with mean zero and variance-covariance matrix $\boldsymbol{D}$. The error terms were uncorrelated and not associated with the random effects

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right).$$

Fieuws and Verbeke obtained the parameters of the model using likelihood based inference. The above mentioned questions can be answered from the covariance matrix of the random effects. In later work, they developed the joint random effects model given in equation (2.1.1) so that outcomes of higher dimension can be accommodated by proposing a pairwise modeling approach. This is obtained by first fitting all possible pairwise bivariate models separately, instead of maximizing the likelihood of the full joint model. Then they obtained estimates for all pair-specific parameters by maximizing each of the likelihoods separately. For some parameters which have multiple estimates, for example, the covariance between random effects from the same outcome, a single estimate is obtained by averaging all corresponding pair-specific maximum likelihood estimates (Fieuws and Verbeke 2005, 2006).

Gueorguieva and Agresti (2001) used an approach similar to Catalano and Ryan (1992) for joint model. They studied a correlated probit model that applies an underlying latent normal variable for the binary outcomes but use a random effects model instead of a conditional model. The focus of their work was on the joint, subject-specific effects on the models. They used a modified Monte Carlo expectation-conditional maximization (ECM) algorithm for finding

maximum likelihood estimates for the multivariate correlated probit model. Later, this method was extended to continuous and ordinal variables (Gueorguieva and Sanacora 2006).

2.2 JOINT MODELING OF LONGITUDINAL AND TIME-TO-EVENT DATA

Phase II and III clinical trials usually use time-to-event as the primary study outcome. In many such trials, patients are also observed longitudinally with respect to potential biomarkers and clinical measurements throughout the follow-up period. Hence, joint modeling of longitudinal and time-to-event data has increasingly been developed for use in clinical trials (Hogan and Laird, 1997a, 1997b; Wang and Taylor, 2001; Xu and Zeger, 2001; Ibrahim, Chen, and Sinha, 2004; Tsiatis and Davidian, 2004; Chi and Ibrahim, 2006; Ding and Wang, 2008; Rizopoulos and Ghosh, 2011). Two areas where both longitudinal and time-to-event outcomes have been modeled jointly are in AIDS and cancer.

Tsiatis, DeGruttola, and Wulfsohn (1995) examined the relationship between the CD4 count and survival time in patients with acquired immune deficiency syndrome (AIDS). They proposed a two-stage procedure by plugging the estimates from longitudinal models into a Cox proportional hazards model. In the first stage, the longitudinal CD4 counts are modeled using a repeated measures random effects model with normal errors for true CD4 counts. In the second stage, the model value is substituted into a proportional hazard model and used Cox regression with CD4 counts as a time-dependent covariate to obtain estimates of the survival parameters. In the two-stage model, only longitudinal data is used to find the parameters for the longitudinal process and then the hazard is a function of this modeled longitudinal process. They gave

considerable thought to how the measurement error and nonrandom missingness of CD4 counts affected the model.

For describing the joint longitudinal and survival data we introduce the following notation. Let T_i denote the survival time for subject i $(i = 1, \cdots, n)$, C_i be the censoring time, and $\delta_i = I(T_i \leq C_i)$ be the event indicator. Let $T_i^* = min(T_i,\ C_i)$ be the observed event time for the subject i. The CD4 counts $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{ij})$ are measured at times $\boldsymbol{t}_i = (t_{ij} :\ t_{ij} \leq T_i^*)$, where t_{ij} is the time from randomization for measurement j on subject i, $j = 1, \cdots, n_i$. Let y_{ij}^* denote the true value of the CD4 count for subject i at time t_{ij}. The log transformed longitudinal outcome CD4 counts were modeled using a linear growth curve model with a random intercept and slope;

$$\begin{aligned} y_{ij} &= y_{ij}^* + \varepsilon_{ij} \\ &= \alpha_{0i} + \alpha_{1i} t_{ij} + \varepsilon_{ij}. \end{aligned} \tag{2.2.1}$$

where ε_{ij} is measurement error assumed to be independent and normally distributed with mean zero, and variance σ_ε^2. The measurement error terms and the random effects were assumed to be independent of each other. It is also assumed that the individual intercepts and slopes followed a bivariate normal distribution, i.e.,

$$\begin{pmatrix} \alpha_{0i} \\ \alpha_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}, \begin{bmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{01} & \sigma_{11} \end{bmatrix} \right).$$

Let $\boldsymbol{\mathcal{Y}}_i(t)$ and $\boldsymbol{\mathcal{Y}}_i^*(t)$ denote the history of observed and true CD4 counts up to time t, respectively, that is, $\boldsymbol{\mathcal{Y}}_i(t) = \{y_i(t_{i1}), \cdots, y_i(t_{ij}); t_{ij} \leq t\}$ and $\boldsymbol{\mathcal{Y}}_i^*(t) = \{y_i^*(t_{i1}), \cdots, y_i^*(t_{ij}); t_j \leq t\}$. For the survival data, the hazard of death is modeled as a function of the conditional

expectation of true CD4 counts given the history of observed CD4 counts through the proportional hazards model and a partial likelihood approach is used to obtain estimates of the survival parameters. Hence, the model relating the hazard to time-varying covariates can be written as

$$h\big(t|\boldsymbol{y}_i(t)\big) = h_0(t)E\big[f(\boldsymbol{y}_i^*(t),\beta) \mid y_i(t_{i1}),\cdots,y_i\big(t_{ij}\big),T_i^* \geq t\big],$$

where $f(\boldsymbol{y}_i^*(t),\beta)$ is a function of the covariate history specified up to an unknown parameter β. Let the conditional expectation part, $E\big[f(\boldsymbol{y}_i^*(t),\beta) \mid y_i(t_{i1}),\cdots,y_i\big(t_{ij}\big),T_i^* \geq t\big]$, be shortened to $E(t,\beta)$. Thus, the partial likelihood is

$$\prod_{i=1}^{n}\left[E_i(T_i^*,\beta)\Big/\sum_{k=1}^{n}E_k(T_i^*,\beta)I(T_i^* \geq t)\right]^{\delta_i}.$$

Wulfsohn and Tsiatis (1997) later developed this method by joint maximization of the likelihood from both the longitudinal CD4 counts and the survival data. Because their method used data from both the longitudinal and the survival data, it made more efficient use of the data. They implemented an EM algorithm to fit this model. Song, Davidian, and Tsiatis (2002a, 2002b) proposed relaxations of the normality assumption for latent process expressed by a set of random effects.

Faucett and Thomas (1996) also assumed a proportional hazards model for survival conditional on a random effects model with normal errors for the CD4 counts similar to that proposed by Tsiatis, DeGruttola, and Wulfsohn (1995), which are defined by (2.2.1). But they assumed a proportional hazards model with the parametric assumption of piecewise constant

baseline hazard and adopted a different estimation procedure using a Bayesian Markov chain Monte Carlo (MCMC) technique of Gibbs sampling to do the estimation.

Henderson, Diggle, and Dobson (2000) proposed a flexible joint model that avoids specifying the class variable. They modeled the longitudinal data including fixed effects, random effects, serial correlation, and pure measurement error. For survival data, they used a proportional hazard model with or without frailty terms. A key feature of their strategy is to connect the longitudinal and survival model with two correlated latent Gaussian processes allowing the trend to vary with time. They assumed that longitudinal and survival data are conditionally independent given the linking latent process and covariates.

There are n subjects with longitudinal measurements $\{y_{ij} : j = 1, \dots, n_i\}$ at times $\{t_{ij} : j = 1, \dots, n_i\}$. When the interval of follow-up is $[0, \tau)$, let $\{N_i(s) : 0 \leq s \leq \tau\}$ denote a counting process for the events and $\{H_i(s) : 0 \leq s \leq \tau\}$ denote an indicator for whether the subject is at risk of an event at time s. Let $W_i(t) = \{W_{1i}(t)\ ,\ W_{2i}(t)\}$ denote a latent zero-mean bivariate Gaussian process, which is realized independently in different subjects. They considered the following for longitudinal model:

$$y_{ij} = \mu_i(t_{ij}) + W_{1i}(t_{ij}) + \varepsilon_{ij},$$

where ε_{ij} is a measurement error term assumed to be mean-zero normally distributed with $\text{var}(\varepsilon_{ij}) = \sigma_\varepsilon^2$ and $\mu_i(t_{ij})$ is the mean response assumed by a linear model

$$\mu_i(t) = \boldsymbol{x}_{1i}(t)^\intercal \boldsymbol{\beta}_1$$

in which the vectors $\boldsymbol{x}_{1i}(t)$ and $\boldsymbol{\beta}_1$ represent possibly time-varying covariate and their corresponding regression coefficients, respectively. For the survival model, they considered a semi-parametric multiplicative model:

$$\lambda_i(t) = H_i(t)\alpha_0(t)\exp\{\boldsymbol{x}_{2i}(t)^{\intercal}\boldsymbol{\beta}_2 + W_{2i}(t)\},$$

with the form of $\alpha_0(t)$ left unspecified. To colligate both the longitudinal and survival models, they introduced the following flexible model for $W_{1i}(t)$:

$$W_{1i}(t) = \boldsymbol{Z}_{1i}(t)^{\intercal}\boldsymbol{U}_{1i} + V_{1i}(t),$$

where $\boldsymbol{Z}_{1i}(t)$ is a vector of covariate values, $\boldsymbol{U}_{1i}$ is a corresponding vector of random effects follows a multivariate normal distribution with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}_1$ and $V_{1i}(t)$ is a stationary Gaussian process with mean zero, variance σ_{v1}^2 and correlation function $r_1(s) = \text{cov}\{V_{1i}(t)\,, V_{1i}(t-s)\}/\sigma_{v1}^2$. Tsiatis, DeGruttola, and Wulfsohn (1995), Faucett and Thomas (1996), and Wulfsohn and Tsiatis (1997) all assume a Laird and Ware (1982) linear random effects model, $W_{1i}(t) = U_{1i} + U_{2i}t$, and assume $W_{2i}(t)$ is proportional to $W_{1i}(t)$. In contrast, Henderson, Diggle, and Dobson (2000) proposed that $W_{2i}(t)$ be specified as

$$W_{2i}(t) = \gamma_1 U_{1i} + \gamma_2 U_{2i} + \gamma_3(U_{1i} + U_{2i}t) + U_{3i},$$

where the frailty term $U_{3i} \sim N(0\,, \sigma_3^2)$ is independent of the $(U_{1i}\,, U_{2i})$. The parameters $\gamma_1, \gamma_2,$ and γ_3 in this model measure the association between the longitudinal and survival models induced through the random intercepts, slopes, and current value of W_{1i} at time t, respectively. They used an EM algorithm proposed by Wulfsohn and Tsiatis (1997) for estimation. In later work, they focused on the use of longitudinal outcome trends as individual-level surrogates for survival time.

They developed a score test for association between longitudinal outcome and survival time (Henderson, Diggle, and Dobson 2002).

Guo and Carlin (2004) used the flexible joint model proposed by Henderson, Diggle, and Dobson (2000). Consequently, the longitudinal and survival outcomes were assumed to be independent given a linking latent bivariate Gaussian process and available covariates. Their joint model avoids specification of a class variable. They developed this model using a fully Bayesian approach through MCMC methods using the WinBUGS software. They compared their results to those obtained from the traditional maximum likelihood approach for this joint model. Also they compared the results from a joint model to separate models.

Rizopoulos, Verbeke, and Molenberghs (2008) developed a shared parameter model for the joint distribution of longitudinal data and time-to-event data. Shared parameter models assume that the longitudinal and survival processes are conditionally independent, given the random effects. This assumption implies that all associations between the longitudinal and survival processes are induced by the random effects. They proposed two separate sets of random effects for the longitudinal and survival processes, linking them using a copula function which was specified as a cumulative multivariate distribution function with a uniform marginal.

Again, let T_i denote the true event time for subject i $(i = 1, \cdots, n)$, C_i be the censoring time, and $T_i^* = \min(T_i,\ C_i)$ be the observed event time for the subject i. Define the event indicator as $\delta_i = I(T_i \leq C_i)$. Let $\boldsymbol{y}_i = \{y_i(t_{ij}), j = 1, \ldots, n_i\}$ denote the observed longitudinal outcome for a subject i taken at time t_{ij}. Denote the time invariant random effects underlying both the longitudinal and survival model by $\boldsymbol{b}_i$. The shared parameter model is

$$p(\boldsymbol{y}_i, T_i^*, \delta_i\,;\,\boldsymbol{\theta}) = \int p\big(\boldsymbol{y}_i|\boldsymbol{b}_i\,;\,\boldsymbol{\theta}_y\big)p(T_i^*, \delta_i|\boldsymbol{b}_i\,;\,\boldsymbol{\theta}_t)p(\boldsymbol{b}_i\,;\,\boldsymbol{\theta}_b)d\boldsymbol{b}_i,$$

where $\boldsymbol{\theta}^{\intercal} = \big(\boldsymbol{\theta}_y^{\intercal}, \boldsymbol{\theta}_t^{\intercal}, \boldsymbol{\theta}_b^{\intercal}\big)$ is the vector containing the parameters of each one of the submodel. Both the longitudinal and survival processes are assumed noninformative, i.e. independent of $\boldsymbol{b}_i$. They proposed that the conditional submodel for longitudinal outcome $\boldsymbol{y}_i$ has the form

$$\boldsymbol{y}_i|\boldsymbol{b}_i \sim N\big(\boldsymbol{X}_{yi}\boldsymbol{\beta} + \boldsymbol{Z}_{yi}\boldsymbol{b}_{yi},\ \ \sigma_y^2\boldsymbol{I}_{n_i}\big),$$

where $\boldsymbol{X}_{yi}$ and $\boldsymbol{Z}_{yi}$ are known fixed and random effects design matrices, respectively. For the conditional submodel of survival outcome T_i^*, they assumed that $\log T_i^*\,|\boldsymbol{b}_i$ follows a parametric distribution with $\mathrm{E}[\log T_i^*\,|\boldsymbol{b}_i] = \boldsymbol{x}_{ti}^{\intercal}\gamma + \boldsymbol{b}_{ti}$ where $\boldsymbol{x}_{ti}$ is a vector of covariates for the survival outcoem. To connect the two separate sets of random effects for the longitudinal and survival processes they used a copula function of the general form:

$$p\big(\boldsymbol{b}_{yi}, \boldsymbol{b}_{ti}\big) = c\big\{F_y\big(\boldsymbol{b}_{yi}\big), F_t(\boldsymbol{b}_{ti})\,;\,\alpha\big\}p\big(\boldsymbol{b}_{yi}\big)p(\boldsymbol{b}_{ti}),$$

where α is a parameter of the copula function and $F_y(\cdot)$ and $F_t(\cdot)$ are the marginal cumulative distribution functions for $\boldsymbol{b}_{yi}$ and $\boldsymbol{b}_{ti}$ respectively. They investigated the impact of misspecifying random-effects but also how affected the estimation of standard errors would be under the misspecified model. Rizopoulos (2010) developed an R package JM that fits a variety of joint models for normal longitudinal outcome and time-to-event data under maximum likelihood based on the shared parameter model.

2.3 PREDICTIVE MODELING OF TIME-TO-EVENT PROBABILITY

The major purpose of prediction is to provide estimates, either point or interval, for future measurements based on the results obtained from previous observations (Dunsmore 1974). Predictive models are particularly useful in a clinical setting (Putter et al., 2006). A number of authors have proposed the prediction of future event probabilities for subjects based on the joint modeling of longitudinal measurements, time-to-event outcome, and other covariates (Taylor, Yu, and Sandler, 2004; Garre et al., 2008; Proust-Lima and Taylor, 2009).

Rizopoulos (2011) provided individualized prediction models of survival in AIDS patients who also had longitudinal CD4 cell count measurements. He specified the joint model of longitudinal and survival processes by

$$y_i(t) = \boldsymbol{x}_i^{\intercal}(t)\boldsymbol{\beta} + \boldsymbol{z}_i^{\intercal}(t)\boldsymbol{b}_i + \varepsilon_i(t), \qquad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2),$$

$$\lambda_i(t) = \lambda_0(t)\exp[\boldsymbol{d}_i^{\intercal}\gamma + \alpha\{\boldsymbol{x}_i^{\intercal}(t)\boldsymbol{\beta} + \boldsymbol{z}_i^{\intercal}(t)\boldsymbol{b}_i\}],$$

where $\boldsymbol{x}_i(t)$ and $\boldsymbol{z}_i(t)$ denote column vectors of time-varying covariates corresponding to fixed and random effects, respectively, $\boldsymbol{d}_i$ is a vector of baseline covariates, and $\boldsymbol{b}_i$ is the vector of random effects. In this model, longitudinal measurements on a variable represent an endogenous time-dependent covariate (Kalbfleisch and Prentice 2002). The model implies that longitudinal measurements are directly related to the survival process. Hence, longitudinal measurements on patients up to time t implies the patients' survival up to time t. Thus, he focused on the conditional probability of surviving time $t^* > t$ given survival up to t, that is

$$\pi_i(t^*|t) = \Pr(T_i \geq t^* | T_i > t, \boldsymbol{\mathcal{Y}}_i(t), \mathcal{D}_n\,;\, \theta),$$

where $\boldsymbol{y}_i(t) = \{y_i(s); 0 \leq s \leq t\}$ denotes a set of longitudinal measurements and $y_i(t)$ denotes the value of the longitudinal outcome at time point t for subject i, and $\mathcal{D}_n = \{T_i^*, \delta_i, y_i ; i = 1, \cdots, n\}$ denotes the sample on which the joint model is fitted and on which the prediction model is based. As before, T_i^* is observed event time, T_i is true event time, and δ_i is the event indicator. Rizopoulos derived a first-order estimate of $\pi_i(t^*|t)$ using the empirical Bayes estimate for $\boldsymbol{b}_i$. To produce valid standard errors for the estimate of $\pi_i(t^*|t)$, a standard asymptotic Bayesian formulation (Cox and Hinkley, 1974, Section 10.6) was used. Accordingly, he obtained a Monte Carlo estimate of $\pi_i(t^*|t)$. To assess the predictive performance of time-to-event models, estimates of sensitivity and specificity measures were derived under the joint modeling framework. Based on these estimates they presented ROC curves and AUC estimates of prediction performance.

Fieuws et al. (2008) investigated predicting renal graft failure using multiple longitudinal outcomes of biochemical and physiological markers. To handle prediction from multiple correlated longitudinal outcomes, they used multivariate mixed model (MMM). They obtained a probability function of the risk that the graft will fail within a 10-year period after transplantation using Bayes rule as

$$H_i(t) = P_i\left(t \leq T_i \leq 120 \middle| \boldsymbol{y}_i(t)\right)$$

$$= \frac{f_i(\boldsymbol{y}_i(t)|t \leq T_i \leq 120)P(T_i \leq 120|T_i \geq t)}{f_i(\boldsymbol{y}_i(t)|t \leq T_i \leq 120)P(T_i \leq 120|T_i \geq t) + f_i(\boldsymbol{y}_i(t)|T_i > 120)P(T_i > 120|T_i \geq t)}$$

where t is time expressed in months and ranges from 12 to 120, T_i denotes the time of failure, and $\boldsymbol{y}_i(t)$ is a vector containing the outcome information collected since 1 year after the transplantation up to and including time, t, for subject i. These probabilities are computed using

the MMM used for the markers. Because of the computational complexity of MMM, they used all possible pairwise mixed models as proposed in their earlier paper (Fieuws and Verbeke, 2006). All pairwise mixed models were fit instead and results averaged across fits. A pattern-mixture approach (Little 1993; Molenberghs and Verbeke 2005) was used to factorize the joint model for the longitudinal outcomes and the time-to-event outcome.

Pauler and Finkelstein (2002) specified a parametric marginal model for longitudinal prostate-specific antigen (PSA) outcomes using a subject-specific change point that allowed for PSA slopes to change and hence, jointly model the data. For the time-to-event outcome, recurrence of prostate cancer, a Bayesian version of Cox proportional hazard model was assumed. The posterior predictive distribution of the time-to-event based on the joint probability model was calculated as

$$
\begin{aligned}
& p(t^*|\boldsymbol{y}^*, \overline{\mathfrak{D}}) \propto p(t^*, \boldsymbol{y}^*|\overline{\mathfrak{D}}) \\
&= \int p(t^*, \boldsymbol{y}^*, \boldsymbol{b}^*, \boldsymbol{b}|\overline{\mathfrak{D}})\, \mathrm{d}\boldsymbol{b}^* \mathrm{d}\boldsymbol{b} \\
&= \int p(t^*|\boldsymbol{b}^*, \boldsymbol{b}) p(\boldsymbol{y}^*|\boldsymbol{b}^*, \boldsymbol{b}) p(\boldsymbol{b}^*|\, \boldsymbol{b}) p(\boldsymbol{b}|\overline{\mathfrak{D}})\, \mathrm{d}\boldsymbol{b}^* \mathrm{d}\boldsymbol{b}
\end{aligned}
$$

where t^* denotes the time-to-event for the patient under consideration with an additional (or new) vector of longitudinal outcomes $\boldsymbol{y}^*$, $\overline{\mathfrak{D}}$ denotes data from all previous patients for which posterior distributions are available, $\boldsymbol{b}$ denotes the patient-specific random effects vector, and $\boldsymbol{b}^*$ denotes the new random effects vector for the additional subject. To estimate prediction probabilities Markov chain Monte Carlo method was used.

Yu, Taylor, and Sandler (2004) performed individual prediction using a longitudinal-survival-cure model (Law, Taylor, and Sandler 2002; Yu, Law, Taylor, and Sandler 2004). They

divided the patients into “cured” or “susceptible” groups depending on whether the patients have their tumor completely killed by the treatment or not. This aspect of the study was incorporated into the joint modeling by using mixture cure models (Farewell 1982; Kuk and Chen 1992; Taylor 1995). The cured fraction was modeled as a logistic function of baseline covariates, measured before the end of the radiation therapy period. The longitudinal outcome was modeled using non-linear hierarchical mixed models with different models for the cured and susceptible groups. The time-to-event outcome was modeled using a time-dependent proportional hazards model for those in the susceptible group where the time dependent covariate include both the current value and the slope of post-treatment longitudinal outcome profile. Estimates of the parameters in the model were obtained by using MCMC method (Lockwood, and Schervish, 2005). The posterior distributions for all the parameters were obtained from the product of full complete data likelihood and prior distributions. Model predictions were approximated using posterior distribution based on the observed data. This model has the disadvantage that it is highly parameterized. With such complicated modeling, interpretation of the parameters can be difficult. The number of parameters to be estimated in such cure models can lead to identifiability problems (Farewell 1986; Li, Taylor, and Sy 2001).

3.0 JOINT MODELING APPROACH

3.1 MODEL DEFINITIONS FOR MULTIVARIATE LONGITUDINAL DATA WITH DIFFERENT TYPES OF OUTCOMES

In this chapter, we first describe the bivariate longitudinal setting with a single continuous response and a single binary response which is used to model a time-to-event outcome. The generalization to higher dimensions and other members of the exponential family of distributions is conceptually straightforward. Let y_{1ij} and y_{2ij} denote the j^{th} outcome which consists of continuous and binary components for subject i, respectively. Further, let $\boldsymbol{y}_i = (\boldsymbol{y}_{1i}, \boldsymbol{y}_{2i})^{\intercal}$ denote the complete bivariate longitudinal outcome vector for the subject i, where $\boldsymbol{y}_{hi} = \left(y_{hi1}, \cdots, y_{hij}\right)^{\intercal}, h = 1,2,\ j = 1, \cdots, n_{hi}$ is a vector of h^{th} longitudinal outcome at time point j. For the longitudinal bivariate outcome vector, $\boldsymbol{y}_i$, with different data types, we assume an underlying generalized linear mixed effects model which can be written as

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i \\ &= g(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i) + \boldsymbol{\varepsilon}_i \end{aligned} \tag{3.1.1}$$

where $g(\cdot)$ denotes a known one-to-one link function that is allowed to change with the characteristics of the different types of outcomes in $\boldsymbol{y}_i$, and $\boldsymbol{X}_i$ and $\boldsymbol{\beta}$ represent a design matrix of known covariate values and a vector of their corresponding regression coefficients, respectively.

Also, the $\boldsymbol{b}_i$ are random effects that are assumed to follow a normal distribution with mean zero and variance-covariance matrix $\boldsymbol{G}$, and the $\boldsymbol{Z}_i$ are design matrices for the random effects. It is assumed that the elements in each outcome in $\boldsymbol{y}_i$ are independent conditional on $\boldsymbol{b}_i$. Finally, $\boldsymbol{\varepsilon}_i$ is a vector of measurement error terms. More specifically, by choosing the identity link for continuous outcome and the logit link for the binary outcome, the generalized linear mixed effects model (3.1.1) can be written in the form

$$\begin{aligned} y_{1ij} &= \boldsymbol{x}_{1ij}^{\intercal}\boldsymbol{\beta}_1 + \boldsymbol{z}_{1ij}^{\intercal}\boldsymbol{b}_{1i} + \varepsilon_{1ij}, \\ y_{2ij} &= \frac{\exp\left(\boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 + \boldsymbol{z}_{2ij}^{\intercal}\boldsymbol{b}_{2i}\right)}{1+\exp\left(\boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 + \boldsymbol{z}_{2ij}^{\intercal}\boldsymbol{b}_{2i}\right)} + \varepsilon_{2ij}, \end{aligned} \tag{3.1.2}$$

where ε_{1ij} and ε_{2ij} are independent (Fitzmaurice, Laird, and Ware, 2004; Fitzmaurice, Davidian, Verbeke, and Molenberghs, 2009). We consider two types of structures for the random effects $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$. First, we assume that $\boldsymbol{b}_{1i}$ follows a normal distribution with a mean vector of zeros and variance-covariance matrix $\boldsymbol{\Sigma}$, and that $\boldsymbol{b}_{2i}$ is proportional to $\boldsymbol{b}_{1i}$, i.e., $\boldsymbol{b}_{2i} = \boldsymbol{A}_0\boldsymbol{b}_{1i}$, where $\boldsymbol{A}_0$ is a diagonal matrix of unknown constants. We call a joint model with this assumption a *shared random effects joint model*. Second, we assume a joint distribution that $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$ follow multivariate normal distribution as;

$$\begin{pmatrix}\boldsymbol{b}_{1i}\\ \boldsymbol{b}_{2i}\end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{pmatrix},\ \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{12}^{\intercal} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right).$$

We call a joint model with the second assumption a *correlated random effects joint model*. The relationship between the continuous and binary outcomes can be investigated through the correlation among the random effects $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$.

3.2 PROPOSED JOINT MODEL: RELATING A TIME-TO-EVENT OUTCOME TO MULTIVARIATE LONGITUDINAL DATA OF DIFFERENT TYPES

As introduced in section 2.2, Henderson, Diggle, and Dobson (2000) proposed a joint model of longitudinal measurements and event time data through a latent zero-mean bivariate Gaussian process, using an EM algorithm for estimation. Guo and Carlin (2004) applied Bayesian version of this joint modeling strategy. They used only a single longitudinal continuous outcome in the joint model. We developed their methods to multivariate longitudinal data of different types and a time-to-event outcome. As mentioned in section 2.2, let T_i denote the true event time for subject i $(i = 1, \cdots, n)$, C_i be the censoring time, and $\delta_i = I(T_i \leq C_i)$ be the event indicator. Let $T_i^* = \min(T_i,\ C_i)$ be the observed event time for the subject i. We assume that censoring is non-informative. A proportional hazard model is assumed, that is,

$$\lambda_i(t) = \lim_{dt \to 0} \frac{\Pr\{t \leq T_i < t + dt \mid T_i \geq t\}}{dt}$$

$$= \lambda_0(t) \exp\{\boldsymbol{x}_{3i}^{\mathrm{T}} \boldsymbol{\beta}_3 + U_i\}, \qquad (3.2.1)$$

where $\boldsymbol{x}_{3i}$ is a vector of covariates, $\boldsymbol{\beta}_3$ is regression coefficients of covariates, and $\lambda_0(t)$ is baseline hazard function, which can be assumed to be of a parametric form or left unspecified. To express the effects of longitudinal outcomes on the time-to-event, we assume that shared parameters, U_i, is associated with the random effects of longitudinal outcomes $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$. There are three submodels of the longitudinal continuous outcome process, the longitudinal binary outcome process, and the time-to-event outcome process presented as (3.1.2) and (3.2.1).

Our proposed joint model connects (3.1.2) and (3.2.1) by taking

$$U_i = \boldsymbol{a}_1^{\intercal}\boldsymbol{b}_{1i} + \boldsymbol{a}_2^{\intercal}\boldsymbol{b}_{2i} + b_{3i}, \tag{3.2.2}$$

where $\boldsymbol{a} = (\boldsymbol{a}_1^{\intercal}, \boldsymbol{a}_2^{\intercal})$ is a set of unknown constants and b_{3i} is a normally distributed frailty term with mean zero and variance σ_3^2, independent of the $\boldsymbol{b}_i = (\boldsymbol{b}_{1i}^{\intercal}, \boldsymbol{b}_{2i}^{\intercal})^{\intercal}$. The hazard function of the time-to-event depends on the longitudinal outcomes through the shared $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$. Thus, the parameter $\boldsymbol{a}$ in the survival model (3.2.2) measures a degree of association explained by the random effects in (3.1.2). The components of $\boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{x}_3$ may not all the same, which allows that the longitudinal continuous and binary outcomes, and the time-to-event outcome to depend on different and/or overlapping covariate information. We assume that the longitudinal outcome vector, $\boldsymbol{y}_i$, and the time-to-event outcome, T, are independent conditional on covariates $\boldsymbol{X}, \boldsymbol{Z}$, and the random effects, $\boldsymbol{b}$. The observed data for the i^{th} subject with n_{hi} repeated measurements for the h^{th} outcome are denoted by

$$\{y_{hi}(t_{ij}),\ X(t_{ij}),\ Z(t_{ij}),\ T_i^*,\ \delta_i\ ;\ h = 1,2,\ \ i = 1,\cdots,n,\ \ j = 1,\cdots,n_{hi}\}.$$

Based on a full conditional independence assumption, we can express the joint distribution of the observed data as

$$f(\boldsymbol{y}_i, T_i^*, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta}) = f\left(\boldsymbol{y}_{1i}\middle|\boldsymbol{b}_{1i}; \boldsymbol{\beta}_1, \sigma_{\varepsilon_1}^2\right) f(\boldsymbol{y}_{2i}|\boldsymbol{b}_{2i}; \boldsymbol{\beta}_2) f(T_i^*, \delta_i|\boldsymbol{b}_i; \boldsymbol{\beta}_3, \boldsymbol{a}, \sigma_3^2)$$

where $\boldsymbol{\theta}$ denotes the complete parameter vector and $f(\cdot)$ denotes a probability density function.

Thus, we can derive the log-likelihood for the observed data as

$$\sum_{i=1}^{n} \log \int_{\boldsymbol{b}_i} \prod_{j=1}^{n_{1i}} \left[\frac{1}{\sqrt{2\pi\sigma_{\varepsilon_1}^2}} \exp\left\{ -\frac{1}{2\sigma_{\varepsilon_1}^2} \left(y_{1ij} - \boldsymbol{x}_{1ij}^{\intercal}\boldsymbol{\beta}_1 - \boldsymbol{z}_{1ij}^{\intercal}\boldsymbol{b}_{1i} \right)^2 \right\} \right]$$

$$\times \prod_{j=1}^{n_{2i}} \left[\left\{ \frac{1}{1+\exp\left(-\boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 - \boldsymbol{z}_{2ij}^{\intercal}\boldsymbol{b}_{2i}\right)} \right\}^{y_{2ij}} \left\{ \frac{\exp\left(-\boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 - \boldsymbol{z}_{2ij}^{\intercal}\boldsymbol{b}_{2i}\right)}{1+\exp\left(-\boldsymbol{x}_{2ij}^{\intercal}\boldsymbol{\beta}_2 - \boldsymbol{z}_{2ij}^{\intercal}\boldsymbol{b}_{2i}\right)} \right\}^{1-y_{2ij}} \right]$$

$$\times \left[\lambda_0(T_i^*)\exp\{\boldsymbol{x}_{3i}^{\intercal}\boldsymbol{\beta}_3 + U_i\}\right]^{\delta_i} \exp\left\{ -\int_0^{T_i^*} \lambda_0(s)\exp\{\boldsymbol{x}_{3i}^{\intercal}\boldsymbol{\beta}_3 + U_i\} ds \right\}$$

$$\times f(\boldsymbol{b}_i|\boldsymbol{\Sigma}) d\boldsymbol{b}_i$$

where $f(\boldsymbol{b}_i|\boldsymbol{\Sigma})$ is the density function of $\boldsymbol{b}_i$ conditional on the covariance parameters **Σ.**

3.3 ESTIMATION AND MODEL SELECTION

We used a Bayesian approach for estimating parameters of the proposed joint model through MCMC methods. The methods here are programmed in the R interface called ‘rbugs’ which accessed the software ‘OpenBUGS’. BUGS is an acronym for ‘Bayesian inference Using Gibbs Sampling’. Among the two most common versions of BUGS, namely WinBUGS and OpenBUGS, we used the more recent version, namely, OpenBUGS. This is the version that has been designated as that for which future versions will be based on.

We considered many different forms of the random effects terms in the joint model ranging from the simplest model with no random effects to the largest model with random intercepts and random slopes in both the longitudinal outcomes and the time-to-event outcome.

Historically, there have been two well-known information criteria for model selection, namely, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Spiegelhalter, et al. (2002) proposed the Deviance Information Criterion (DIC) which can be viewed as the Bayesian version of the AIC. The DIC is a generalization of the AIC for hierarchical models based on the *deviance* of the posterior distribution. For the parameter vector, $\boldsymbol{\theta}$, and observed data vector, $\boldsymbol{y}$, let $D(\boldsymbol{\theta})$ be the deviance,

$$D(\boldsymbol{\theta}) = -2 \log f(\boldsymbol{y}|\boldsymbol{\theta}) + 2 \log h(\boldsymbol{y}),$$

where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function and $h(\boldsymbol{y})$ is a standardizing function of the data alone (Carlin and Louis, 2009). We define $D(\boldsymbol{\theta})$ to be $D(\boldsymbol{\theta}) = -2 \sum_{i=1}^{n} \log f(\boldsymbol{y}_i, T_i^*, \delta_i; \boldsymbol{\theta})$ in our joint model. $\overline{D(\boldsymbol{\theta})} = E_{\theta|y}[D(\boldsymbol{\theta})]$ denotes the posterior expected deviance and $\overline{\boldsymbol{\theta}} = E_{\theta|y}[\boldsymbol{\theta}]$ denotes the posterior means of the parameters. The effective number of parameters, p_D, which can capture the complexity of a model is defined as

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}}).$$

Then, the DIC is defined as

$$\text{DIC} = D(\overline{\boldsymbol{\theta}}) + 2\, p_D$$

$$= \overline{D(\boldsymbol{\theta})} + p_D.$$

Smaller values of the DIC indicate better fitting models. As with the AIC, differences of the DIC between models are a tool used for model selection. Differences of 3 to 5 are considered to be meaningful. Conveniently, OpenBUGS provides the DIC values after running an MCMC. In this dissertation, we use the DIC for model selection criterion.

3.4 PREDICTION OF AN EVENT PROBABILITY

In this section, we illustrate how predicted future event probabilities are calculated based on our proposed joint model of the longitudinal outcomes and the time-to-event outcome. We applied the dynamic prediction of a clinical event occurring within a fixed window for a subject still at risk just before time t as proposed by van Houwelingen and Putter (2011). Specifically, we considered the conditional survival function of time $v \geq t$ given that one survives up to time t:

$$S(v|t) = P(T > v|T \geq t, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}; \boldsymbol{\theta}, v \geq t\ ) = \frac{S(v|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}; \boldsymbol{\theta}, v \geq t)}{S(t - |\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}; \boldsymbol{\theta})}. \qquad (3.4.1)$$

Equation (3.4.1) can also be expressed using a hazard function as

$$S(v|t) = \exp[-\{\Lambda(v|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}; \boldsymbol{\theta}) - \Lambda(t|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}; \boldsymbol{\theta})\}]$$

$$= \exp\left(-\int_t^v \lambda(s|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{b}; \boldsymbol{\theta})ds\right) \qquad (3.4.2)$$

where $\Lambda(\cdot)$ denotes the cumulative hazard function and $\lambda(\cdot)$ denotes the instantaneous hazard function. Equation (3.4.2) implies that only the hazard on the interval $[t, v]$ is necessary to predict the probability of event up to time v for a subject at risk just before time t. Let $v = t + w$ be a fixed window of width w. We can relate the survival function to the cumulative distribution function (c.d.f.) as follows:

$$F_w(t) = 1 - S(t + w|t).$$

This function is called the *fixed width failure function*. It is evaluated at all-time points t where the estimates change value.

The variance of this function is based on the Nelson-Aalen estimate of that cumulative hazard that is given as

$$se^2\left[\ln\left\{\hat{F}_w(t)\right\}\right] = \sum_{\substack{t \le t_i \le t+w \\ t_i \in \mathfrak{D}}} \frac{1}{R(t_i)^2}$$

where $\mathfrak{D}$ denotes the set of event times and $R(t)$ denotes the size of the risk set, i.e., the number of subjects with no event and still being followed just before time t.

The dynamic prediction of probabilities of event occurring within a fixed window given all baseline covariates and longitudinal measurements is applied to IPF data in chapter 5 with survival function obtained from our proposed joint models.

4.0 APPLICATION TO MAINTENANCE THERAPIES IN A LATE-LIFE DEPRESSION STUDY

4.1 DESCRIPTION OF DATASET

In this chapter, we present the analysis of the maintenance therapies in a late-life depression study, introduced in Chapter 1, using the proposed joint model for mixed multivariate longitudinal responses and time-to-event. We are mainly interested in associating recurrence of major depression with the longitudinal outcomes of neuropsychological functioning and C-IADL. From the 130 patients considered in the study, 30 patients relapsed with major depression that corresponds to a 23% recurrence rate. For full details regarding the conduct of the trial the reader is referred to Reynolds et al. (2011). Patient characteristics are summarized in Table 1. Of the 130 patients, 77% of the patients were female and 44% were adjudicated to have mild cognitive impairment (MCI) with or without amnesia. Almost 90% of the patients were white. Roughly, 51% of patients were assigned to maintenance antidepressant pharmacotherapy and donepezil, and 49% were assigned to maintenance antidepressant pharmacotherapy and placebo. The average age of the cohort was 73.5 ± 6.15 years.

Table 1. Patient characteristics of MTLD data.

		N(=130)	%
Gender			
	Male	30	23.1
	Female	100	76.9
Race			
	White	117	90.0
	Black	13	10.0
Base Diagnosis			
	MCI Amnestic - Multiple Domain	30	23.1
	MCI Amnestic - Single Domain	5	3.8
	MCI Non-Amnestic - Multiple Domain	11	8.5
	MCI Non-Amnestic - Single Domain	11	8.5
	No Cognitive Disorder	73	56.2
Treatment			
	Donepezil Hydrochloride	67	51.5
	Placebo	63	48.5
		Mean	**SD**
Age (years)			
	Overall	73.5	6.15
	Male	74.2	6.56
	Female	73.3	6.04
	Donepezil Hydrochloride	73.1	6.50
	Placebo	73.9	5.82

4.2 DATA STRUCTURE

In measuring cognitive impairment in the late-life depression example, there are five domains with 17 individual tests to assess neuropsychological functioning: language, delayed memory, executive, visuospatial, and speed of information processing domain. A global performance score is calculated by averaging over all 17 tests. Each subject had baseline scores of neuropsychological functioning and C-IADL at time 0. Then, subjects were randomized to treatment of donepezil or placebo and had repeated measurements of neuropsychological functioning and C-IADL at one and two years. Individual profiles for the global score and five

domains of neuropsychological functioning are shown in Figure 1 and Figure 2. We observed that in both treatment groups subjects show similar variability in their longitudinal profiles.

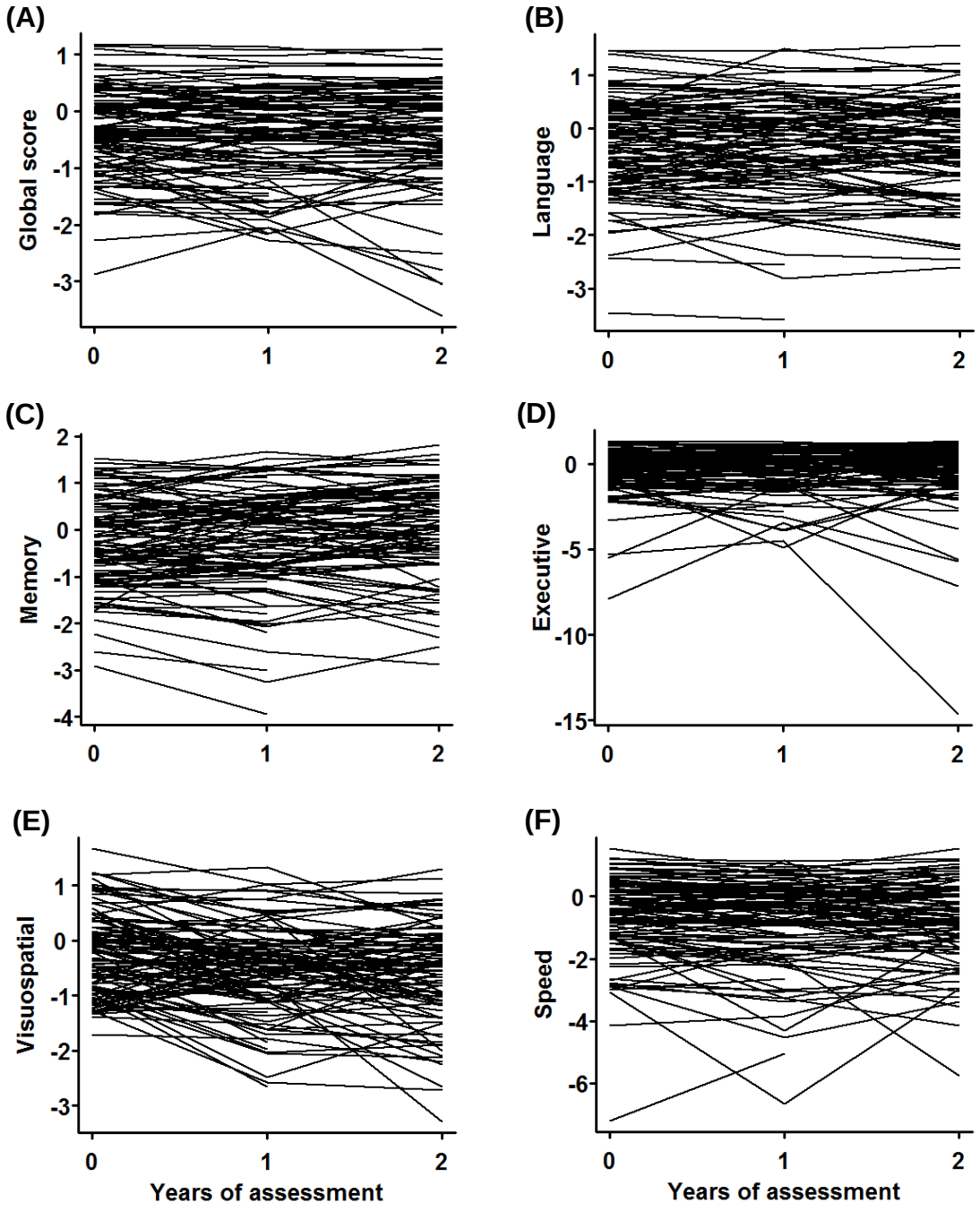

Figure 1. Subject-specific evolutions in time of the Neuropsychological Functioning, showing global score (A), the language domain (B), the memory domain (C), the executive domain (D), the visuospatial domain (E), and speed of information processing domain (F).

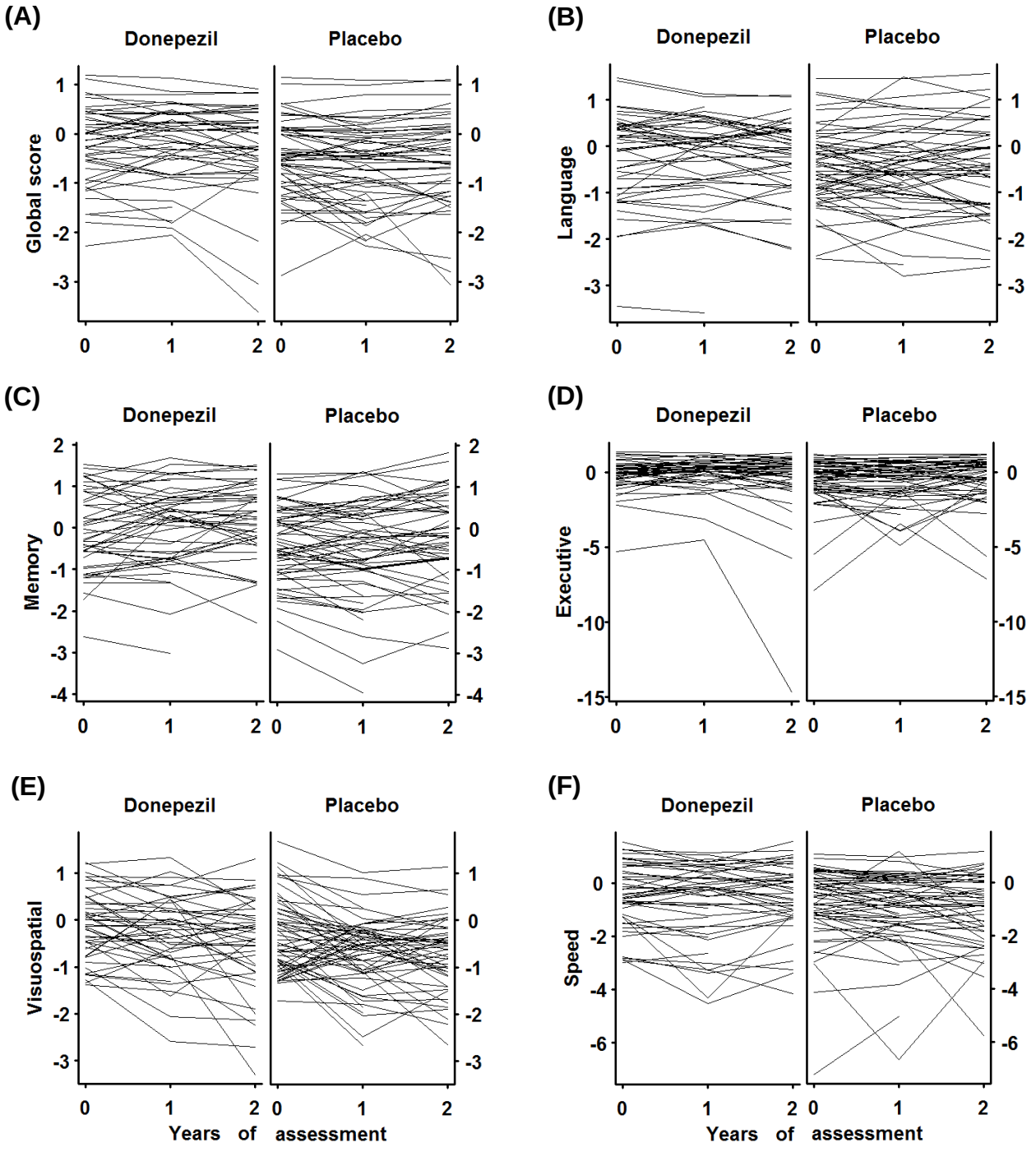

Figure 2. Subject-specific evolutions in time of the Neuropsychological Functioning, separately for Donepezil and Placebo groups, showing global score (A), the language domain (B), the memory domain (C), the executive domain (D), the visuospatial domain (E), and speed of information processing domain (F).

The composite measure of C-IADL was dichotomized. The included covariates are age, years of education, baseline MCI, and follow-up time in years, which were chosen because they were significant variables from the preliminary analyses fitting the separate outcomes. Although no significant effect was observed, treatment is included because the treatment effect was the primary interest in the study. The time to recurrence of major depression was used as time-to-event outcome. The Kaplan-Meier plot of recurrence of major depression by treatment is shown in Figure 3. From the Kaplan-Meier estimates in Figure 3 and the associated log-rank test, the Donepezil treatment group has a marginally significant higher recurrence of major depression than the Placebo group.

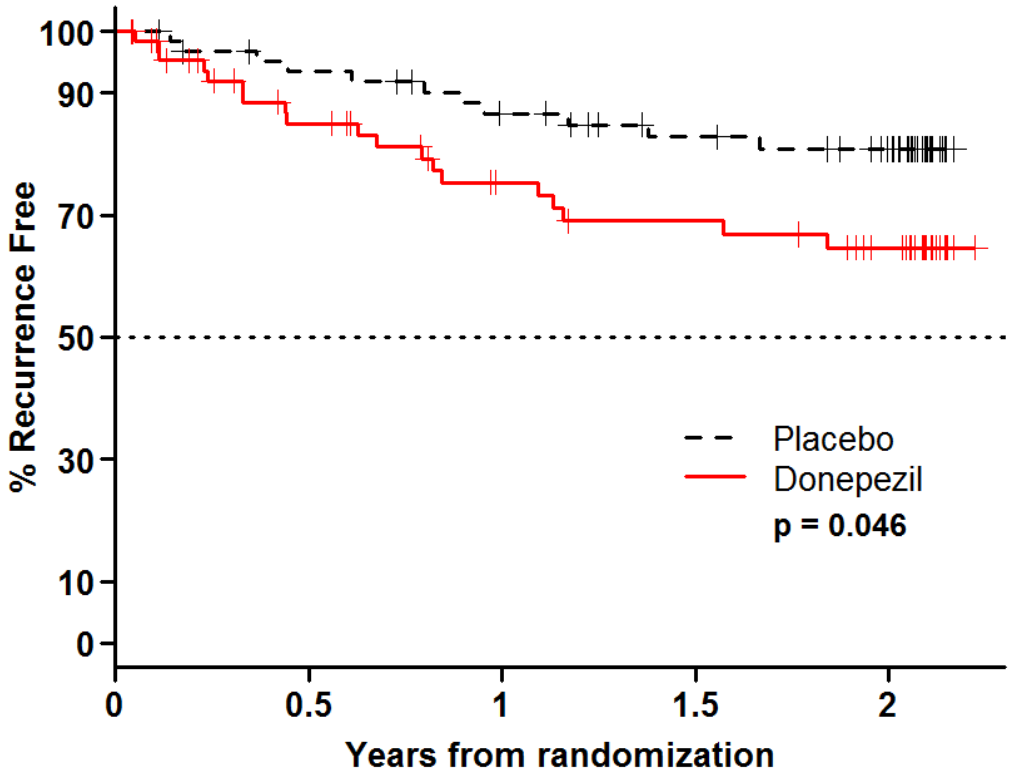

Figure 3. The Kaplan-Meier plot of recurrence of major depression for Donepezil (*solid red curve*) and Placebo (*broken black curve*) groups.

4.3 MODEL DEFINITIONS

4.3.1 Shared Random Effects Joint Model

The subject-specific random intercepts and slopes, b_{10i} and b_{11i}, of the continuous longitudinal outcome, are shared in the joint model. We assume $\boldsymbol{b}_{1i}$ follow normal distribution with mean vector zero and variance-covariance matrix $\boldsymbol{\Sigma}$, and that $\boldsymbol{b}_{2i}$ is proportional to $\boldsymbol{b}_{1i}$, i.e., $\boldsymbol{b}_{2i} = \boldsymbol{A}_0\boldsymbol{b}_{1i}$, where $\boldsymbol{A}_0$ is a diagonal matrix of unknown constants as introduced in section 3.1. The model entities are defined as follows:

y_{1ij} - global scores of neuropsychological functioning for the *i*th subject at time *j*(continuous).

y_{2ij} - composite measure of C-IADL for the *i*th subject at time *j* (binary).

T_i - time to recurrence of major depression for the *i*th subject.

$$y_{1ij}|\boldsymbol{b}_{10}, \boldsymbol{b}_{11} \sim indep.\ N\left(\mu_{1i}(t_{ij}),\ \sigma^2_{\epsilon 1}\right)$$

$$y_{1ij} = \mu_{1i}(t_{ij}) + U_{1i}(t_{ij}) + \epsilon_{1ij}$$

- $\mu_{1i}(t_{ij}) = \mathbf{x}^T_{1i}(t_{ij})\boldsymbol{\beta}_1 = \beta_{11} + \beta_{12}\text{Age}_i + \beta_{13}\text{Education}_i + \beta_{14}\text{Baseline}_{\text{MCI}_i} + \beta_{15}\text{Time}_{ij} + \beta_{16}\text{Treatment}_i$
- $U_{1i}(t_{ij}) = \mathbf{z}^T_{1i}(t_{ij})\mathbf{b}_{1i} = b_{10i} + b_{11i}\text{Time}_{ij}$
- $\mathbf{b}_{1i} = \begin{pmatrix} b_{10i} \\ b_{11i} \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2_{b_{10}} & \rho_{b_1}\sigma_{b_{10}}\sigma_{b_{11}} \\ \rho_{b_1}\sigma_{b_{10}}\sigma_{b_{11}} & \sigma^2_{b_{11}} \end{bmatrix} \right)$
- $\epsilon_{1ij} \sim N\left(0,\ \sigma^2_{\epsilon_1}\right)$

$y_{2ij}|\boldsymbol{b}_{20}, \boldsymbol{b}_{21} \sim \ indep.\ Bernoulli(\Pr(y_{2ij} = 1))$

$$\text{logit}(\Pr(y_{2ij} = 1)) = \mu_{2i}(t_{ij}) + U_{2i}(t_{ij})$$

- $\mu_{2i}(t_{ij}) = \mathbf{x}_{2i}^{T}(t_{ij})\boldsymbol{\beta}_2 = \beta_{21} + \beta_{22}\text{Age}_i + \beta_{23}\text{Baseline}_{\text{MCI}_i} + \beta_{24}\text{Time}_{ij} + \beta_{25}\text{Treatment}_i$
- $U_{2i}(t_{ij}) = \mathbf{z}_{2i}^{T}(t_{ij})\mathbf{b}_{2i} = \mathbf{z}_{2i}^{T}(t_{ij})\boldsymbol{A}_0\mathbf{b}_{1i} = \alpha_1 b_{10i} + \alpha_2 b_{11i}\text{Time}_{ij}$

$T_i \sim \text{Weibull}(p, \mu_{3i})\ ,\ p > 0$

$$\log(\mu_{3i}) = \mathbf{x}_{3i}^{T}\boldsymbol{\beta}_3 + U_{3i}$$

- $\mathbf{x}_{3i}^{T}\boldsymbol{\beta}_3 = \beta_{31} + \beta_{32}\text{Treatment}_i$
- $U_{3i} = \boldsymbol{a}_1^{\intercal}\boldsymbol{b}_{1i} + \boldsymbol{a}_2^{\intercal}\boldsymbol{b}_{2i} + b_{3i} = \alpha_3 b_{10i} + \alpha_4 b_{11i} + b_{3i}$
- $b_{3i} \sim N(\ 0,\ \sigma_{b_3}^2\ )$
- Hazard at time t : $\lambda_i(t) = pt^{p-1}\mu_{3i} = pt^{p-1}\exp\{\mathbf{x}_{3i}^{T}\boldsymbol{\beta}_3 + U_{3i}\}$

For the continuous longitudinal outcome, we included the covariates of age, years of education, baseline MCI, time in years, and treatment. For the binary longitudinal outcome, the same covariates are included except years of education. Only treatment is included in the model for time-to-event outcome as other covariates did not reach statistical significance. The longitudinal continuous and binary outcomes and survival outcome are associated through shared random effects of longitudinal continuous outcome. Thus, the association between both longitudinal outcomes and time-to-event outcome can be explained by α_3 and α_4.

4.3.2 Correlated Random Effects Joint Model

We assume that the subject-specific random effects of continuous and binary longitudinal outcomes, $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$, are multinormally distributed as

$$\begin{pmatrix}\boldsymbol{b}_{1i}\\ \boldsymbol{b}_{2i}\end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{pmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{12}^{\intercal} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right).$$

The model assumptions are same with section 4.3.1, while the random effects terms now take the form

- $U_{1i}(t_{ij}) = b_{10i} + b_{11i}\text{Time}_{ij}$
- $U_{2i}(t_{ij}) = b_{20i} + b_{21i}\text{Time}_{ij}$
- $U_{3i} = \boldsymbol{a}_1^{\intercal}\boldsymbol{b}_{1i} + \boldsymbol{a}_2^{\intercal}\boldsymbol{b}_{2i} + b_{3i} = \alpha_1 b_{10i} + \alpha_2 b_{11i} + \alpha_3 b_{20i} + \alpha_4 b_{21i} + b_{3i}$

In this case, the association between both longitudinal outcomes and time-to-event outcome can be explained by $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

4.3.3 Assumptions About The Prior Distributions

We selected non-informative priors for all parameters, that is, priors that had very large variance components. For the coefficients of the fixed effects $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$, we use multivariate normal distributions with mean zero and variance-covariance matrices $100\boldsymbol{I}_6$, $100\boldsymbol{I}_5$, and $100\boldsymbol{I}_2$, respectively, where $\boldsymbol{I}_\text{k}$ indicates an $k \times k$ identity matrix. For error variance $\sigma^2_{\epsilon_1}$, we take an inverse gamma(0.1, 0.1), for frailty term variance σ^2_3, we take an inverse gamma(20, 5), for the association coefficients $\boldsymbol{\alpha}$, we use a normal distribution $N(0, 100)$. For the parameters of random effects, we use an inverse Wishart which is a conjugate prior for the variance-covariance matrix

in the multivariate normal likelihood (Carlin and Louis 2009). The methods are programed in the R interface called 'rbugs' to access the software OpenBUGS. As was previously mentioned, in our analyses, a total of 20,000 MCMC iterations were used. We discarded the first 10,000 iterations for the burn-in period.

4.4 RESULTS

4.4.1 Shared Random Effects Joint Model

We applied various models with all possible combinations of forms of subject-specific random effects for the three processes as the longitudinal continuous and binary outcomes, and time-to-event outcome. We considered 37 differently expressed shared random effects joint models. First, we assumed Weibull distribution for the time to recurrence of major depression outcome. But the estimated values of shape parameter in Weibull distribution were not significantly different with one in all 37 models. Thus, we fit the models assuming exponential distribution for the time to recurrence of major depression.

Table 2 summarizes DICs for each submodel, the posterior expected deviance, $\overline{D}$, the effective number of parameters, p_D, and total DIC scores, DIC_{total}, for each joint model. Note that DIC_{y1}, DIC_{y2}, and DIC_T denote DIC from the longitudinal continuous outcome submodel, from the longitudinal binary outcome submodel, and from the time-to-event outcome submodel, respectively. These scores show relative contributions to the overall model DIC. The total DIC score, DIC_{total}, was used to choose the best model. The joint models with subject-specific random intercepts in both the longitudinal continuous outcome submodel and binary outcome submodel

(Model 7 ~ Model 9) show the decreased DICs in both the longitudinal outcome submodels, and hence the DIC_{total} of the overall model. When subject-specific random slopes were included in the longitudinal continuous submodel (Model 11 further), DIC_{y1} for the longitudinal continuous submodel increased. Although there were reductions in DIC_{y2} for the longitudinal binary submodel when subject-specific random slopes were allowed in the longitudinal binary submodel (Model 29 ~ Model 37), it could not cover the increasing amount of DIC_{y1} to improve in DIC_{total}. The model with the smallest DIC_{total} is Model 8. Unfortunately, this model did not assume any association between the longitudinal outcomes and the time-to-event outcome. We investigated the associations between the longitudinal outcomes and the time-to-event outcome comparing the Model 37 assumed that the longitudinal binary outcome shares the random intercept and slope of the longitudinal continuous outcome and the time-to-event outcome is related to both the subject-specific random intercept and slope of the longitudinal continuous outcome and has frailty term.

Table 3 reported the posterior estimates of fixed effects, random effects, and the association of the longitudinal outcomes and the time-to-event outcome, the standard errors, and their 95% credibility intervals for both Model 8 and Model 37. The results from the Model 8 and Model 37 are similar in the longitudinal continuous submodel. It seems that even though the subject-specific random slope of the longitudinal continuous outcome caused the longitudinal continuous submodel to increase in DIC_{y1}, it does not effect on the posterior estimates of longitudinal continuous outcome. In the longitudinal binary submodel, the estimates of Model 37 are slightly smaller than Model 8. The association parameters between the longitudinal outcomes are significant in both the subject-specific random intercept and slope showing positive relationship between the both longitudinal outcomes.

Table 2. Bayesian Model Selection of shared random effects joint model – MTLD data.

Model	$U_1(t)$	$U_2(t)$	U_3	DIC_{y1}	DIC_{y2}	DIC_T	$\bar{D}$	p_D	DIC_{total}
	no random effects								
1	0	0	0	570.1	339.7	169.1	1065	14.0	1079
2	0	0	b_3	569.9	339.7	166.7	1056	20.3	1076
	random intercepts								
3	b_{10}	0	0	224.1	339.7	169.1	609.6	123.3	732.9
4	b_{10}	0	b_3	225.8	339.9	166.6	602.0	130.3	732.3
5	b_{10}	0	$\alpha_3 b_{10}$	224.9	339.7	170.4	610.7	124.3	735.0
6	b_{10}	0	$\alpha_3 b_{10} + b_3$	225.1	339.8	167.7	602.0	130.5	732.6
7	b_{10}	$\alpha_1 b_{10}$	0	225.4	310.7	169.1	579.9	125.3	705.2
8	b_{10}	$\alpha_1 b_{10}$	b_3	224.3	310.8	166.8	570.1	131.7	701.8
9	b_{10}	$\alpha_1 b_{10}$	$\alpha_3 b_{10}$	224.8	310.7	170.3	580.2	125.7	705.9
10	b_{10}	$\alpha_1 b_{10}$	$\alpha_3 b_{10} + b_3$	225.0	310.6	167.9	571.3	132.2	703.5
	random intercepts and random slopes								
11	$b_{10} + b_{11} t$	0	0	256.7	339.9	169.3	604.2	161.6	765.8
12	$b_{10} + b_{11} t$	0	b_3	258.1	339.8	166.7	596.5	168.1	764.6
13	$b_{10} + b_{11} t$	0	$\alpha_3 b_{10}$	259.9	339.8	169.6	607.3	161.9	769.2
14	$b_{10} + b_{11} t$	0	$\alpha_4 b_{11}$	256.5	339.6	165.5	600.2	161.4	761.6
15	$b_{10} + b_{11} t$	0	$\alpha_3 b_{10} + b_3$	260.3	339.6	167.0	599.5	167.5	767.0
16	$b_{10} + b_{11} t$	0	$\alpha_4 b_{11} + b_3$	259.1	339.6	162.0	594.1	166.6	760.7
17	$b_{10} + b_{11} t$	0	$\alpha_3 b_{10} + \alpha_4 b_{11}$	262.0	339.8	159.8	597.5	164.2	761.7
18	$b_{10} + b_{11} t$	0	$\alpha_3 (b_{10} + b_{11})$	257.1	339.7	170.2	605.1	161.9	767.0
19	$b_{10} + b_{11} t$	0	$\alpha_3 b_{10} + \alpha_4 b_{11} + b_3$	263.0	339.8	155.3	591.2	166.9	758.1
20	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	0	264.2	306.2	169.2	574.6	165.1	739.7
21	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	b_3	263.0	306.5	166.6	564.4	171.8	736.2
22	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 b_{10}$	263.0	307.4	170.1	574.2	166.4	740.6
23	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_4 b_{11}$	263.4	307.0	166.1	572.6	163.9	736.6
24	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 b_{10} + b_3$	264.7	306.5	167.1	566.3	172.0	738.3
25	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_4 b_{11} + b_3$	263.9	306.8	163.7	563.9	170.4	734.4
26	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 b_{10} + \alpha_4 b_{11}$	267.1	308.0	159.7	569.6	165.2	734.8
27	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 (b_{10} + b_{11})$	264.8	306.8	170.3	576.0	165.8	741.9
28	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 b_{10} + \alpha_4 b_{11} + b_3$	267.4	307.7	157.9	563.1	169.9	733.0
29	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	0	259.3	298.7	169.2	563.1	164.1	727.2
30	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	b_3	260.6	298.1	166.6	554.8	170.6	725.4
31	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 b_{10}$	258.4	299.2	169.7	561.6	165.7	727.3
32	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_4 b_{11}$	261.0	298.1	164.4	558.3	165.3	723.6
33	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 b_{10} + b_3$	260.5	297.2	167.1	554.7	170.1	724.8
34	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_4 b_{11} + b_3$	262.7	299.2	162.4	553.5	170.8	724.3
35	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 b_{10} + \alpha_4 b_{11}$	263.1	299.8	160.2	551.7	171.4	723.1
36	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 (b_{10} + b_{11})$	261.1	297.6	170.5	565.1	164.1	729.2
37	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 b_{10} + \alpha_4 b_{11} + b_3$	264.8	298.6	156.0	545.6	173.8	719.4

Table 3. Joint Bayesian analysis results for shared random effects joint model – MTLD data.

Parameter	Model 8			Model 37		
	Posterior Mean	Std. Error	95% CI	Posterior Mean	Std. Error	95% CI
Longitudinal continuous submodel						
Intercept (β_{11})	1.994	0.638	(0.741, 3.259)	1.722	0.487	(0.789, 2.671)
Age (β_{12})	-0.045	0.008	(-0.061, -0.029)	-0.041	0.006	(-0.054, -0.030)
Education (β_{13})	0.088	0.017	(0.054, 0.122)	0.088	0.014	(0.060, 0.116)
Baseline MCI (β_{14})	-0.839	0.103	(-1.040, -0.638)	-0.842	0.081	(-0.993, -0.679)
Time (β_{15})	-0.062	0.021	(-0.103, -0.022)	-0.064	0.033	(-0.129, 0.002)
Treatment (β_{16})	0.043	0.098	(-0.147, 0.232)	0.043	0.077	(-0.104, 0.202)
$\sigma^2_{b_{10}}$	0.266	0.040	(0.197, 0.352)	0.115	0.014	(0.090, 0.145)
$\sigma^2_{b_{11}}$	-	-	-	0.064	0.007	(0.052, 0.078)
ρ_{b_1}	-	-	-	0.094	0.076	(-0.054, 0.240)
σ^2_{ϵ}	0.083	0.009	(0.068, 0.103)	0.088	0.012	(0.067, 0.114)
Longitudinal binary submodel						
Intercept (β_{21})	9.359	2.198	(5.165, 13.920)	9.255	2.205	(5.065, 13.680)
Age (β_{22})	-0.112	0.029	(-0.173, -0.057)	-0.110	0.029	(-0.168, -0.053)
Baseline MCI (β_{23})	-0.791	0.353	(-1.507, -0.115)	-0.935	0.365	(-1.662, -0.235)
Time (β_{24})	-0.559	0.183	(-0.919, -0.210)	-0.709	0.255	(-1.248, -0.238)
Treatment (β_{25})	-0.381	0.336	(-1.065, 0.275)	-0.420	0.346	(-1.113, 0.254)
α_1	1.815	0.382	(1.097, 2.596)	1.707	0.676	(0.483, 3.138)
α_2	-	-	-	4.869	2.214	(1.305, 9.783)
Time-to-event submodel						
Intercept (β_{31})	-2.354	0.323	(-3.035, -1.768)	-2.659	0.438	(-3.655, -1.908)
Treatment (β_{32})	0.794	0.400	(0.028, 1.586)	0.844	0.437	(0.002, 1.735)
α_3	-	-	-	-1.388	0.835	(-3.108, 0.226)
α_4	-	-	-	3.856	2.051	(0.209, 8.139)
$\sigma^2_{b_3}$	0.272	0.067	(0.172, 0.428)	0.268	0.064	(0.170, 0.417)

The treatment covariate is significant only in the time-to-event submodel. We observe that the hazard of the recurrence of major depression is significantly higher for Donepezil treatment group as we showed in Figure 3. The association between the longitudinal outcomes and the time-to-event outcome can be explained by parameter α_3 and α_4. We observe the negative association between the subject-specific random intercept of the global performance score of

neuropsychological functioning and the time to recurrence of major depression although α_3 is not significant in level 0.05. And the subject-specific random slope of the global score is positively associated with the risk of for a recurrence of major depression significantly.

4.4.2 Correlated Random Effects Joint Model

We considered all possible combinations of differently expressed correlated random effects joint models. Because we showed there is no gain in the DICs by including the subject-specific random slope in the longitudinal continuous outcome in the section 4.4.1, we reported here only the models allowed the subject-specific random intercept in the longitudinal continuous submodel in Table 4. We observe that there is great decrease in the DIC_T with the association between the subject-specific random intercepts and slopes of the longitudinal binary outcome and the time-to-event outcome in the time-to-event submodel. But inconclusive results emerge for the correlated random effects joint modeling for this data set. The sub-components of p_D estimates for the time-to-event submodel were negative in some models, particularly in the model included the association between the subject-specific random intercept of the longitudinal binary outcome and the time-to-event outcome in the time-to-event submodel. Also the sub-components of p_D estimates for the longitudinal binary submodel were strangely small in all models. The small numbers of repeated measurements (one, two, or three) of binary outcome appears to be causing problems with the estimation. In such cases, we cannot trust the improvement of total DIC obtained for these models (Carlin and Louis 2009). Thus, we cannot choose the best model for the correlated random effects joint models nor report the results of the estimates.

Table 4. Bayesian Model Selection of Correlated random effects joint model – MTLD data.

Mode	$U_1(t)$	$U_2(t)$	U_3	DIC_{y1}	DIC_{y2}	DIC_T	$\bar{D}$	p_D	DIC_{total}
	random intercepts								
1	b_{10}	b_{20}	0	232.1	330.6	169.0	610.1	121.6	731.7
2	b_{10}	b_{20}	b_3	232.9	330.7	166.6	601.8	128.4	730.2
3	b_{10}	b_{20}	$\alpha_1 b_{10}$	232.8	330.7	170.3	611.5	122.3	733.8
4	b_{10}	b_{20}	$\alpha_1 b_{10} + b_3$	232.3	330.4	167.7	601.7	128.7	730.3
5	b_{10}	b_{20}	$\alpha_3 b_{20}$	233.5	334.1	122.3	585.4	104.5	689.9
6	b_{10}	b_{20}	$\alpha_3 b_{20} + b_3$	236.1	332.4	141.6	587.1	123.0	710.0
7	b_{10}	b_{20}	$\alpha_1 b_{10} + \alpha_3 b_{20}$	233.9	331.1	130.2	572.3	122.8	695.2
8	b_{10}	b_{20}	$\alpha_1 b_{10} + \alpha_3 b_{20} + b_3$	232.1	330.9	123.4	574.3	112.1	686.4
9	b_{10}	$b_{20} + b_{21} t$	0	232.2	323.6	169.0	596.4	128.5	724.9
10	b_{10}	$b_{20} + b_{21} t$	b_3	231.7	323.7	166.6	586.8	135.3	722.1
11	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10}$	231.4	323.4	170.2	595.7	129.2	724.9
12	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + b_3$	231.8	324.1	167.9	588.4	135.4	723.8
13	b_{10}	$b_{20} + b_{21} t$	$\alpha_3 b_{20}$	231.8	326.1	120.9	572.5	106.4	678.9
14	b_{10}	$b_{20} + b_{21} t$	$\alpha_3 b_{20} + b_3$	232.9	325.8	129.5	568.8	119.4	688.2
15	b_{10}	$b_{20} + b_{21} t$	$\alpha_4 b_{21}$	232.8	326.0	131.0	573.4	116.3	689.7
16	b_{10}	$b_{20} + b_{21} t$	$\alpha_4 b_{21} + b_3$	232.0	325.3	140.2	566.1	131.3	697.4
17	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_3 b_{20}$	230.9	322.7	145.6	553.9	145.3	699.2
18	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_3 b_{20} + b_3$	233.5	324.3	139.7	560.2	137.2	697.5
19	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_4 b_{21}$	231.0	322.0	146.4	547.8	151.6	699.4
20	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_4 b_{21} + b_3$	230.0	323.9	133.6	552.6	134.9	687.5
21	b_{10}	$b_{20} + b_{21} t$	$\alpha_3 b_{20} + \alpha_4 b_{21}$	232.6	325.9	88.4	552.1	94.8	646.9
22	b_{10}	$b_{20} + b_{21} t$	$\alpha_3 b_{20} + \alpha_4 b_{21} + b_3$	232.6	325.1	102.4	554.1	105.9	660.0
23	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_3 b_{20} + \alpha_4 b_{21}$	233.2	323.5	104.6	552.4	108.9	661.3
24	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_3 b_{20} + \alpha_4 b_{21} + b_3$	232.4	323.9	100.7	544.5	112.4	657.0

5.0 APPLICATION TO MORTALITY IN IDIOPATHIC PULMONARY FIBROSIS OUTCOMES STUDY

5.1 DESCRIPTION OF DATASET

In this chapter, we present the analysis of mortality in the idiopathic pulmonary fibrosis (IPF) outcomes study. IPF is a chronic, progressive lung disease characterized by fibrosis of unknown etiology (Richards et al, 2012). Our primary concern is to model the relationship between longitudinally measured pulmonary function tests (PFT) and a time-to-event outcome which included two survival time random variables, overall survival and transplant-free survival. For overall survival analysis, only death without lung transplantation is treated as an event. For the transplant-free survival outcome, lung transplants were counted as events in addition to death. Among PFTs, forced vital capacity (FVC) measurement was used as a longitudinal continuous outcome. A disease progression indicator variable defined as a decline of 5% or more in FVC from the baseline FVC measurement was used as a longitudinal binary outcome. From the 125 patients considered in the study, 64 (51.2%) patients died or were transplanted. Patient characteristics are summarized in Table 5. Of the 125 patients, 67.2% of patients were male. Almost (97.6%) patients were white, 29.6% of patients had never smoked, 50.4% of patients had a clinically confirmed diagnosis, and 18.4% were transplanted. The average age at diagnosis was

65.2 ± 9.34 years. This data set was comprised of irregular follow-up times across patients. Thus, the number and timing of longitudinal measurements are different for each patient. The PFTs of each patient were repeatedly measured about 11 times on average, up to 43 times. The average follow-up time is 3.7 years among all patients and 4.1 years among patients who do not die nor were transplanted. In the data analysis, age was mean-centered.

Table 5. Patient characteristics of IPF outcome data.

		N(=125)	**%**
Gender			
	Male	84	67.2
	Female	41	32.8
Race			
	White	122	97.6
	Black	1	0.8
	American Indian	1	0.8
	Oriental	1	0.8
Smoking			
	Ever	88	70.4
	Never	37	29.6
Diagnosis made			
	Clinically	63	50.4
	Historically	62	49.6
Transplant			
	Yes	23	18.4
	No	102	81.6
		Mean	**SD**
Age (years)			
	Overall	65.2	9.34
	Male	65.7	9.00
	Female	64.2	10.04
Follow-up (years)			
	All patients	3.7	2.32
	Alive and not transplanted	4.1	2.35
Baseline PFTs			
	FVC	2.7	0.85
	FEV1	2.2	0.65
	DLCO	12.7	4.36

Definition of abbreviations: PFT = pulmonary function tests; FVC = forced vital capacity; FEV1 = forced expiratory volume in 1 second; DLCO = diffusing capacity of carbon monoxide.

Individual profiles for the pulmonary function assessed by FVC are shown in Figure 4. With higher baseline FVC, the FVC measurement trajectory seems to decline in time but with moderate baseline FVC, it tends to maintain the degree or to decrease only slightly. In contrast, FVC shows to increase in time with lower baseline FVC for some patients. We can see subject-specific random intercepts and slopes. Covariates included in the model are centered age at diagnosis, baseline FVC, smoking, gender, and time in years. The stepAIC approach for variable selection was applied in the preliminary analyses fitting separate outcomes.

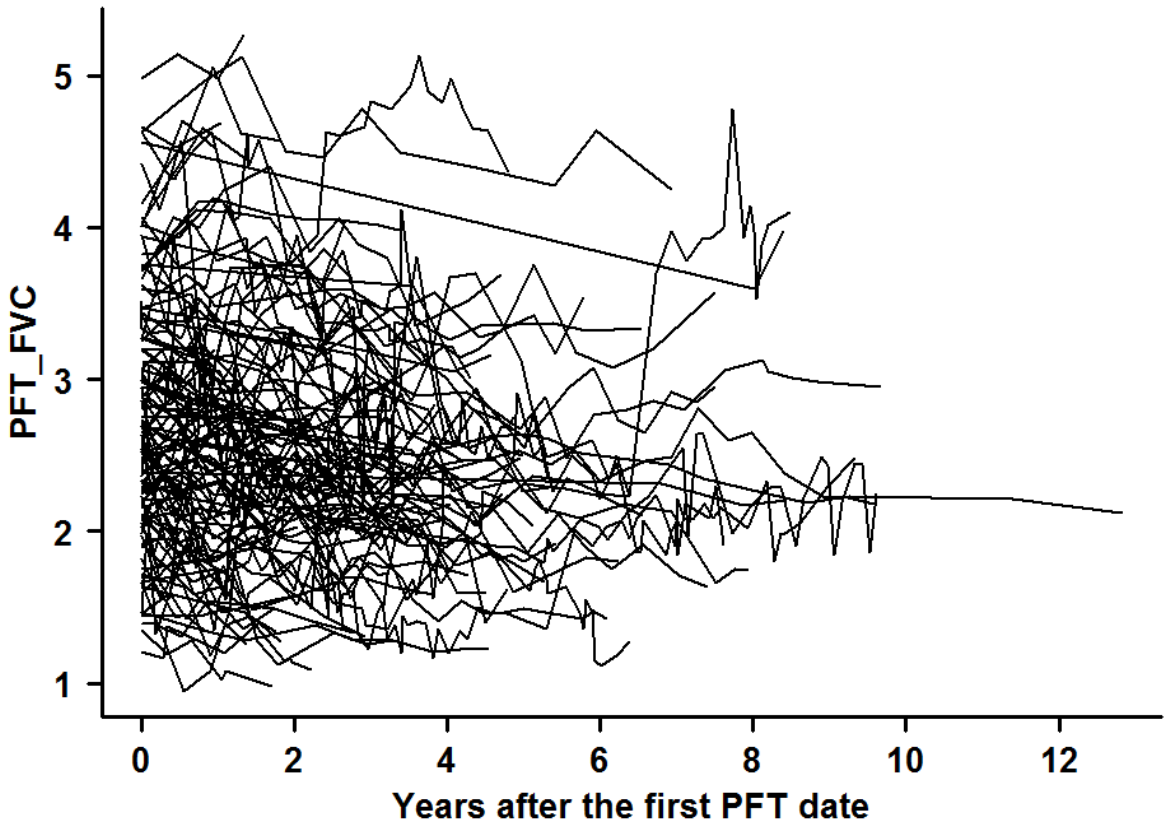

Figure 4. Subject-specific evolutions in time of FVC measurements.

Survival curves were estimated using the Kaplan-Meier method and fitted survival curves of the Weibull model and exponential model are shown in Figure 5. The survival curves of the Weibull model and exponential model are almost indistinguishable and both models show a good fit to the marginal survival function. The estimated median mortality from the initial visit date was 7.9 years (Figure 5, *red line*); the median transplant-free survival time was 4.4 years (Figure 5, *blue line*). All IPF patients were evaluated at the University of Pittsburgh Medical Center and clinical data were obtained from the Simmons Center for Interstitial Lung Disease.

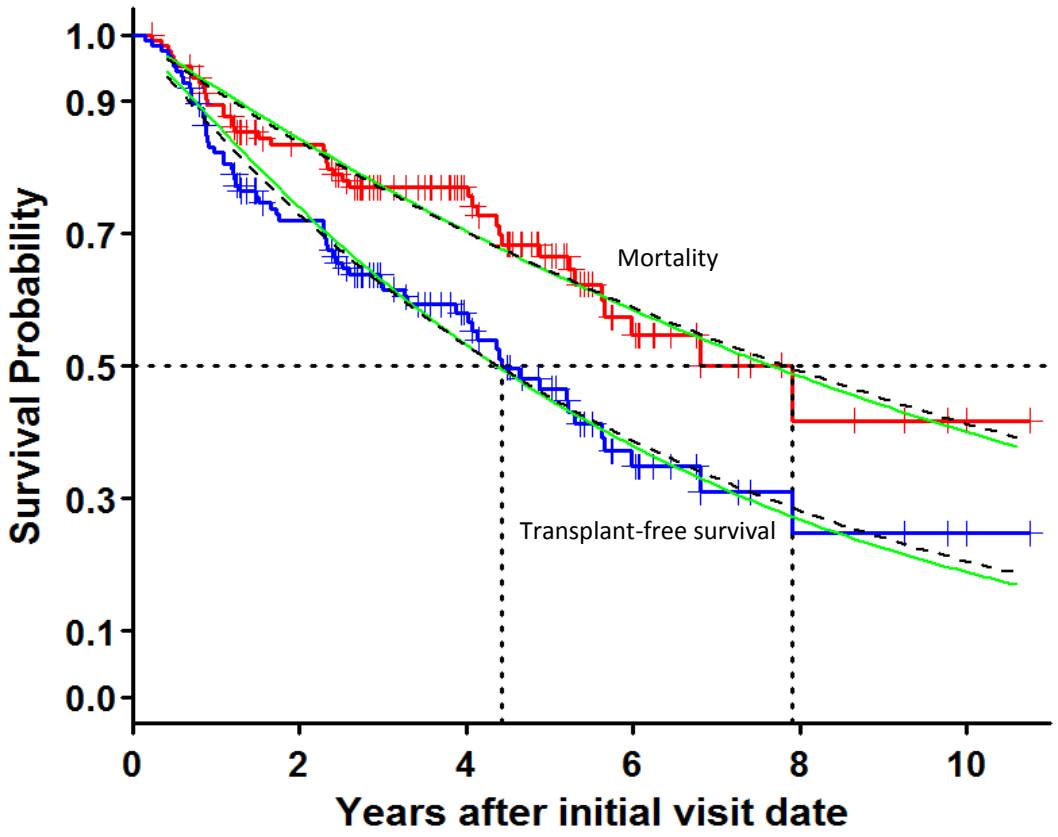

Figure 5. The Kaplan-Meier plot of overall survival (*red curve*) and transplant-free survival (*blue curve*); Weibull survival curve (*green curve*) and exponential survival curve (*black broken curve*).

5.2 MODEL DEFINITIONS

5.2.1 Shared Random Effects Joint Model

The subject-specific random intercepts and slopes, b_{10i} and b_{11i}, of the continuous longitudinal outcome are shared in the joint model. We assume $\boldsymbol{b}_{1i}$ follow normal distribution with mean vector zero and variance-covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{b}_{2i}$ is proportional to $\boldsymbol{b}_{1i}$, i.e., $\boldsymbol{b}_{2i} = \boldsymbol{A}_0\boldsymbol{b}_{1i}$, where $\boldsymbol{A}_0$ is a diagonal matrix of unknown constants same with section 4.2.1. The model is defined as follows:

y_{1ij} – FVC measurement of PFT for the *i*th subject at time *j* (continuous).

y_{2ij} – Disease progression indicator of decline of at least 5 % in FVC from the baseline FVC measurement for the *i*th subject at time *j* (binary).

T_i – time to transplant or death for the *i*th subject.

$$y_{1ij}|\boldsymbol{b}_{10}, \boldsymbol{b}_{11} \sim \ indep.\ N\big(\ \mu_{1i}(t_{ij}),\ \ \sigma^2_{\epsilon 1}\ \big)$$

$$y_{1ij} = \mu_{1i}(t_{ij}) + U_{1i}(t_{ij}) + \epsilon_{1ij}$$

- $\mu_{1i}(t_{ij}) = \mathbf{x}^T_{1i}(t_{ij})\boldsymbol{\beta}_1 \ = \beta_{11} + \beta_{12}\text{Baseline}_{\text{FVC}_i} + \beta_{13}\text{Time}_{ij}$
- $U_{1i}(t_{ij}) = \mathbf{z}^T_{1i}(t_{ij})\mathbf{b}_{1i} = b_{10i} + b_{11i}\text{Time}_{ij}$
- $\mathbf{b}_{1i} = \begin{pmatrix} b_{10i} \\ b_{11i} \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2_{b_{10}} & \rho_{b_1}\sigma_{b_{10}}\sigma_{b_{11}} \\ \rho_{b_1}\sigma_{b_{10}}\sigma_{b_{11}} & \sigma^2_{b_{11}} \end{bmatrix} \right)$
- $\epsilon_{1ij} \sim N\big(\ 0,\ \sigma^2_{\epsilon_1}\ \big)$

$y_{2ij}|\boldsymbol{b}_{20}, \boldsymbol{b}_{21} \sim \ indep.\ Bernoulli(\Pr(y_{2ij} = 1))$

$$\text{logit}(\Pr(y_{2ij} = 1)) = \mu_{2i}(t_{ij}) + U_{2i}(t_{ij})$$

- $\mu_{2i}(t_{ij}) = \mathbf{x}_{2i}^T(t_{ij})\boldsymbol{\beta}_2 = \beta_{21} + \beta_{22}\text{Age}_i + \beta_{23}\text{Time}_{ij}$
- $U_{2i}(t_{ij}) = \mathbf{z}_{2i}^T(t_{ij})\mathbf{b}_{2i} = \mathbf{z}_{2i}^T(t_{ij})\boldsymbol{A}_0\mathbf{b}_{1i} = \alpha_1 b_{10i} + \alpha_2 b_{11i}\text{Time}_{ij}$

$T_i \sim \text{Exponential}(\mu_{3i})$

$$\log(\mu_{3i}) = \mathbf{x}_{3i}^T\boldsymbol{\beta}_3 + U_{3i}$$

- $\mathbf{x}_{3i}^T\boldsymbol{\beta}_3 = \beta_{31} + \beta_{32}\text{Gender}_i + \beta_{33}\text{Smoking}_i + \beta_{34}\text{Baseline}_{\text{FVC}_i}$
- $U_{3i} = \boldsymbol{a}_1^\intercal\boldsymbol{b}_{1i} + \boldsymbol{a}_2^\intercal\boldsymbol{b}_{2i} + b_{3i} = \alpha_3 b_{10i} + \alpha_4 b_{11i} + b_{3i}$
- $b_{3i} \sim N(\ 0,\ \sigma_{b_3}^2\ )$
- Hazard function : $\lambda_i = \mu_{3i} = \exp\{\mathbf{x}_{3i}^T\boldsymbol{\beta}_3 + U_{3i}\}$

For the continuous longitudinal outcome, we included the covariates of baseline FVC and time in years. For the binary longitudinal outcome, age and time in years were found to be significant covariates. Gender, smoking, and baseline FVC were significant in the model for transplant-free survival and exponential model was assumed. Because the Weibull regression model and the exponential regression model are almost identical for this data set (Figure 5), we assumed the exponential model for the transplant-free survival. The longitudinal continuous and binary outcomes and survival outcome are associated through the shared random effects of the continuous longitudinal outcome.

5.2.2 Correlated Random Effects Joint Model

The subject-specific random effects of continuous and binary longitudinal outcomes, $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$, are assumed to follow multivariate normal distribution as

$$\begin{pmatrix}\boldsymbol{b}_{1i}\\ \boldsymbol{b}_{2i}\end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{12}^{\intercal} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right).$$

The model assumptions are same with section 5.2.1, while the random effects terms now take the form

- $U_{1i}(t_{ij}) = b_{10i} + b_{11i}\text{Time}_{ij}$
- $U_{2i}(t_{ij}) = b_{20i} + b_{21i}\text{Time}_{ij}$
- $U_{3i} = \boldsymbol{a}_1^{\intercal}\boldsymbol{b}_{1i} + \boldsymbol{a}_2^{\intercal}\boldsymbol{b}_{2i} + b_{3i} = \alpha_1 b_{10i} + \alpha_2 b_{11i} + \alpha_3 b_{20i} + \alpha_4 b_{21i} + b_{3i}$

The association between both longitudinal outcomes and time-to-event outcome can be explained by the parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

5.2.3 Assumptions About The Prior Distributions

The prior assumptions for IPF data set are similar with MTLD data set in section 4.3.3. For the coefficients of the fixed effects $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$, we assume a multivariate normal distribution with mean zero and variance-covariance matrices $100\boldsymbol{I}_3$, $100\boldsymbol{I}_3$, and $100\boldsymbol{I}_4$, respectively, where $\boldsymbol{I}_k$ indicates an identity matrix with dimension *k*. For error variance, $\sigma_{\epsilon_1}^2$, and frailty term variance, σ_3^2, we take an inverse gamma distribution. For the association coefficients $\boldsymbol{\alpha}$, we take a normal distribution with mean zero and variance equal to 10^2. For the parameters of random effects, we take invers Wishart distribution. A total of 15,000 MCMC iterations were used discarding the first 5,000 iterations for the burn-in period.

5.3 RESULTS

5.3.1 Shared Random Effects Joint Model

As with the application of MTLD data set in the chapter 4, we applied various models for IPF data set. We considered 37 differently expressed shared random effects joint models. Table 6 summarizes the DICs of each submodel, the resulting fit, $\overline{\mathrm{D}}$, complexity, p_D, and the total DIC for each joint model. The joint models with both subject-specific random intercepts and slopes in both the longitudinal continuous outcome submodel and binary outcome submodel (Models 29 ~ 37) have the smallest DICs for both the longitudinal outcome submodels, and hence the $\mathrm{DIC}_{\mathrm{total}}$ of the overall model compared with other joint models with different structures of random effects. In the time-to-event submodel, adding a frailty term leads to a slight decrease in the $\mathrm{DIC}_{\mathrm{T}}$. The joint model connecting the time-to-event submodel to only the longitudinal submodels through the random intercepts does not improve the DICs at all. When both the random intercept and slopes, and frailty term were included to the association between the longitudinal submodels and the time-to-event submodel, there is the biggest decrease in $\mathrm{DIC}_{\mathrm{T}}$. However, there is no gain in the $\mathrm{DIC}_{\mathrm{T}}$ or the $\mathrm{DIC}_{\mathrm{total}}$ by including only the subject-specific random intercept of the longitudinal continuous outcome to W_3. Also there is miniscule loss in the $\mathrm{DIC}_{\mathrm{T}}$ in removing the frailty term from the time-to-event submodel. On the basis of model comparisons, Model 32 with the smallest $\mathrm{DIC}_{\mathrm{total}}$ was selected as our best model. Under Model 32, it seems that the longitudinal binary outcome shares the subject-specific random intercept and slope of the longitudinal continuous outcome and the time-to-event outcome is related to the subject-specific random slope of the longitudinal continuous outcome.

Table 6. Bayesian Model Selection of shared random effects joint model – IPF data.

Model	$U_1(t)$	$U_2(t)$	U_3	DIC_{y1}	DIC_{y2}	DIC_T	$\bar{D}$	p_D	DIC_{total}
	no random effects								
1	0	0	0	1795	1715	356.2	3856	11.02	3867
2	0	0	b_3	1795	1715	354.3	3842	23.11	3865
	random intercepts								
3	b_{10}	0	0	763.8	1715	356.1	2719	115.7	2835
4	b_{10}	0	b_3	763.9	1715	354.5	2706	127.9	2833
5	b_{10}	0	$\alpha_3 b_{10}$	763.6	1715	357.3	2720	116.3	2836
6	b_{10}	0	$\alpha_3 b_{10} + b_3$	763.6	1715	355.2	2706	128.2	2834
7	b_{10}	$\alpha_1 b_{10}$	0	719.5	1189	356.2	2143	122.5	2265
8	b_{10}	$\alpha_1 b_{10}$	b_3	718.8	1189	354.5	2128	134.3	2263
9	b_{10}	$\alpha_1 b_{10}$	$\alpha_3 b_{10}$	719.4	1189	357.8	2143	122.9	2266
10	b_{10}	$\alpha_1 b_{10}$	$\alpha_3 b_{10} + b_3$	718.8	1189	355.3	2128	134.7	2263
	random intercepts and random slopes								
11	$b_{10} + b_{11} t$	0	0	241.8	1715	356.2	2113	200.4	2313
12	$b_{10} + b_{11} t$	0	b_3	241.9	1715	354.2	2099	212.5	2311
13	$b_{10} + b_{11} t$	0	$\alpha_3 b_{10}$	241.7	1715	355.2	2111	200.8	2312
14	$b_{10} + b_{11} t$	0	$\alpha_4 b_{11}$	241.9	1715	350.4	2104	203.1	2307
15	$b_{10} + b_{11} t$	0	$\alpha_3 b_{10} + b_3$	241.9	1715	353.0	2098	212.2	2310
16	$b_{10} + b_{11} t$	0	$\alpha_4 b_{11} + b_3$	241.0	1715	351.5	2095	212.1	2307
17	$b_{10} + b_{11} t$	0	$\alpha_3 b_{10} + \alpha_4 b_{11}$	242.3	1715	351.1	2106	202.1	2308
18	$b_{10} + b_{11} t$	0	$\alpha_3 (b_{10} + b_{11})$	238.9	1715	352.7	2106	200.7	2307
19	$b_{10} + b_{11} t$	0	$\alpha_3 b_{10} + \alpha_4 b_{11} + b_3$	238.9	1715	349.7	2092	211.4	2304
20	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	0	374.3	1227	355.9	1742	215.1	1957
21	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	b_3	373.7	1227	354.6	1728	227.3	1955
22	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 b_{10}$	374.4	1226	357.5	1743	215.8	1958
23	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_4 b_{11}$	372.6	1226	346.6	1727	217.7	1945
24	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 b_{10} + b_3$	375.0	1228	355.1	1729	228.6	1958
25	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_4 b_{11} + b_3$	372.6	1227	348.0	1719	228.6	1947
26	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 b_{10} + \alpha_4 b_{11}$	372.8	1225	347.0	1726	218.9	1945
27	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 (b_{10} + b_{11})$	373.8	1227	356.2	1741	216.2	1957
28	$b_{10} + b_{11} t$	$\alpha_1 b_{10}$	$\alpha_3 b_{10} + \alpha_4 b_{11} + b_3$	373.6	1226	348.9	1719	229.4	1949
29	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	0	167.2	923.2	356.1	1235	211.6	1447
30	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	b_3	167.0	923.2	354.4	1221	223.9	1445
31	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 b_{10}$	168.6	922.4	356.3	1235	211.9	1447
32	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_4 b_{11}$	164.1	923.0	352.9	1229	210.6	1440
33	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 b_{10} + b_3$	166.0	923.1	353.4	1218	224.0	1442
34	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_4 b_{11} + b_3$	167.2	924.0	351.8	1219	224.3	1443
35	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 b_{10} + \alpha_4 b_{11}$	166.2	923.8	352.8	1230	212.7	1443
36	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 (b_{10} + b_{11})$	165.3	922.5	354.8	1232	211.0	1443
37	$b_{10} + b_{11} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} t$	$\alpha_3 b_{10} + \alpha_4 b_{11} + b_3$	167.6	924.3	351.0	1218	224.5	1443

Table 7. Joint Bayesian analysis results for shared random effects joint model – Model 32.

Parameter	Posterior Mean	Std. Error	95% CI
Longitudinal continuous submodel			
Intercept (β_{11})	0.167	0.048	(0.072, 0.263)
BaselineFVC (β_{12})	0.933	0.014	(0.906, 0.961)
Time (β_{13})	-0.094	0.016	(-0.122, -0.061)
$\sigma^2_{b_{10}}$	0.081	0.008	(0.066, 0.098)
$\sigma^2_{b_{11}}$	0.068	0.007	(0.056, 0.084)
ρ_{b_1}	-0.068	0.069	(-0.200, 0.067)
σ^2_{ϵ}	0.060	0.002	(0.055, 0.064)
Longitudinal binary submodel			
Intercept (β_{21})	-1.424	0.325	(-2.043, -0.796)
Age (β_{22})	-0.015	0.014	(-0.044, 0.012)
Time (β_{23})	1.019	0.180	(0.6513, 1.341)
α_1	-10.67	0.835	(-12.38, -9.145)
α_2	-10.31	0.727	(-11.82, -8.961)
Time-to-event submodel			
Intercept (β_{31})	-1.967	0.530	(-3.003, -0.939)
Male (β_{32})	0.903	0.360	(0.214, 1.648)
Smoking (β_{33})	0.719	0.341	(0.087, 1.422)
BaselineFVC (β_{34})	-0.405	0.203	(-0.807, -0.013)
α_4	-1.613	0.810	(-3.201, -0.025)

In Table 7, we present the posterior estimates for Model 32 of the fixed effects, random effects, and the association of the longitudinal outcomes and the time-to-event outcome along with their standard errors and 95% credibility intervals. For the longitudinal outcome, both the baseline FVC and time are statistically significant. Hence, patients with high values of baseline FVC have high FVC measurements and FVC measurements tend to decrease as time progresses. For the longitudinal binary outcome, we see that there is a significant time effect and it negatively shares both random intercept and slope of longitudinal continuous outcome. For the time-to-event outcome, male smokers with lower baseline FVC have a higher risk of death or transplant. The value of the association parameter, α_4 indicates that there is a negative association between the subject-specific random slopes of FVC measurements and the hazard of the transplant-free survival outcome significantly.

5.3.2 Correlated Random Effects Joint Model

We fit joint models similar to the shared random effects models. However, we excluded the frailty term in the time-to-event submodel because there was no improvement in adding this parameter in the shared random effects joint models. As with shared random effects joint models, the joint models with both subject-specific random intercepts and slopes in both longitudinal submodels (Models 21 ~ 36) have the smallest DICs, and the smallest DIC_{total} of the overall model (Table 8). The biggest decrease in DIC_T occurs when both the random intercepts and slopes in both longitudinal responses were included in the association between the longitudinal submodels and the time-to-event submodel. However, in this model, the parameter $\boldsymbol{\alpha_3}$ was not significant. Thus, we selected Model 33 with the second smallest DIC_{total} and all significant association parameters as our best model. Under Model 33, the time-to-event outcome is related to both the subject-specific random intercepts and slopes of the longitudinal continuous outcome and to the subject-specific random slope of the longitudinal binary outcome.

In Table 9 we present the posterior estimates, the standard errors, and their 95% credibility intervals for Model 33. We observe similar results with best shared random effects joint model for estimates of the fixed covariate effects although intercepts show lower estimates for all three submodels. For the time-to-event outcome submodel, focusing on the association parameters, we can see negative associations between the hazard of the transplant-free survival and the subject-specific random intercepts, and slopes of FVC measurements, respectively. From this model we also see the subject-specific random slopes of the disease progression indicator are associated with a higher hazard for the transplant-free survival.

Table 8. Bayesian Model Selection of Correlated random effects joint model – IPF data.

Model	$U_1(t)$	$U_2(t)$	U_3	DIC_{y1}	DIC_{y2}	DIC_T	$\bar{D}$	p_D	DIC_{total}
	random intercepts								
1	b_{10}	b_{20}	0	756.5	1266	356.0	2210	168.3	2379
2	b_{10}	b_{20}	$\alpha_1 b_{10}$	756.0	1267	357.3	2212	168.4	2380
3	b_{10}	b_{20}	$\alpha_3 b_{20}$	756.4	1268	357.4	2213	168.9	2382
4	b_{10}	b_{20}	$\alpha_1 b_{10} + \alpha_3 b_{20}$	755.7	1269	351.8	2206	170.3	2377
5	b_{10}	$b_{20} + b_{21} t$	0	756.9	1176	356.3	2111	178.4	2290
6	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10}$	757.4	1176	357.4	2112	179.3	2291
7	b_{10}	$b_{20} + b_{21} t$	$\alpha_3 b_{20}$	757.0	1177	353.3	2109	175.5	2288
8	b_{10}	$b_{20} + b_{21} t$	$\alpha_4 b_{21}$	756.9	1178	356.9	2113	179.0	2292
9	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_3 b_{20}$	757.9	1177	345.9	2105	176.1	2281
10	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_4 b_{21}$	755.8	1180	350.0	2105	180.5	2286
11	b_{10}	$b_{20} + b_{21} t$	$\alpha_3 b_{20} + \alpha_4 b_{21}$	757.7	1178	347.7	2095	188.4	2283
12	b_{10}	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_3 b_{20} + \alpha_4 b_{21}$	757.0	1178	344.2	2094	185.2	2280
	random intercepts and random slopes								
13	$b_{10} + b_{11} t$	b_{20}	0	240.3	1288	356.1	1627	257.1	1884
14	$b_{10} + b_{11} t$	b_{20}	$\alpha_1 b_{10}$	239.4	1289	355.6	1626	257.8	1884
15	$b_{10} + b_{11} t$	b_{20}	$\alpha_2 b_{11}$	239.8	1289	350.9	1621	258.3	1880
16	$b_{10} + b_{11} t$	b_{20}	$\alpha_3 b_{20}$	239.0	1290	357.1	1630	256.8	1886
17	$b_{10} + b_{11} t$	b_{20}	$\alpha_1 b_{10} + \alpha_2 b_{11}$	242.0	1288	351.7	1623	258.6	1881
18	$b_{10} + b_{11} t$	b_{20}	$\alpha_1 b_{10} + \alpha_3 b_{20}$	241.0	1292	354.7	1629	258.4	1888
19	$b_{10} + b_{11} t$	b_{20}	$\alpha_2 b_{11} + \alpha_3 b_{20}$	238.9	1292	349.0	1620	259.7	1880
20	$b_{10} + b_{11} t$	b_{20}	$\alpha_1 b_{10} + \alpha_2 b_{11} + \alpha_3 b_{20}$	239.6	1290	344.2	1613	261.3	1874
21	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	0	239.8	1188	356.2	1517	267.0	1784
22	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_1 b_{10}$	240.4	1190	355.3	1518	267.1	1785
23	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_2 b_{11}$	239.4	1190	351.1	1513	267.4	1780
24	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_3 b_{20}$	243.0	1190	352.4	1517	268.4	1785
25	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_4 b_{21}$	239.7	1190	356.6	1519	267.2	1787
26	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_2 b_{11}$	239.1	1190	351.3	1512	268.2	1780
27	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_3 b_{20}$	239.7	1192	347.1	1515	263.6	1779
28	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_4 b_{21}$	238.6	1190	355.4	1516	268.4	1784
29	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_2 b_{11} + \alpha_3 b_{20}$	239.9	1191	348.5	1512	267.4	1780
30	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_2 b_{11} + \alpha_4 b_{21}$	236.0	1191	347.9	1504	270.2	1775
31	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_3 b_{20} + \alpha_4 b_{21}$	239.5	1191	348.4	1504	275.1	1779
32	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} + \alpha_3 b_{20}$	241.2	1191	344.3	1506	270.4	1776
33	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} + \alpha_4 b_{21}$	238.8	1189	343.2	1497	273.6	1771
34	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_3 b_{20} + \alpha_4 b_{21}$	242.0	1190	340.5	1507	265.7	1772
35	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_2 b_{11} + \alpha_3 b_{20} + \alpha_4 b_{21}$	240.6	1188	343.2	1494	278.3	1772
36	$b_{10} + b_{11} t$	$b_{20} + b_{21} t$	$\alpha_1 b_{10} + \alpha_2 b_{11} + \alpha_3 b_{20} + \alpha_4 b_{21}$	238.2	1190	338.1	1495	271.0	1766

Table 9. Joint Bayesian analysis results correlated random effects joint model – Model 33.

Parameter	Posterior Mean	Std. Error	95% CI
Longitudinal continuous submodel			
Intercept (β_{11})	0.045	0.101	(-0.151, 0.254)
BaselineFVC (β_{12})	0.976	0.035	(0.904, 1.045)
Time (β_{13})	-0.091	0.020	(-0.132, -0.052)
σ_{ϵ}^2	0.061	0.002	(0.056, 0.065)
Longitudinal binary submodel			
Intercept (β_{21})	-0.810	0.144	(-1.093, -0.53)
Age (β_{22})	0.015	0.012	(-0.009 0.039)
Time (β_{23})	0.466	0.084	(0.306, 0.633)
Time-to-event submodel			
Intercept (β_{31})	-2.111	0.576	(-3.276, -1.004)
Male (β_{32})	0.989	0.390	(0.233, 1.752)
Smoking (β_{33})	0.741	0.357	(0.062, 1.476)
BaselineFVC (β_{34})	-0.398	0.215	(-0.813, 0.024)
α_1	-1.409	0.651	(-2.701, -0.125)
α_2	-3.866	1.195	(-6.227, -1.498)
α_4	-1.121	0.456	(-2.004, -0.223)
Variance components of Random effect			
$\sigma_{b_{10}}^2$	0.079	0.008	(0.064, 0.097)
$\sigma_{b_{11}}^2$	0.068	0.007	(0.055, 0.083)
$\sigma_{b_{20}}^2$	0.171	0.036	(0.114, 0.258)
$\sigma_{b_{21}}^2$	0.197	0.034	(0.139, 0.273)
ρ_{12}	-0.044	0.070	(-0.183, 0.093)
ρ_{13}	-0.230	0.079	(-0.382, -0.069)
ρ_{14}	-0.113	0.078	(-0.264, 0.044)
ρ_{23}	-0.109	0.080	(-0.264, 0.050)
ρ_{24}	-0.231	0.073	(-0.371, -0.082)
ρ_{34}	0.310	0.107	(0.085, 0.511)

5.4 COMPARISON OF THE RANDOM EFFECT STRUCTURE

We directly compare the random effect structures of the shared random effects joint models and the correlated random effects joint models from the results of DIC and estimations. The Model 7 of the shared random effects joint models named as 'shared_Model 7' and the Model 1 of the correlated random effects joint model named as 'corr_Model 1' allowed only subject-specific random intercept in both the longitudinal submodels.

Table 10 reports the estimation results for both shared_Model 7 and corr_Model 1. We observed consistent estimation results in all three submodels. In shared_Model 7, the association parameter, α_1, is negative and significant at $\alpha = 0.05$ indicating that the disease progression indicator shares the subject-specific random intercept of longitudinal FVC measurements negatively. The correlation between subject-specific random intercepts of both the longitudinal continuous and binary outcome shows a negative value in the corr_Model 1. The shared_Model 7 has smaller DICs in both the longitudinal continuous and binary submodels, and the total DIC. This difference is caused by complexity of the correlated random effects joint model. Similar comparisons exist between shared_Model 9 and corr_Model 2, both of which include association between the longitudinal outcomes and the time-to-event outcome in the time-to-event submodel (not shown).

When we considered models that allow both the subject-specific random intercept and slope in both the longitudinal continuous and binary outcome, we were able to investigate comparisons between shared_Model 29 and corr_Model 21, between shared_Model 31 and corr_Model 22, between shared_Model 32 and corr_Model 23, and between shared_Model 35 and corr_Model 26 which depended on the shared parameter in the time-to-event submodel. In all four pairs of models, the shared random effects joint models have smaller DICs than the

correlated random effects joint models. However, there is a restriction using the shared random effects joint model. We cannot use the shared random effects models that allow subject-specific random slopes in the longitudinal binary submodel without subject-specific random slopes in the longitudinal continuous submodel. Thus, if one considers the longitudinal continuous outcome with only subject-specific random intercept and longitudinal binary outcome with both the subject-specific random intercept and slope, it should be used with the correlated random effects joint model. A disadvantage of the correlated random effects joint model is that convergence issues may arise due to the large number of parameters. It is hard to control convergences for all parameters simultaneously.

Table 10. Joint Bayesian analysis results for the shared random effects joint model 7 and the correlated random effects model 1 – IPF data.

Parameter	Shared_Model 7			Corr_Model 1		
	Posterior Mean	Std. Error	95% CI	Posterior Mean	Std. Error	95% CI
Longitudinal continuous submodel						
Intercept (β_{11})	0.134	0.056	(0.023, 0.241)	0.112	0.101	(-0.090, 0.307)
BaselineFVC (β_{12})	0.923	0.017	(0.891, 0.955)	0.933	0.034	(0.865, 0.999)
Time (β_{13})	-0.039	0.004	(-0.048, -0.030)	-0.041	0.005	(-0.051, -0.031)
$\sigma^2_{b_{10}}$	0.096	0.013	(0.0723, 0.125)	0.099	0.010	(0.081, 0.121)
σ^2_{ϵ}	0.093	0.003	(0.086, 0.100)	0.093	0.003	(0.086, 0.101)
Longitudinal binary submodel						
Intercept (β_{21})	-0.424	0.259	(-0.934, 0.085)	-0.409	0.147	(-0.702, -0.125)
Age (β_{22})	-0.006	0.012	(-0.030, 0.017)	0.033	0.012	(0.010, 0.056)
Time (β_{23})	0.253	0.041	(0.175, 0.333)	0.204	0.038	(0.130, 0.280)
α_1	-7.27	0.517	(-8.324, -6.304)	-	-	-
$\sigma^2_{b_{20}}$	-	-	-	0.793	0.151	(0.540, 1.128)
$\rho_{b_{10}b_{20}}$	-	-	-	-0.537	0.055	(-0.638, -0.421)
Time-to-event submodel						
Intercept (β_{31})	-2.028	0.512	(-3.057, -1.040)	-2.018	0.450	(-3.006, -1.072)
Male (β_{32})	0.964	0.345	(0.288, 1.653)	0.959	0.347	(0.315, 1.672)
Smoking (β_{33})	0.725	0.345	(0.078, 1.431)	0.737	0.330	(0.115, 1.405)
BaselineFVC (β_{34})	-0.387	0.196	(-0.771, -0.0005)	-0.391	0.194	(-0.785, -0.018)
Information of DIC						
DIC_{y1}	719.5	-	-	756.5	-	-
DIC_{y2}	1189	-	-	1266	-	-
DIC_T	356.2	-	-	356.0	-	-
$\overline{D}$	2143	16.55	(1941, 2177)	2210	27.34	(2160, 2266)
p_D	122.5	-	-	168.3	-	-
DIC_{total}	2265	-	-	2379	-	-

5.5 PREDICTION OF PROBABILITIES OF EVENT OCCURING WITHIN A FIXED WINDOW

As discussed in section 3.4, Figure 6 shows the estimated fixed width failure probabilities with four year width for the IPF data set. We used estimated survival functions obtained from the best models selected in section 5.3 for both the shared random effects joint model and the correlated random effects joint model. For comparison, we also show the curves of two Cox models with or without the longitudinal measurements as a time-dependent covariate.

The two proposed-joint models show similar trends in the dynamic predictions of death or transplant within a window of four years. In general, the probability of event within the next four years decreases slowly. However, if a patient is still surviving and transplant-free after 5.25 years, the probability of event occurring within next four years rises higher by approximately ten percent. This reflects the very low ten-year IPF survival probability and the high likelihood of disease progression and therefore lung transplant after five years of follow-up in patients with a disease for which no FDA-approved therapy yet exists. The shared random effects joint model predicts a somewhat higher probability of transplant or death within the next four years than the correlated random effects joint model.

The dynamic predictions using Cox models show little differences. The curves oscillate and increase until four years, and then drop to zero after a steep rise at four years. When we used a time-dependent Cox model, the probability of transplant or death within the next four years was lower than the fixed Cox model not taking into account longitudinal measurements. Furthermore, the Cox model with the time-dependent covariates is also closer to those of our joint models.

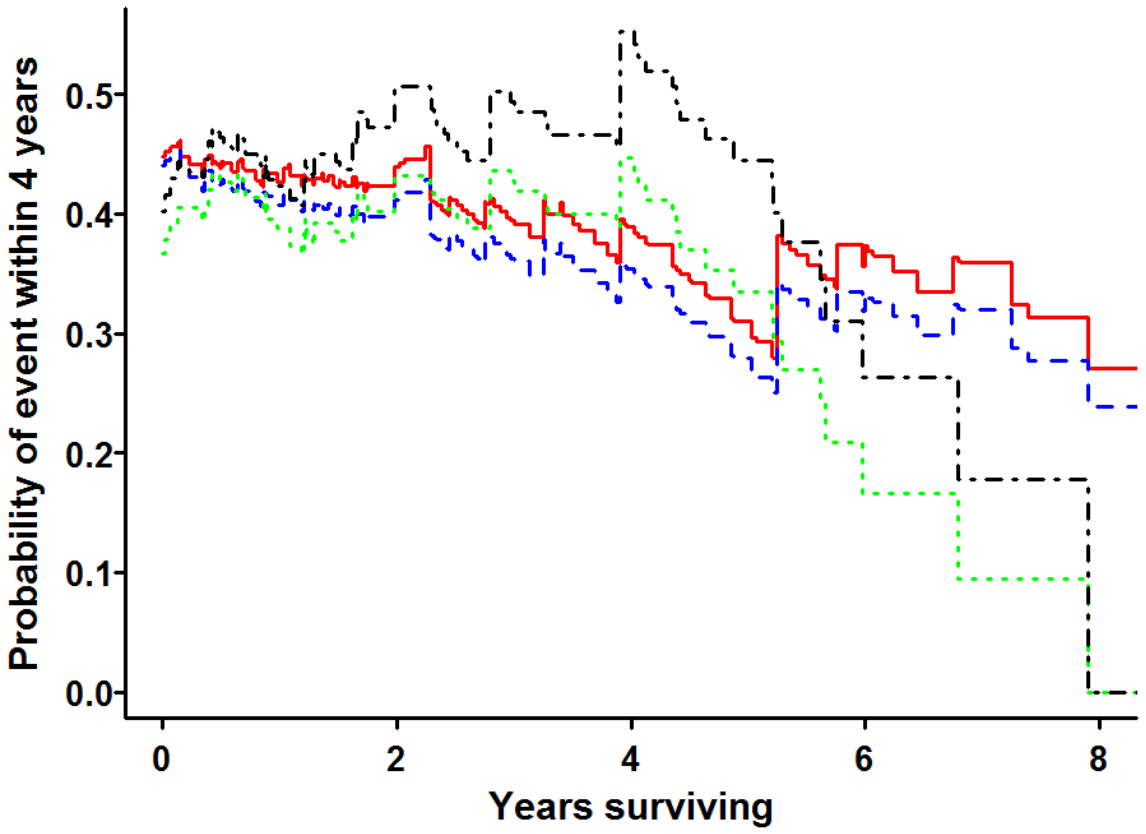

Figure 6. Probability of transplant or death within the next four years using shared random effects joint model 32 (*red solid curve*); the correlated random effects joint model 33 (*blue dashed curve*); Cox model with both the FVC measurements and disease progression indicator as a time-dependent covariate (*green dotted curve*); Cox model not taking into account longitudinal measurements at all (*black dot-dashed curve*).

6.0 SIMULATION STUDY

6.1 DESCRIPTION OF SIMULATED DATA

We conducted simulation studies to investigate the validity and comparison of the proposed joint models with the longitudinal continuous and binary outcomes and the time-to-event outcome. We considered two sets of simulation studies corresponding to the two random effects structures (shared random effects and correlated random effects) presented in section 3.1. The simulated longitudinal data consists of a quantitative outcome and of a dichotomous outcome with seven repeated measurements at fixed times 0, 0.5, 1, 1.5, 2, 2.5, and 3 years. We considered two sample sizes, N = 200 and N = 500. For both longitudinal outcomes, we considered the same fixed covariates. In particular, a quantitative covariate centered age (x_1) generated from the normal distribution, $x_1 \sim N(0, 5^2)$, a dichotomous covariate treatment (x_2) sampled from the Bernoulli distribution with the equal probability of 0.5, $x_2 \sim Bernoulli(0.5)$, and time (t), $t = \{0, 0.5, 1, 1.5, 2, 2.5, 3\}$, were included as fixed covariates and the subject-specific random intercept and random slope were assumed. For the longitudinal continuous outcome, the measurement error term is assumed to be normally distributed with mean zero and variance $\sigma^2_{\epsilon_1} = 1$.

The continuous longitudinal outcome was generated from the model

$$y_1(t) = \beta_{11} + \beta_{12}x_1 + \beta_{13}x_2 + \beta_{14}t + b_{10} + b_{11}t + \epsilon_1(t)$$

and the time-to-event outcome was generated from an exponential distribution, $T \sim exp\big(\mu_3(t)\big)$, where $\mu_3(t) = \exp(\beta_{31} + \beta_{32}x_2 + \alpha_3 b_{10} + \alpha_4 b_{11} + b_3)$ and $b_3 \sim N\big(0, \sigma^2_{b_3}\big)$ with $\sigma^2_{b_3} = 0.25$ for both the shared random effects joint model and the correlated random effects model. The binary longitudinal outcome was generated differently for the shared random effects joint model and the correlated random effects model. First, for the shared random effects joint model, the binary longitudinal outcome was generated from the model

$$\text{logit}(\Pr(y_2(t) = 1)) = \beta_{21} + \beta_{22}x_1 + \beta_{23}x_2 + \beta_{24}t + \alpha_1 b_{10} + \alpha_2 b_{11}t$$

and the random effects were assumed as,

$$\begin{pmatrix} b_{10} \\ b_{11} \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \mathbf{\Sigma} = \begin{bmatrix} \sigma^2_{b_{10}} & \rho_{12}\sigma_{b_{10}}\sigma_{b_{11}} \\ & \sigma^2_{b_{11}} \end{bmatrix} \right).$$

This shared random effects joint model has the same form with Model 37 in Table 2 and Table 6. Second, for the correlated random effects joint model, the binary longitudinal outcome was generated from the model

$$\text{logit}(\Pr(y_2(t) = 1)) = \beta_{21} + \beta_{22}x_1 + \beta_{23}x_2 + \beta_{24}t + b_{20} + b_{21}t$$

and the random effects were assumed as,

$$\begin{pmatrix} b_{10} \\ b_{11} \\ b_{20} \\ b_{21} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ \mathbf{\Sigma} = \begin{bmatrix} \sigma^2_{b_{10}} & \rho_{12}\sigma_{b_{10}}\sigma_{b_{11}} & \rho_{13}\sigma_{b_{10}}\sigma_{b_{20}} & \rho_{14}\sigma_{b_{10}}\sigma_{b_{21}} \\ & \sigma^2_{b_{11}} & \rho_{23}\sigma_{b_{11}}\sigma_{b_{20}} & \rho_{24}\sigma_{b_{11}}\sigma_{b_{21}} \\ & & \sigma^2_{b_{20}} & \rho_{34}\sigma_{b_{20}}\sigma_{b_{21}} \\ & & & \sigma^2_{b_{21}} \end{bmatrix} \right).$$

This correlated random effects joint model has similar form with Model 26 in Table 8 adding a frailty term in the time-to-event submodel. A non-informative censoring time C was generated from a uniform distribution on [0.2, 2] which resulted in roughly 35% censoring on average. The true parameter values for simulated data are presented in Table 11. For each random effect simulation study, 200 replications were conducted. In each analysis, a total of 15,000 MCMC iterations were used. However, we discarded the first 5,000 iterations as a burn-in period for each simulated sample.

Table 11. True parameter values for simulated data.

Continuous longitudinal outcome		Binary longitudinal outcome		Time-to-event outcome		Random effects	
Parameter	True	Parameter	True	Parameter	True	Parameter	True
β_{11}	5.0	β_{21}	0.3	β_{31}	0.2	$\sigma^2_{b_{10}}$	0.2
β_{12}	0.8	β_{22}	0.3	β_{32}	1.0	$\sigma^2_{b_{11}}$	0.25
β_{13}	-0.2	β_{23}	-0.2	α_1	1.6	$\sigma^2_{b_{20}}$	0.2
β_{14}	-0.2	β_{24}	-0.25	α_2	1.2	$\sigma^2_{b_{21}}$	0.25
$\sigma^2_{\epsilon_1}$	1.0			α_3	2.0	ρ_{12}	0.5
				α_4	3.5	ρ_{13}	0.5
				$\sigma^2_{b_3}$	0.25	ρ_{14}	0.5
						ρ_{23}	0.5
						ρ_{24}	0.5
						ρ_{34}	0.5

6.2 RESULTS OF SIMULATED DATA

The results of simulation study for the two random effects joint model are presented in Table 12 and Table 13, respectively, for N = 200 and N = 500. The results include the true parameters, the bias defined as the true parameters minus the estimated parameters, the standard errors of the

parameter estimates, the mean squared error, and the coverage probability of the estimated 95% credibility intervals. Comparing the estimates of the correlated random effects joint model, the estimates of shared random effects joint model are in general less biased and have the smaller standard errors and hence the mean squared errors. The shared random effects joint model shows better empirical coverage probability. In the simulation with large sample size (500), as expected, better results are obtained as indicated by the smaller Bias, SE, MSE, and CP. In particular, the coverage probabilities of variance of random effects are in a reasonable range.

Table 12. Simulation results of the shared random effect joint model.

		N = 200				N = 500			
Parameter	*True*	*Bias*	*SE*	*MSE*	*CP*	*Bias*	*SE*	*MSE*	*CP*
β_{11}	5.0	-0.015	0.078	0.014	0.915	0.003	0.051	0.005	0.935
β_{12}	0.8	0.008	0.107	0.029	0.880	-0.010	0.071	0.010	0.955
β_{13}	-0.2	0.001	0.009	0.000	0.925	0.000	0.006	0.000	0.950
β_{14}	-0.2	-0.005	0.044	0.004	0.910	-0.004	0.028	0.002	0.945
β_{21}	0.3	-0.009	0.164	0.054	0.945	-0.003	0.103	0.022	0.960
β_{22}	0.3	0.013	0.198	0.084	0.930	0.001	0.125	0.029	0.965
β_{23}	-0.2	-0.001	0.020	0.001	0.970	-0.001	0.012	0.000	0.940
β_{24}	-0.25	-0.015	0.083	0.015	0.920	-0.004	0.052	0.006	0.925
β_{31}	0.2	-0.074	0.266	0.156	0.940	-0.021	0.171	0.059	0.950
β_{32}	1.0	0.031	0.307	0.216	0.915	-0.002	0.194	0.072	0.985
α_1	1.6	0.280	0.331	0.296	0.890	0.147	0.198	0.098	0.870
α_2	1.2	0.040	0.161	0.055	0.935	-0.012	0.100	0.022	0.930
α_3	2.0	0.164	0.585	0.595	0.980	0.052	0.344	0.211	0.995
α_4	3.5	0.153	0.457	0.379	0.985	0.087	0.291	0.158	0.960
$\sigma^2_{\epsilon_1}$	1.0	0.023	0.043	0.004	0.940	0.008	0.027	0.001	0.975
$\sigma^2_{b_3}$	0.25	0.001	0.055	0.003	1.000	-0.011	0.049	0.003	1.000
$\sigma^2_{b_{10}}$	0.2	-0.024	0.026	0.002	0.865	0.019	0.026	0.001	0.955
$\sigma^2_{b_{11}}$	0.25	-0.016	0.025	0.001	0.895	0.013	0.021	0.001	0.950
ρ_{12}	0.5	-0.027	0.067	0.007	0.995	0.043	0.052	0.006	0.945

Table 13. Simulation results of correlated random effect joint model.

		N = 200				N = 500			
Parameter	*True*	*Bias*	*SE*	*MSE*	*CP*	*Bias*	*SE*	*MSE*	*CP*
β_{11}	5.0	0.011	0.080	0.015	0.895	-0.006	0.052	0.006	0.915
β_{12}	0.8	-0.018	0.110	0.026	0.930	0.003	0.073	0.011	0.945
β_{13}	-0.2	0.000	0.004	0.000	0.960	0.000	0.003	0.000	0.930
β_{14}	-0.2	-0.006	0.043	0.004	0.915	0.003	0.028	0.002	0.920
β_{21}	0.3	-0.014	0.155	0.054	0.910	-0.001	0.101	0.022	0.920
β_{22}	0.3	-0.026	0.176	0.071	0.905	-0.010	0.117	0.031	0.925
β_{23}	-0.2	0.014	0.012	0.000	0.775	0.008	0.008	0.000	0.790
β_{24}	-0.25	0.020	0.078	0.014	0.910	0.006	0.051	0.005	0.960
β_{31}	0.2	-0.014	0.268	0.173	0.910	-0.008	0.172	0.059	0.955
β_{32}	1.0	-0.012	0.315	0.211	0.950	0.011	0.200	0.076	0.975
α_3	2.0	0.086	0.650	0.741	0.970	0.071	0.390	0.272	0.985
α_4	3.5	0.163	0.505	0.469	0.965	0.045	0.325	0.190	0.970
$\sigma^2_{\epsilon_1}$	1.0	0.010	0.043	0.004	0.955	0.012	0.027	0.002	0.930
$\sigma^2_{b_3}$	0.25	0.003	0.056	0.003	1.000	-0.009	0.050	0.003	1.000
$\sigma^2_{b_{10}}$	0.2	-0.015	0.028	0.002	0.910	0.026	0.027	0.002	0.930
$\sigma^2_{b_{11}}$	0.25	-0.015	0.026	0.001	0.895	0.013	0.021	0.001	0.940
$\sigma^2_{b_{20}}$	0.2	-0.064	0.026	0.005	0.250	-0.032	0.033	0.003	0.885
$\sigma^2_{b_{21}}$	0.25	-0.077	0.030	0.007	0.230	-0.030	0.033	0.003	0.880
ρ_{12}	0.5	-0.022	0.068	0.007	1.000	0.052	0.054	0.007	0.925
ρ_{13}	0.5	-0.218	0.114	0.064	0.540	-0.087	0.094	0.020	0.985
ρ_{14}	0.5	-0.116	0.088	0.025	0.855	-0.007	0.067	0.007	0.995
ρ_{23}	0.5	-0.179	0.116	0.049	0.850	-0.068	0.092	0.017	0.990
ρ_{24}	0.5	-0.066	0.079	0.014	0.955	0.011	0.059	0.005	0.995
ρ_{34}	0.5	-0.226	0.107	0.066	0.385	-0.077	0.087	0.016	0.985

7.0 CONCLUSIONS

7.1 SUMMARY

In this dissertation, we proposed a joint model of mixed types of multivariate longitudinal continuous and binary outcomes and a time-to-event outcome. We assume that the longitudinal outcomes and the time-to-event outcome depend on shared parameters induced from the subject-specific random effects of the longitudinal outcomes. We considered two types of random effects structure, that is, shared random effects and correlated random effects to characterize the relationship between the longitudinal continuous and binary outcomes. We used a Bayesian approach for estimating parameters of the proposed joint model through MCMC methods. Through various model comparisons, we selected a "best" model using the DIC for model selection criterion. Our joint models were illustrated by application on the MTLD data set and IPF data set. Although we could not get satisfying results for the correlated random effects joint model in the MTLD data set analysis, we observed that both the shared random effects joint model and the correlated random effects model provided consistent results for the IPF data application. A disadvantage of the shared random effects joint model is that the structure of the random effects for both the longitudinal outcomes is limited. However, we can gain efficiency by using a smaller number of random effects parameters. In contrast, the correlated random effects

joint model can characterize the dependency between the longitudinal continuous and binary outcome more freely. We illustrated the dynamic prediction of the probabilities of events occurring within a fixed window of time. Given a subject is at risk just before time t, the probability of an event occurring within the next fixed window of time is predicted using the survival function obtained from our proposed joint models. From two sets of simulation studies of differing sample sizes, we found that the proposed joint models performed reasonably under both the random effects structures and larger sample size.

7.2 FUTURE RESEARCH

One future research direction of this research is assessing the predictive accuracy of the joint models and the prediction of event probability. There are two main approaches. One is focused on calibration measures (Schemper and Henderson, 2000; Henderson et al., 2002). Another approach is focused on discrimination measures (Heagerty et al., 2000; Heagerty and Zheng, 2005). To assess the predictive accuracy, we will derive the estimates of time-dependent sensitivity and specificity measures and the time-dependent receiver operating characteristic (ROC) curves and area under the curve (AUC) estimates of prediction performance under the joint modeling framework.

A further direction of this research would be to evaluate the adequacy of the DIC used as the model selection criterion. As we presented in section 4.4.2, there were inconclusive results where negative estimates of effective numbers of parameters were obtained in sub-components of DIC. Thus, to check the adequacy of the DIC for testing model fits would be useful to further develop the Bayesian paradigm for addressing joint modeling problems.

APPENDIX

POSTERIOR ESTIMATION R CODE INTERFACE TO OPENBUGS

```
## Model_33 : correlated random effects joint model
##           ( b10+b11*t / b21+b22*t / a1b10+a2b11+a4b22 )
model.name <- paste("Model_33.txt",sep="")
  write("model{
    for (j in 1:NJ) {
     # Continuous outcome
         Y[j] ~ dnorm(muy[j], tauz)
         muy[j] <- beta1[1]+beta1[2]*blFvc[j]+beta1[3]*time[j]
                   +U[u[j],1]+U[u[j],2]*time[j]
     # Binary outcome
         Y_b[j] ~ dbern(b[j])
         logit(b[j]) <- beta2[1]+beta2[2]*Age_at_Dx_ct[j]+beta2[3]*time[j]
                        +U[u[j],3]+U[u[j],4]*time[j]
    } # end of j loop

    for (i in 1:N) {
    # Survival Model
         surt[i] ~ dweib(p,mut[i]) I(surt.cen[i],)
         log(mut[i]) <- beta3[1]+beta3[2]*SEX[i]+beta3[3]*SMOKING[i]+beta3[4]*BLFVC[i]
                        +a1*U[i,1]+a2*U[i,2]+a4*U[i,4]

     # Subject-specific parameters
      U[i,1:4]  ~ dmnorm(U0[],tau[,])
    }  # end of i loop

p <- 1

sigmaz <- 1/tauz

sigma[1:4,1:4]<-inverse(tau[,])
sigma1<-sigma[1,1]
sigma2<-sigma[2,2]
sigma3<-sigma[3,3]
sigma4<-sigma[4,4]
sigma12<-sigma[1,2]
sigma13<-sigma[1,3]
sigma14<-sigma[1,4]
sigma23<-sigma[2,3]
sigma24<-sigma[2,4]
sigma34<-sigma[3,4]
```

```
cor12<-sigma12/(sqrt(sigma1*sigma2))
cor13<-sigma13/(sqrt(sigma1*sigma3))
cor14<-sigma14/(sqrt(sigma1*sigma4))
cor23<-sigma23/(sqrt(sigma2*sigma3))
cor24<-sigma24/(sqrt(sigma2*sigma4))
cor34<-sigma34/(sqrt(sigma3*sigma4))

#priors
tauz~dgamma(0.1, 0.1)
tau[1:4,1:4] ~ dwish(R[,], 100)
beta1[1:3]~dmnorm(betamu1[],Sigma1[,])
beta2[1:3]~dmnorm(betamu2[],Sigma2[,])
beta3[1:4]~dmnorm(betamu3[],Sigma3[,])
a1~dnorm(0, 0.01)
a2~dnorm(0, 0.01)
a4~dnorm(0, 0.01)
  } # end of BUGS code",file=model.name)

  ## data set
DATA <- list(NJ=nrow(PA_PFT),N=nrow(PA),
betamu1=c(0,0,0),
betamu2=c(0,0,0),
betamu3=c(0,0,0,0),
Sigma1=diag(0.01,3,3),
Sigma2=diag(0.01,3,3),
Sigma3=diag(0.01,4,4),
U0=c(0,0,0,0),
R=diag(10,4,4),
Y=PA_PFT$PFT_FVC,
Y_b=PA_PFT$FvcEvent05,
u=PA_PFT$PTnum,
Age_at_Dx_ct=PA_PFT$Age_at_Dx_ct,
blFvc=PA_PFT$blFvc,
time=PA_PFT$time,
SEX=as.integer(PA$Sex=="M"),
surt=surt,
surt.cen=surt.cen,
SMOKING=PA$Smoking, BLFVC=PA$blFvc
)

INT1 <- list(beta1=c(0,0,0),beta2=c(0,0,0),beta3=c(0,0,0,0),tauz=1,
             a1=1,a2=1,a4=1,U=matrix(0,nrow=nrow(PA),ncol=4))

params <- c("beta1","beta2","beta3","sigmaz","tauz","sigma1","sigma2","sigma3",
            "sigma4","cor12","cor13","cor14","cor23","cor24","cor34","a1","a2","a4")

library(rbugs)
Corr_Model_33 <- rbugs(data=DATA,inits=list(INT1),paramSet=params,
                       model='Model_33.txt',n.chains=1,
                       n.iter=15000,n.burnin=5000,n.thin=1,dic=TRUE,seed=0,
                       bugs=OpenBUGS,OpenBugs=T,bugsWorkingDir=getwd())
```

BIBLIOGRAPHY

Carlin, B. P., and Louis, T. A. (2009) *Bayesian Methods for Data Analysis* (3rd edition). Boca Raton, FL: Chapman and Hall/CRC Press

Catalano, P. J. (1997). Bivariate modelling of clustered continuous and ordered categorical outcomes. *Statistics in Medicine* **16**, 883-900

Catalano, P. J., and Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *American Statistical Association* **87**, 651-658

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

Cox, D. R., and Hinkley, D. (1974). *Theoretical Statistics*. London: Chapman and Hall

Cox, D. R., and Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika* **79**, 441-461

Ding, J., and Wang, J. L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* **64**, 546-556

Dunsmore, I. R. (1974). The bayesian predictive distribution in life testing models. *Technometrics* **16**, 455-460

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, 1041−1046

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk. *The Canadian Journal of Statistics* **14**, 257-262

Faucett, C. J., and Thomas, D. C. (1996). Simultaneously modeling censored survival data and repeated measured covariates: A Gibbs sampling approach. *Statistics in Medicine* **15**, 1663-1685

Fieuws, S., and Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statistics in Medicine* **23**, 3093-3104

Fieuws, S., and Verbeke, G. (2005). Evaluation of the pairwise approach for fitting joint linear mixed models: A simulation study. Technical Report TR0527, Biostatistical Centre, Katholieke Universiteit Leuven, Belgium

Fieuws, S., and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* **62**, 424-431

Fieuws, S., Verbeke G., Maes, B., and Vanrenterghem, Y. (2008). Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics* **9**, 419-431

Fitzmaurice, G. M., Davidian, M., Verbeke, G., and Molenberghs, G. (2009). *Longitudinal Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC Press

Fitzmaurice, G. M., and Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *American Statistical Association* **90**, 845-852

Fitzmaurice, G. M., Laird, N. M., and Ware, J. (2004). *Applied Longitudinal Data*. Hoboken, New Jersey: Wiley

Garre, F. G., Zwinderman, A. H., Geskus, R. B., Sijpkens, Y. W. J. (2008). A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of Royal Statistical Society: Series A* **171**, 299-308

Gueorguieva, R. V., and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *American Statistical Association* **96**, 1102-1112

Gueorguieva, R. V., and Sanacora, G. (2006). Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. *Statistics in Medicine* **25**, 1307-1322

Guo, Xu, and Carlin, B. P. (2004) Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* **58**, 16-24

Heagerty, P. J., Lumley, T., and Pepe, M. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**, 337-344

Heagerty, P. J. and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92-105

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

Henderson, R., Diggle, P., and Dobson, A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics* **3**, 33-55

Hogan, J. W., and Laird, N. M. (1997a). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16**, 239-257

Hogan, J. W., and Laird, N. M. (1997b). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16**, 259-272

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine studies. *Statistica Sinica* **14**, 847-867

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data* (2nd edition). New York: Wiley

Kuk, A. Y. C. and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531–541

Lauritzen, S. L., and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The annals of Statistics* **17**, 31-57

Laird, N. M., and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974

Law, N. J., Taylor, J. M. G., and Sandler, H. (2002). The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics* **3**, 547-563

Li, C. S., Taylor, J. M. G., and Sy, J. P. (2001). Identifiability of cure models. *Statistics and Probability Letters* **54**, 389-395

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125-134

Liu, X., Daniels, M. J., and Marcus, B. (2010). Joint models for the association of longitudinal binary and continuous processes with application to a smoking cessation trial. *Journal of American Statistical Association* **104**, 429-438

Lockwood, J. R., and Schervish, M. J. (2005). MCMC strategies for computing bayesian predictive densities for censored multivariate data. *Journal of Computational and Graphical Statistics* **14**, 395-414

Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer

Molenberghs, G., Verbeke, G., Demetrio, C. G. B., and Vieira, A. M. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science* **25**, 325-347

Olkin, I., and Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann of Math Stat* **32**, 448-465

Pauler, D.K., and Finkelstein, D.M. (2002). Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Medicine* **21**, 3897-3911

Proust-Lima, C., Taylor, J. M. G. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* **10**, 535-549

Putter, H., van der Hage, J., de Bock, G.H., Elgalta, R., and van der Velde, C. J. H. (2006). Estimation and prediction in a multistate model for breast cancer. *Biometrical Journal* **48**, 366–380

Reynolds, C. F., Butters, M. A., Lopez, O., Pollock, B. G., Dew, M. A., Mulsant, B. H., Lenze, E. J., Holm, M., Rogers, J. C., Mazumdar, S., Houck, P. R., Begley, A., Anderson, S., Karp, J. F., Miller, M. D., Whyte, E. M., Stack, J., Gildengers, A., Szanto, K., Bensasi, S., Kaufer, D. I., Kamboh, M. I., and DeKosky, S. T. (2011) Maintenance treatment of depression in old age. *Arch Gen Psychiatry* **68** (1), 51-60

Richards, T. J., Kaminski,N., Baribaud, F., Flavin, S., Brodmerkel, C., Horowitz, D., Li, K., Choi, J., Vuga, L. J., Lindell, K. O., Klesen, M., Zhang, Y., Gibsonm K. F. (2012) Peripheral Blood Proteins Predict Mortality in Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine* **185**, 67-76Rizopoulos, D. (2010). An R package for the joint modelling of longitudinal and time-to-event data. *Journal of statistical software* **35** (9), 1-33

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data Biometrics. *Biometrics* **67**, 819-829

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

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika* **95**, 63-74

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2010). Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* **66**, 20-29

Roeder, K., Lynch, K. G., and Nagin, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *American Statistical Association* **94**, 766-776

Schemper, M. and Henderson, R. (2000). Predictive accuracy and explained variation in Cox regression. *Biometrics* **56**, 249-255

Song, X., Davidian, M., and Tsiatis, A. A. (2002a). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742-753

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society*, Series. B., **64**, 583–639

Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics* **51**, 899-907

Taylor, J., Yu, M., and Sandler, H. M. (2004). Individualized predictions of disease progression following radiation therapy for prestate cancer. The university of Michigan department of biostatistics working paper series 2004

Tsiatis, A. A., and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* **14**, 809-834

Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M. S. (1995). 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

van Houwelingen H. C., Putter H. (2011). *Dynamic Prediction in Clinical Survival Analysis*. Boca Raton, FL: Chapman and Hall/CRC Press

Wang, Y. and Taylor, J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* **96**, 895-905

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

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375-387

Yu, M., Law, N. J., Taylor, J. M. G., and Sandler, H. M. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica* **14**, 835-862

Yu, M., Taylor, J., and Sandler, H. M. (2004). Individual prediction in prostate cancer studies using a joint longitudinal-survival-cure model. The university of Michigan department of biostatistics working paper series 2004

Zeger, S. L., and Liang K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130