BAYESIAN HIERARCHICAL JOINT MODELING OF REPEATEDLY MEASURED MIXED BIOMARKERS OF DISEASE SEVERITY AND TIME-TO-EVENT

by

Olive D. Buhule

MSc. Biostatistics, University of Minnesota, 2010

MStat., Makerere University, Uganda, 2006

BSc. Quantitative Economics, Makerere University, Uganda, 2003

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

2014

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Olive D. Buhule

It was defended on

May 28, 2014

and approved by

Ada O. Youk, PhD

Assistant Professor of Biostatistics

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Abdus S. Wahed, PhD

Associate Professor of Biostatistics

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Vincent C. Arena, PhD

Associate Professor of Biostatistics

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Daniel E. Weeks, PhD

Professor of Human Genetics and Biostatistics,

Department of Human Genetics

Graduate School of Public Health

University of Pittsburgh

Dissertation Advisors: Ada O. Youk, PhD

Assistant Professor of Biostatistics

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

,

Abdus S. Wahed, PhD

Associate Professor of Biostatistics

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

BAYESIAN HIERARCHICAL JOINT MODELING OF REPEATEDLY MEASURED MIXED BIOMARKERS OF DISEASE SEVERITY AND TIME-TO-EVENT

Olive D. Buhule, PhD

University of Pittsburgh, 2014

ABSTRACT

In many clinical follow-up studies, patients are observed at irregular intervals for more than one biomarker of disease severity. Although these biomarkers are often meant to measure the same disease severity, they may differ due to the instruments or reagents used as well as the scale of measurements. They could show different patterns for treatment because clinicians prescribe medications based on the severity of disease. Moreover, if these markers are modeled separately to determine the factors that are associated with disease progression over time or to predict the event of interest given different treatments, they may yield misleading or inefficient results. Joint modeling of correlated biomarkers alone or with time-to-event data leads to efficient results, hence better clinical decisions.

In this study, we have first developed a joint model to analyze multivariate unbalanced repeatedly measured outcomes of mixed types, in particular, continuous and ordinal outcomes. Secondly, we have extended the first model to include time-to-event data. The postulated models assumes that the outcomes are from distributions that are in the exponential family and hence modeled as a multivariate generalized linear mixed effects model linked through random effects. The Markov Chain Monte Carlo (MCMC) Bayesian approach is used to approximate the posterior distribution and draw inference on the parameters. These joint models provide a flexible framework to account for the hierarchical structure of the highly unbalanced data as well as the association between the multiple mixed types of outcomes and

time-to-event. Moreover, the simulation studies show that estimates obtained from the joint models are consistently less biased and more efficient than those obtained from the separate models. We applied our models to diabetes data from an observational study.

Diabetes and its associated complications such as heart attack and stroke are of serious public health concerns across the globe. Proper treatment can help control and prevent the development of these complications and hence improve the quality of life of millions of people. This work proposes to efficiently estimate the treatment effect by introducing state-of-the-art statistical methods. This will help researchers identify effective treatments that can slow down the disease progression.

Keywords: Diabetes; Generalized Linear Mixed Effects Models; Hierarchical Modeling; Joint Modeling; Mixed Biomarkers; MCMC; Time-To-Event.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First and foremost I would like to express my gratitude to my advisors, Dr. Abdus S. Wahed and Dr. Ada O. Youk, to whom I am much obliged. Throughout the entire period, they have provided candid and useful ideas, suggestions and comments, and have insisted on the highest standards for the conducted research. The quality of this dissertation is thus largely attributable to their supervision. I am extremely grateful to them for their immeasurable support and guidance that has helped reshape my future. I am also indebted to my committee members, Dr. Vincent C. Arena and Dr. Daniel E. Weeks for their unexplainable kindness accorded to me. They have been there for me when everything seemed to come to an end, and they have provided valuable comments and suggestions that have enriched this dissertation work. Special thanks go to Dr. Stewart Anderson and Dr. Sati Mazumdar for their support during the early stages of this work. I am indeed very grateful.

In addition, I thank all the Professors of this institution for the knowledge imparted, which has been helpful in this research, and not forgetting the Systems Administrator, Steve Patterson for his invaluable assistance with the virtual lab to run my simulations. I am especially grateful to the Chair, Dr. Sally Morton for all the support rendered to me. She has always been there for me every time I needed her help.

I wish to thank the administration of Mulago, Rubaga, and Nsambya hospitals in Uganda for providing the data used in this study. In addition, I am grateful to the Uganda National Council for Science and Technology for granting permission to use these data.

I would like also to extend my sincere thanks to all my friends especially, Cyprian, Emmanuel, Margaret, and Susan for their encouragement and support that has helped me stay focused. I am exceedingly grateful to Margaret and Susan for their endless and motherly love accorded to my son. Thank you for being there for him on my behalf. I am very thankful

to my parents too, for bringing me into this world and for their kind support. Also not forgetting other persons who directly or indirectly rendered assistance to me. I am extremely grateful. Above all, I thank the Almighty God who made everything possible for me.

Finally, I dedicate this work to my son Emmanuel. Thank you for being a loving, caring, and an understanding child. May God be with you always!

1.0 INTRODUCTION

In the first part of this dissertation, we present a Bayesian hierarchical joint model of repeatedly measured continuous and ordinal markers of diseases severity for highly unbalanced data. The two outcomes are assumed to be from distributions that are in the exponential family and hence modeled as a multivariate generalized linear mixed effects model linked through correlated and/or shared random effects. The Markov Chain Monte Carlo (MCMC) Bayesian approach is used to approximate the posterior distribution and draw inference on the parameters. In the second part of this dissertation, we extend the joint model to include time-to-event data. We employ same Bayesian methods for parameter estimation.

This dissertation is organized as follows. In Chapter 1, we give a background to our study that includes both the clinical and statistical motivation. Chapter 2 shows the methods for longitudinal data modeling and time-to-event or survival analysis in the univariate and multivariate settings. We plan to write Chapter 3 and Chapter 4 of this dissertation as independent papers, and because Chapter 4 builds on Chapter 3, there are likely to be some repetitions in some sections.

1.1 BACKGROUND

In many clinical studies, more than one biomarker of disease severity is obtained and some may be easier and cheaper to obtain than others. Although these biomarkers are often meant to measure the same disease severity, they may differ due to the instruments/reagents used as well as the scale of measurements. They could show different patterns for treatment because clinicians prescribe medications based on the severity of disease. Moreover, if these markers

are modeled separately to determine the factors that are associated with disease progression over time or to predict the event of interest (i.e., time-to-remission) given different treatments, they may yield different, misleading or inefficient results. Modeling these markers jointly while accounting for their correlation is likely to provide more accurate and efficient results.

The motivation for our study is based on data collected retrospectively from medical registries of diabetic patients in three Ugandan hospitals. These patients were recruited in the diabetic clinics between January 1992 and December 2004. Diabetes is a progressive illness occurs when the pancreas does not produce enough insulin or when the body does not respond properly to varying levels of insulin. This results in increased concentrations of glucose in the blood, which in turn damages many of the body's systems, in particular the blood vessels and nerves. Thus, the amount of glucose in the blood determines the state of the disease at a point in time. In addition, the amount of glucose in the urine is used to detect if the individual's blood glucose level is above the renal threshold of 180 mg/dl. The amount of glucose in the urine is interpreted using the + symbolic method or the actual amount in the urine(mg/dl) depending on the manufacturer of the urine glucose reagent strips. That is, Nil (no urine sugar), $+(\approx 100$ mg/dl), $++$ (≈ 250 mg/dl), $+++(\approx 500$ mg/dl), and $++++$ (≈ 1000 mg/dl), respectively. Thus, the two main biomarkers used to determine the severity of the disease at any given time were blood glucose and urine glucose levels. Of the two biomarkers, blood glucose which is measured by the fasting plasma glucose test (FPG) or the oral glucose tolerance test (OGTT) is more accurate and hence recommended by both the World Health Organization (WHO) and the American Diabetes Association (ADA). The urine glucose tests to detect Glycosuria/Glucosuria (glucose in urine) are used as an alternative to blood glucose tests especially in developing countries because they are fast, do not require many reagents, easy to carry out and generally economical (Carter and Lema, 2003)[10]. However, the urine glucose test for diabetes may be contaminated by drugs and individual variations in the renal threshold for glucose. Thus, making a clinical decision based on a urine test alone may be invalid or misleading.

According to the ADA, normal blood glucose level for diabetics is between 70 and 180 mg/dl. Specifically, normal fasting blood glucose (before a meal) is between 70 and 130 mg/dl and after a meal is less than 180 mg/dl. Blood glucose level below 70 mg/dl is referred to as

hypoglycemia and above 180 mg/dl is referred to as hyperglycemia. Thus, the clinical interest is to detect that the blood glucose lies in the normal range of 70-180 mg/dl. Thus, a diabetic person is said to be well if the blood glucose level is in the normal range. In addition, having normal blood pressure can circumvent most of the common diabetes complications that include diabetic retinopathy (which is damage to the back of the eye) and kidney damage medically known as diabetic nephropathy. Studies have indicated that individuals with adequately controlled blood pressure possess lower risk of mortality related to diabetes complications such as heart attack and stroke. Moreover, body mass index (BMI) which is a measure of body fat, is positively associated with Type 2 diabetes mellitus and hypertension (high blood pressure). Thus, having normal blood pressure for diabetics is as important as having good control of blood glucose levels and BMI.

During their hospital visits, the patients in the Ugandan study were periodically tested for the amount of glucose in the blood or urine or both to determine the severity of the disease so as to prescribe appropriate medication. Other measures that are associated with diabetes like diastolic and systolic blood pressure, and body mass index (BMI) were also taken. The data are highly unbalanced because this was an observational study where patients reported for checkup at irregular intervals with the number of hospital visits varying from patient to patient. Specifically, the number of hospital visits per patient varied from 2 to 78 with a mean of 29 and standard deviation of 15. Out of 1010 patients, 301 were treated with Sulphonyureas, 299 with Biguanides, 402 with Insulin, and 8 were on diet and exercise at baseline. Clinically, blood glucose and urine glucose measure the same diabetes severity although blood glucose is continuous and urine glucose is ordinal with five levels. In addition, blood glucose, which is expensive to measure is associated with inexpensive biomarkers such as blood pressure levels and body mass index. Joint modeling of the two markers (blood glucose and urine glucose) simultaneously will produce efficient estimates because the two markers are highly correlated. Moreover, a joint model that combines time to normalization of blood glucose, blood pressure levels, and body mass index will lead to a more optimal way of caring for diabetic patients. This will be beneficial to the patients, and to the health-care personnel and institutions.

Statistically, joint modeling allows for the assessment of the overall impact as well as the separate and joint effects of a risk factor or treatment on all the outcomes while adjusting for the correlation that exists between or among these outcomes. Joint modeling avoids multiple testing by calculating an overall test of the effect of the predictor without having to resort to ad hoc methods such as Bonferroni adjustment. Overall, joint modeling leads to more efficient estimates than separate analyses.

The goal of this study is to propose a joint model that handles unbalanced repeatedly measured outcomes of mixed types and time-to-event. Specifically, we propose to develop two models:

Aim 1: Joint model for unbalanced repeatedly measured continuous and ordinal outcomes that are measures of disease severity.

Aim 2: Joint model for repeatedly measured mixed outcomes and time-to-event data.

In the next chapter (Chapter 2), we give a critical review of the literature related to modeling of longitudinal outcomes of mixed types and time-to-event, and briefly describe our contribution. Chapters 3 and 4 shows the completed work on Aim 1 and Aim 2, respectively.

2.0 METHODOLOGICAL REVIEW

Joint modeling of continuous and ordinal response variables are of primary interest in Part 1 (Chapter 3) of this study. The two outcome variables (e.g., urine glucose and blood glucose) are measured repeatedly on each study participant at the same or varying time-points. The important feature of our data is that it is highly unbalanced, and therefore, the methods that handle these kinds of data are of paramount interest. Part 2 (Chapter 4) of this study extends the joint model in Part 1 (Chapter 3) to include time-to-event. Thus, an understanding of the treatment effects on both markers over time and time-to-event as well as modeling the correlation between and/or among them are of interest. In this chapter, methods for analyzing longitudinal continuous and discrete (binary, ordinal) outcomes, and time-to-event (separately and jointly) are reviewed. These include marginal, generalized linear mixed effects, survival, joint modeling of continuous and discrete (binary or ordinal) outcomes, and joint modeling of longitudinal outcomes and time-to-event.

2.1 METHODS FOR LONGITUDINAL DATA MODELING

In longitudinal or repeated measured studies, the key issues are to capture the change in a response over time as well as the within subject change or to account for the correlation between the measurements. In addition, follow-up studies may have staggered entry, dropout, intermittent missing data, and mistimed visits, which results in unbalanced datasets. Thus, in longitudinal data analysis, appropriate or realistic methods that handle unbalanced data should be employed to obtain reliable results. Modern linear mixed effects models (Laird and Ware, 1982)[57] or generalized estimating equations (Zeger and Liang, 1986)[103]

approaches accommodate these unbalanced data sets, subject to assumptions about the missing data mechanism (Little and Rubin, 1987)[62]. Older standard methods for univariate and multivariate analysis of repeated measures data are more restrictive in that they require balanced datasets.

2.1.1 Notation

Let y_{ij} be the response variable and x_{ij} a vector of length p of explanatory variables observed at the j^{th} time-point, $j = 1, \ldots, n_i$ on the i^{th} subject, $i = 1, \ldots, n$. The set of repeated outcomes for subject i are collected into an n_i-vector and can be written as $\mathbf{y}_i' = [\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{i\mathbf{n}_i}]'$, with mean $E(\mathbf{y}_i) = \mu_i$. The set of explanatory variables or covariates are grouped into an $n_i \times p$ matrix $\mathbf{x}_i = [\mathbf{x}_{i1}', \mathbf{x}_{i2}', \ldots, \mathbf{x}_{i\mathbf{n}_i}']'$.

2.1.2 Marginal models

Marginal or population averaged models are widely used in the biomedical sciences and are very flexible in that they require no distributional assumption for the vector of responses, only a model for the mean response. They are marginal in that the mean response depends only on the covariates of interest and not on any random effects. The frequentist estimation methods, Generalized Estimating equations (GEE) by Liang and Zeger (1986)[60] were specifically developed for parameter estimation in these models. In case of independent repeated measurements, the classical score equations for the estimation of β are given as :

$$S_k(\beta) = \sum_{i=1}^{k} \frac{d\mu_i}{d\beta_k} \nu_i^{-1} (\mathbf{y}_i - \mu_i) = 0 \quad k = 1, \ldots, p, \tag{2.1.21}$$

where β is a vector of unknown regression coefficients and ν_i is a diagonal matrix with $\nu_{ij} = Var(y_{ij})$ on the main diagonal. For longitudinal data, Liang and Zeger (1986)[60] extended the score equations (2.1.21) to multivariate setting:

$$\sum_{i=1}^{k} \frac{d\mu_i}{d\beta_k} (A_i^{1/2} R_i(\alpha) A_i^{1/2} / \phi)(\mathbf{y}_i - \mu_i) = 0, \tag{2.1.22}$$

where $R_i(\alpha)$ is an $n_i \times n_i$ fully specified "working correlation" matrix which may depend on a vector of unknown parameters α, which is assumed to be the same for all subjects,

$\nu_i = (A_i^{1/2} R_i(\alpha) A_i^{1/2}/\phi)$ is the "working covariance" matrix for y_i , A_i is an $n_i \times n_i$ diagonal matrix consisting of a function of the mean $g(\mu_{ij})$ along the main diagonal, and ϕ is a scale parameter. The term "working" is used to imply that the model assumes that the form of the covariance may not be correctly specified (Zeger and Liang, 1986)[103]. The GEE estimator of β is the solution to

$$\sum_{i=1}^{k} \frac{d\mu_i}{d\beta_k} (A_i^{1/2} R_i(\hat{\alpha}) A_i^{1/2}/\hat{\phi})(\mathbf{y}_i - \mu_i) = 0, \tag{2.1.23}$$

where $\hat{\alpha}$ and $\hat{\phi}$ are consistent estimates of α and ϕ, respectively. The solution to Equation 2.1.41 is obtained via iteratively weighted least squares method (for more details see McCullagh and Nelder, 1983[68]).

The strength of the GEE method is that it is robust to the choice of the "working correlation" structure and only requires that the mean response be correctly specified. This robustness property holds if data are complete or missing completely at random (MCAR) (Rubin, 1976)[81]. A major limitation is that the GEE approach is not a likelihood-based method and hence it is difficult to determine the goodness of fit, compare models, and to draw statistical conclusions on the model parameters.

2.1.3 Generalized linear mixed effects model (GLMM)

In Section 2.1.2 we introduced marginal models where the GEE accounts for the within individual correlation. Alternatively, random effects are incorporated into the mean model to account for the association between repeated measurements within an individual. The difference between random effects and marginal models is that the latter is population-specific while random effects are subject-specific. Marginally, conditional on the random effects, the repeated measurements within an individual are assumed to be independent observations from a distribution belonging to the exponential family. This is known as the "conditional independence assumption" (Laird and Ware, 1982)[57]. In general, the model with random effects (i.e., Generalized linear mixed effects model) has three parts:

1. The conditional distribution of the j^{th} response y_{ij}, given a vector of random effects b_i belongs to the exponential family, $Var(y_{ij}|\mathbf{b}_i) = Var\{E[y_{ij}|\mathbf{b}_i]\}\phi$ is a function of the conditional mean, where ϕ is a scale parameter, and given $\mathbf{b}_i$ the $y'_{ij}s$ are independent.

2. The conditional mean depends on fixed covariates and random effects through the linear predictor: $\eta_{ij} = x'_{ij}\beta + z'_{ij}\mathbf{b}_i$ with $g\{E(y_{ij}|b_i)\} = \eta_{ij} = x'_{ij}\beta + z'_{ij}\mathbf{b}_i$ for a pre-specified link function, $g(.)$.
3. The random effects are assumed to have a multivariate distribution with mean vector 0 and $q \times q$ covariance matrix $\mathbf{G}$ and independent of the covariates.

2.1.4 Estimation in Generalized Linear Mixed Models

In contrast with marginal models the joint distribution of the vector of responses and the vector of random effects is fully specified, hence inference and estimation are based on the likelihood function. Based on the conditional independence assumption, the joint distribution of the response $\mathbf{y}_i$ and the random effects, $\mathbf{b}_i$ can be written as

$$f(\mathbf{y}_i|\mathbf{b}_i) = \prod_{i=1}^{n} f(\mathbf{y}_i|\mathbf{b}_i)f(\mathbf{b}_i). \tag{2.1.41}$$

The frequentist maximum likelihood estimates are then obtained by integrating out or averaging over the unobserved random effects $\mathbf{b}_i$ from (2.1.41) to obtain the marginal likelihood which does not depend on $\mathbf{b}_i$ (Equation 2.1.42):

$$L(\beta, \phi, \mathbf{G}) = \prod_{i=1}^{n} \int f(\mathbf{y}_i|\mathbf{b}_i)f(\mathbf{b}_i)\mathbf{db}_i, \tag{2.1.42}$$

The maximum likelihood (ML) estimates are therefore the estimates of $\mathbf{G}$, β and ϕ that maximize Equation 2.1.42. In most cases, the likelihood in (2.1.42) can not be evaluated analytically, therefore, numerical approximations are required. Numerical iterative methods such as the Newton-Raphson (NR) method, the Fisher scoring method, and the EM and modified EM algorithms are employed to obtain the ML estimates. The NR method is faster but very sensitive to starting values hence unstable compared to EM algorithms. All in all, these methods are analytically and computationally very intensive. In the next section, we introduce the Bayesian Markov Chain Monte Carlo (MCMC) methods of estimation that are computationally easier.

2.1.5 Bayesian Methods

In the Bayesian approach, the unknown parameters are treated as random quantities, and therefore assigned a prior probability distribution that describes the uncertainty about the parameter values. Bayesian inferences are then based on the posterior distribution, the conditional probability distribution of the parameters of interest θ, given the observed data $\mathbf{y}$. The posterior distribution is given by

$$\pi(\theta|\mathbf{y}) = \frac{L(\theta|\mathbf{y})\pi(\theta)}{\int_0^\infty \pi(\theta)L(\theta|\mathbf{y})d\theta}, \tag{2.1.51}$$

where $\pi(\theta)$ is the prior distribution of θ and $L(\theta|\mathbf{y})$ is the likelihood function. By ignoring the normalizing constant (denominator) in 2.1.51, the posterior distribution is

$$\pi(\theta|\mathbf{y}) \propto L(\theta|\mathbf{y})\pi(\theta),$$

which is a function of the likelihood and prior information/distribution. Originally, the EM and modified algorithms were employed in Bayesian estimations but they can only estimate the posterior mode and because Bayesian functions are more complex with inclusion of prior distribution, these estimation methods made Bayesian methods very unattractive. However, the introduction of MCMC methods has made Bayesian methods very appealing.

The MCMC methodology provides enormous scope for realistic statistical modeling of complex models because they do not use direct integration methods such as Gaussian quadrature and Laplace approximation (Tierney and Kadane,1986[91]; Shun and McCullagh,1995[84]), which are computationally very intensive. The MCMC methods simulate direct draws from some complex distribution of interest, where previous sample values are used to randomly generate the next sample value, generating a Markov chain (as the transition probabilities between samples values are only a function of the most recent sample value). From the Markov chain theory, this chain in the long run converges to a stationary or equilibrium distribution which is precisely the posterior distribution. There are many ways of constructing these chains, but most of them, including the Gibbs sampler (Geman and Geman, 1984[29]; Gelfand and Smith,1990[23]), are special cases of the general framework of Metropolis et al. (1953)[70] and Hastings (1970)[35]. Many MCMC algorithms are hybrids or generalizations of the simplest methods: the Gibbs sampler and the Metropolis-Hasting algorithm.

2.1.5.1 The Gibbs Sampler Many statistical applications of MCMC have used the Gibbs sampler, which is easy to implement. Gelfand and Smith (1990)[23] gave an overview, and suggested the approach for Bayesian computation. The Gibbs sampling algorithm is described as follows: Let $\theta = (\theta_1, \ldots, \theta_q)$ be the parameters in the model of interest $p(\theta)$. Given an arbitrary set of initial values $\theta^{\mathbf{0}} = \left(\theta_1^0, \ldots, \theta_q^0\right)$, we draw $\theta_1^{(1)}$ from conditional distribution $P\left(\theta_1|\theta_2^{(0)}, \ldots, \theta_q^{(0)}\right)$, then $\theta_2^{(1)}$ from conditional distribution $P\left(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_q^{(0)}\right)$ and so on up to $\theta_q^{(1)}$ from $P\left(\theta_q|\theta_1^{(1)}, \ldots, \theta_{q-1}^{(1)}\right)$ to complete one iteration of the scheme. This scheme is a Markov chain, with equilibrium distribution $p(\theta)$. After t such iterations, we would arrive at the t^{th} value $\theta^{(t)} = \left(\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_q^{(t)}\right)$. Thus, for t large enough, $\theta^{(\mathbf{t})}$ can be viewed as a simulated observation from $p(\theta)$. In essence, sampling long enough from this scheme will result in sampling from the posterior distribution itself. So, after discarding an initial set of samples (called burn-in) the remaining samples constitute the posterior sample from which all inferences can be drawn.

2.1.5.2 The Metropolis-Hasting algorithm Although, the Gibbs sampler works well for complex hierarchical models, it is limited to sampling from the full conditionals. When the conditional distribution is not in closed form, a more general and powerful algorithm is the Metropolis-Hastings (MH) algorithm formulated by Hastings (1970)[35], which is a generalization of the method first proposed by Metropolis et al. (1953)[70]. This algorithm also constructs a Markov Chain, but does not necessarily care about full conditionals. Let $p(\theta)$ be the distribution of interest but suppose it is hard to sample from. Suppose at time t, θ_{t+1} is chosen by first sampling a candidate point v from a proposal distribution $q(\cdot|\theta_t)$, which is easy to sample from. The candidate v is accepted with probability

$$\alpha(\theta, v) = \min\left(1, \frac{p(v)q(\theta|v)}{p(\theta)q(v|\theta)}\right)$$

If the candidate point is accepted, the next state becomes $\theta_{t+1} = v$. If it is rejected, the chain does not move. The proposal distribution can be any kind of continuous probability density, however, empirical evidence suggests that the more it incorporates the structure of the problem the faster is the convergence. Several possible proposals are discussed and

compared by Tierney (1994)[90]. Again, for large enough t and sufficient burn-in period, the stationary distribution of the chain will be $p(\theta)$.

2.1.5.3 Other sampling schemes In addition to these algorithms in pure form, a number of hybrid schemes are available. For instance, one such scheme consists of combining the Metropolis steps within the Gibbs sampler (Muller, 1991)[71] when the full conditionals are formed but difficult to sample from. Here, a Metropolis step can be used to draw samples from $p(\theta)$ by forming proposal densities $q(\theta|v)$ and acceptance probabilities $\alpha(\theta, v)$ based on the proposal and conditional posterior distributions of each parameter of interest. In addition, several algorithms have been developed to improve the convergence of the MCMC iterations. These include blocking of components (Liu et al.,1994; Cowles,1996; Gamerman,1997)[63, 13, 20]. Gamerman [20], notes that for Gibbs sampler in pure forms blocking correlated quantities generally speeds up convergence but the same is not necessarily true for the Metropolis-Hastings algorithms. Accordingly, he developed an MCMC approach that uses Metropolis-Hasting algorithm to sample from the posterior distributions of blocks of correlated parameters (based on their conditional independence structure), which we adopt in our study. This approach incorporates the structure of the model, that is, the form of the likelihood and prior, leading to an algorithm requiring a single iterative procedure. In addition, prior distributions for the regression coefficients and random effects distribution are not restricted to normality with non-informative cases providing a link with frequentist approaches. The resulting inference is based on samples from the posterior distribution of all model parameters and standard assessments such as parameter significance and residual analysis can be made without having to resort to asymptotic normality results.

2.2 METHODS FOR SURVIVAL ANALYSIS

In survival analysis, the exact survival times of the subjects are not known in most cases. These are called censored observations or censored times which require special statistical techniques to handle them. There are different forms of censoring but the most common form

is right-censoring where an individual may withdraw from the study, be lost to follow-up, or for economic or practical reasons it may require that the study ends before the outcome has occurred. This is the censoring we are concerned about in this study.

2.2.1 Notation

Let $T_i = \min(T_i^*, C_i)$ be the failure time and $\delta_i = I(T_i^* \leq C_i)$ an event indicator which indicates whether the observed failure time is a true failure time, T_i^*, or a censoring time C_i for the i^{th} individual. In addition, let $\mathbf{x}_i' = (x_{i1}, x_{i2}, \cdots, x_{ip})$ be a vector of baseline covariates associated with the i^{th} individual.

2.2.2 Likelihood for Right Censored Data

Given that $\mathbf{x}_i'$ and the pairs of random variables (T_i, δ_i), $i = 1, \cdots, n$, are independent, the likelihood for (T_i, δ_i) conditional on $\mathbf{x}_i'$ can be expressed as:

$$L(\theta) \propto \prod_{i=1}^{n} f(t_i, \theta, x_i)^{\delta_i} S(t_i, \theta, x_i)^{1-\delta_i}, \tag{2.2.21}$$

where θ are the parameters to be estimated, $f(t_i, \theta, x_i)$ is the probability density function for a failure time, and $S(t_i, \theta, x_i)$ is the survival distribution for censored time. Because we can write $f(t_i, \theta, x_i) = h(t_i, \theta, x_i)S(t_i, \theta, x_i)$ and $S(t_i, \theta, x_i) = \exp\{-H(t_i, \theta, x_i)\}$ the likelihood (2.2.21) can written as

$$L(\theta) \propto \prod_{i=1}^{n} h(t_i, \theta, x_i)^{\delta_i} \exp\{-H(t_i, \theta, x_i)\}, \tag{2.2.22}$$

where $h(t_i, \theta, x_i)$ and $H(t_i, \theta, x_i)$ are the hazard and cumulative hazard functions for the i^{th} individual, respectively. Simplifying (2.2.22) further gives the likelihood as

$$L(\theta) \propto \prod_{i=1}^{n} h(t_i, \theta, x_i)^{\delta_i} \exp\left(- \int_0^{t_i} h(s, \theta, x_i) ds \right). \tag{2.2.23}$$

2.2.3 Cox proportional hazard model

The Cox proportional hazards model developed by Cox in 1972 is the most widely used analytic tool for survival analysis [14]. This model has a strong assumption called the "proportional hazard assumption" which is often violated. The class of accelerated failure time models is an alternative when the proportionality assumption does not hold. Under the proportional hazard assumption, the hazard function for the i^{th} individual is

$$h_i(t) = h_0(t) \exp\left(\mathbf{x}_i'\beta\right), \tag{2.2.31}$$

where β is the vector of regression coefficients and $h_0(t)$ is the baseline hazard, which can be fully parametric, or left unspecified. Maximum likelihood estimates of β in Equation 2.2.31 are obtained from the Cox's partial likelihood function, $L(\beta)$(2.2.32), assuming independence of failure times.

$$L(\beta) = \prod_{j=1}^{D} \frac{\exp\left(\mathbf{x}_j'\beta\right)}{\sum_{l \in R_j} \exp\left(\mathbf{x}_l'\beta\right)}, \tag{2.2.32}$$

where D is the number distinct event times, R_j is called the risk set (individuals who are at risk at time j), and $\mathbf{x}_j$ denotes the $p \times 1$ vector of covariates for the individual who has the event at time j. The estimator $\hat{\beta}$ has been shown to be a consistent estimator for β and is asymptotically normal as the marginal models are correctly specified (Lin, 1994 [61]). The partial likelihood 2.2.32 can be extended to include time-dependent covariates and rewritten as

$$L(\beta) = \prod_{j=1}^{D} \frac{\exp\left(\mathbf{x}_j'\left(t_j\right)\beta\right)}{\sum_{l \in R_j} \exp\left(\mathbf{x}_l'\left(t_l\right)\beta\right)}, \tag{2.2.33}$$

which can be maximized by iterative techniques, such as the Newton Raphson algorithm but the Bayesian MCMC methods work best. However, this would imply complete knowledge of the covariate at each unique event time, which is problematic when one would wish to include a covariate measured longitudinally over time and examine its effect on an outcome. The solution to this problem is one of our main goals for this study (see Chapter 4).

2.2.4 Frailty Models

In the traditional survival analysis models, observations are assumed to be heterogeneous and the population they come from is assumed to be homogeneous with respect to failure. In a situation where this is questionable i.e., where some members are more failure-prone (frail) than others due to unobserved heterogeneity, these models can lead to under- or over-estimated standard errors of estimates.

Frailties are unobserved effects or unmeasurable genetic factors of an individual (individual-specific or unshared) or shared by all members of the cluster or group (group-specific or shared). Hougaard (1995)[47] pointed out that the impact of unmeasured covariates can lead to transformation of the hazard function and the coefficients of the measured covariates. There is also strong evidence that the hazard functions often converge in contradiction to the proportional hazards assumption of the traditional Cox model. Thus, an introduction of a frailty parameter in the traditional model to handle dependence between survival times is much realistic (Keiding et al., 1997 [55] and Vaupel et al., 1979 [95]). The hazard for the j^{th} subject in i^{th} cluster or subgroup, given the frailty $\mathbf{w}_i = (w_{i1}, \cdots, w_{in_i})'$, is defined as

$$h_{ij}(t) = h_0(t) \exp\left(\sigma \mathbf{w}_i + x'_{ij}\beta\right), i = 1, \cdots, G, j = 1, \cdots, n_i, \tag{2.2.41}$$

where $h_0(t)$ is an arbitrary baseline hazard rate, σ is a vector of parameters associated with the frailties, x_{ij} is a vector of covariates, and β is the vector of coefficients. The frailties are assumed to be from some distribution with mean zero and variance 1. The Gamma distribution is the most common (where $w_{ij} = 1, i.e., \exp(\sigma) \sim gamma(\zeta, \zeta)$) due to its mathematical convenience but other distributions like Uniform, inverse Gaussian and Log-normal can be considered. We note that when $\sigma = 0$ 2.2.41 reduces to the proportional hazards model 2.2.31.

Alternatively, 2.2.41 can be written as

$$h_{ij}(t) = h_0(t)\mathbf{w}_i \exp\left(x'_{ij}\beta\right), i = 1, \cdots, G, j = 1, \cdots, n_i. \tag{2.2.42}$$

From 2.2.42, it is clearly seen that when $w_i > 1$ individuals within a given group tend to fail faster than those with $w_i < 1$. Because the $w'_i s$ are unobserved, estimation methods in GLMM (Sections 2.1.4 and 2.1.5) are employed to estimate the parameters in frailty models.

2.2.5 Common Parametric Survival Distributions

While in a Cox model (nonparametric), $h_0(t)$ is left unparameterized, in the parametric approach a functional form for $h_0(t)$ is specified. The only requirement is that the survival distribution be bounded between 0 ($S(\infty) = 0$) and 1 ($S(0) = 1$). For instance, if

$$h_0(t) = \exp(\alpha),$$

for some α then we have an exponential distribution. Here the baseline hazard is assumed constant over time. If we assume

$$h_0(t) = \lambda t^{\lambda} \exp(\alpha),$$

then we have the Weibull model (see Klein and Moeschberger[56] for a more distributions). Parameter estimation in parametric models is much easier than in semi-parametric ones. By assigning prior distributions to parameters, the Bayesian methods in Section 2.1.5 are also employed to estimate parameters by sampling from their respective full conditionals.

2.3 JOINT MODELING

In Sections 2.1 and 2.2, we introduced the general methods for modeling longitudinal and survival data in a univariate setting, respectively. However, in biomedical studies where more than one biomarker of the disease is measured over time on each individual as well as a set of random times at which events of interest occur (time-to-event), joint modeling of either longitudinal biomarkers together or longitudinal biomarker(s) with time-to-event has been employed to improve the efficiency of the parameter estimates as they tend to account for the variability that exists between or among the different processes.

2.3.1 Joint modeling of multiple longitudinal outcomes of mixed types

Joint modeling of longitudinal outcomes of mixed types has been shown to lead to efficient estimates. For instance, Guerguieva and Sanacora (2006)[33] who studied joint models of repeatedly observed continuous and ordinal measures of the same underlying disease severity noted that when the trajectories over time may be related but measure distinct underlying

trends, separate analyses may be more appropriate. However, accounting for the associations between various outcomes through joint modeling leads to more efficient estimates compared to separate analyses. Similar results were observed by McCulloch (2008)[69] who showed that joint modeling leads to efficiency gain while separate analyses of mixed types longitudinal outcomes can be inefficient. Also, when the data on one of the outcomes are more complete than another then joint modeling can accommodate data that are missing at random instead of the stronger assumption of missing completely at random (MCAR).

Several approaches have been proposed to jointly model multiple outcomes of mixed types but there are two key approaches. The first approach uses the product of marginal and conditional distributions. Letting $\mathbf{y}_1$ and $\mathbf{y}_2$ represent the continuous and discrete outcome, respectively. Using the product of marginal and conditional distributions, the joint distribution of $\mathbf{y}_1$ and $\mathbf{y}_2$ can be written as

$$f(\mathbf{y}_1, \mathbf{y}_2) = f(\mathbf{y}_1)f(\mathbf{y}_2|\mathbf{y}_1) = f(\mathbf{y}_2)f(\mathbf{y}_1|\mathbf{y}_2).$$

In this formulation, it is possible to have different results depending on whether the conditioning variable is discrete or continuous. A major drawback of this method is that it is hard to get easy expressions for the association between both continuous and discrete outcomes, and it does not directly lead to marginal inference. In addition, in case of more than two outcomes, there will be many more possible factorizations instead of only the two associated with two outcomes. In this regard, the conditional model may not be the best choice in high dimensional longitudinal data.

The second approach is that of random effects. In this method, different outcomes are joined by imposing a common distribution for their random effects. These can be shared (i.e., $\mathbf{b}_{i2} = \gamma\mathbf{b}_{i1}$) where $\mathbf{b}_{i2}$ is assumed proportional to $\mathbf{b}_{i1}$ with a restrictive correlation structure between the two outcomes or correlated where $\mathbf{b}_{i1}$ and $\mathbf{b}_{i2}$ are assumed to follow a multivariate distribution with a nonrestrictive covariance structure, which can be unstructured, Toeplitz, exchangeable, etc. In general, the correlated random effects model allows for flexible correlation structure but it has a disadvantage of high-dimensional vector of random effects as the number of outcome variables gets large.

Catalano and Ryan (1992)[12] used the concept of a latent variable to derive the joint distribution of a continuous and binary outcome for clustered data. The joint distribution

was a product of a standard random effects model for the continuous variable and a correlated Probit model for the discrete (binary) variable. Thus, they considered a linear regression mode1 (2.3.11) on the latent variable y_{ij}:

$$y_{ij} = \beta_0 + \beta_1 d_i + \epsilon_{ij}, \epsilon_{ij} \sim N(0, \sigma^2), \tag{2.3.11}$$

where d_i is a vector of covariates for the i^{th} individual. The observed binary variable y^*_{ij} and the latent variable y_{ij} are such that:

$$y^*_{ij} = \begin{cases} 1 & \text{if} \quad y_{ij} > 0 \\ 0 & \text{if} \quad y_{ij} \leq 0, \end{cases} \tag{2.3.12}$$

Then from the normal model (2.3.11), it follows that y^*_{ij} follows a probit model

$$P(y^*_{ij} = 1|d_i) = \Phi\left(\frac{\beta_0 + \beta_1 d_i}{\sigma}\right) \tag{2.3.13}$$

They considered the bivariate model (2.3.14) for the observed continuous variable y_{1ij} and unobserved latent variable y_{2ij}:

$$\begin{aligned} y_{1ij} &= \alpha_0 + \alpha_1 d_i + \epsilon_{1ij} \\ y_{2ij} &= \beta_0 + \beta_1 d_i + \epsilon_{2ij} \end{aligned} \tag{2.3.14}$$

where,

$$\epsilon_{ij} = \begin{pmatrix} \epsilon_{1ij} \\ \epsilon_{2ij} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right).$$

And a probit model (2.3.15) for the unobserved latent variable y^*_2

$$P(y^*_{2ij} = 1|d_i) = \Phi\left(\frac{\beta_0 + \beta_1 d_i}{\sigma}\right) \tag{2.3.15}$$

The joint distribution of the observed continuous y_{1ij} and observed binary y^*_{2ij} was formed as a product of marginal and conditional distributions,

$$f_{y_{1ij}, y^*_{2ij}}(y_1, y^*_2) = f_{y_{1ij}}(y_1) f_{y^*_{2ij}}(y^*_2|y_1),$$

The two processes were linked through the residual errors. Parameter estimation was done in two steps using Generalized Estimating equations (Liang and Zeger, 1986; and Zeger and Liang, 1986)[60, 103]. In the first step they estimated the parameters of the marginal

distribution and in the second step they estimated the parameters of the correlated probit model. Catalano (1997)[11] extended this approach to jointly model continuous and ordinal outcomes, where the ordinal response was modeled using a correlated Probit model.

Fitzmaurice and Laird (1995)[19] also used the same product method to jointly model binary and continuous outcomes with clustering but conditioned on the discrete outcome. They assumed a Bernoulli distribution for the binary response $\mathbf{Y}_i$ and a Gaussian distribution for the continuous response $\mathbf{X}_i$. The joint model was a product of marginal distribution of binary and conditional distribution of the continuous response, i.e.,

$$f_{\mathbf{X}_i,\mathbf{Y}_i}(\mathbf{x}_i,\mathbf{y}_i) = f_{\mathbf{Y}_i}(\mathbf{y}_i) f_{\mathbf{X}_i|\mathbf{Y}_i}(\mathbf{x}_i|\mathbf{y}_i).$$

The marginal distribution of the binary response was related to covariates using a logit link function, whereas the conditional distribution of the continuous response was related to covariates using a linear link function with a conditional mean that depends on the binary response. Thus, this dependence induced association or correlation between the two responses. The parameters were estimated using GEE (Liang and Zeger, 1986)[60]. Here, the regression parameters have a marginal interpretation because the models do not include random (subject–specific) effects.

Furthermore, Gueorguieva and Agresti (2001)[32] proposed a correlated probit model for joint modeling of clustered binary and continuous responses by employing a linear mixed effects model for the continuous observed and unobserved latent variable as suggested by Catalano and Ryan [12] but instead modeled the two outcomes through shared random effects. In addition, they used Monte Carlo Expectation-Conditional maximization (ECM) algorithm which is a modification of EM algorithm for parameter estimation. While Gueorguieva and Sanocora (2006)[33] extended the method of Gueorguieva and Agresti to model the ordinal and continuous responses. These models were rewritten so as to be fit using maximum likelihood with standard software using procedures like NLMIXED in SAS or GLLAMM in STATA.

In summary, other than Fitzmaurice and Laird who modeled the binary outcome as a Bernoulli linked to its covariates by a logit link and the continuous outcome as a Gaussian with a linear link function, the other authors employed the latent variable approach where the

binary/ordinal observed response is modeled with a Probit/correlated Probit model and its underlying latent continuous variable and the observed continuous response with linear mixed effects model. This enabled the two responses to be modeled jointly as a bivariate normal, linked through either correlated residual errors or shared random effects. However, much as the Probit and Logistic models are both used in modeling binary or ordinal data and both have symmetric S-shape cumulative distributions, the logistic places more probability in the tails than does the Probit and hence more stable when dealing with outlying data. In addition, the logistic link function is more popular in the biomedical field and the interpretation is easier than the probit link function.

2.3.2 Joint modeling of longitudinal outcomes and time-to-event

In Section 2.3.1, we gave a review of literature in relation to joint modeling of longitudinal outcomes of mixed types, however, the concept of joint modeling has been widely used in simultaneous modeling of longitudinal outcomes and time-to-event data. The goal of joint modeling in this context include (1) Modeling the distribution of the time to a terminal event conditional on a longitudinal measurement sequence. This kind of modeling was first used in AIDS research, where CD4 cell count or estimated viral load was used to predict the time to onset of clinical AIDs (e.g. Tsiatis, DeGruttola, and Wulfsohn, 1995)[94]. The primary interest here was in survival time but the longitudinal measurements were used as time-varying covariate. (2) Adjusting inference about a longitudinal measurement sequence to allow for informative dropout. That is, the absence of longitudinal observations beyond the event time is a form of non-ignorable missingness, so that a joint distribution is specified for the longitudinal and missingness (survival) processes (e.g. Hu and Sale, 2003)[48]. (3) Modeling the joint evaluation of a measurement and an event-time process. For instance, in the Diabetes study, our interest is to model jointly the time to when blood glucose reaches normal range and the longitudinal evolution of cheaply measured markers that are associated with diabetes progression such as systolic and diastolic blood pressure and body mass index (BMI).

Joint models that combine the longitudinal and time-to-event processes have been widely studied by different authors. Hogan and Laird (1997a)[44], Tsiatis and Davidian (2004)[93], and Ibrahim, Chen, and Sinha (2001, Chapter 7)[51] give a detailed discussion of joint modeling. Pawitan and Self (1993)[73], DeGruttola and Tu (1994)[16], Tsiatis, DeGruttola, and Wulfsohn (1995)[94], Faucett and Thomas (1996)[18], Lavalley and De Gruttola (1996)[58], Wulfsohn and Tsiatis (1997)[100], Henderson, Diggle, and Dobson (2000)[42], Xu and Zeger (2001a)[101], Tsiatis and Davidian (2001)[92], Wang and Taylor (2001)[97], Guo and Carlin (2004)[34], Brown and Ibrahim (2003) [6, 7], Ibrahim, Chu, and Chen(2010)[52], Wang, Shen, and Boye (2012)[96], Huang, Hu, and Dagne (2014)[49] all have worked on one longitudinal outcome and time-to-event processes. Rizopoulos and Ghosh (2010)[79] and Hatfield, Boye, and Carlin (2011)[36] extended the longitudinal outcome to the multivariate case.

In most of the literature cited above, the joint modeling of the survival and longitudinal components is usually done by assuming that the longitudinal model follows a linear mixed effects model and that the survival model depends on the random effects from this process. Inference is then based on the integrated conditional joint likelihood where the random effects usually follow a multivariate normal distribution. Initially the two processes are assumed to be conditionally independent given the data and parameters of interest and only correlated through the induced random effects or the underlying latent process. For instance, Henderson, Diggle, and Dobson (2000)[42] linked 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. Given there are n subjects with longitudinal measurements $\{y_{ij} : j = 1, \cdots, n_i\}$ at times $\{t_{ij} : j = 1, \cdots, n_i\}$. 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 model for longitudinal data:

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

where ϵ_{ij} is a measurement error term assumed to be mean-zero normally distributed with

$var(\epsilon_{ij}) = \sigma_\epsilon^2$ and $\mu_i(t_{ij})$ is the mean response assumed by a linear model

$$\mu_i(t_{ij}) = \mathbf{x}_{1i}(t)^T \beta_1$$

in which the vectors $\mathbf{x}_{1i}(t)$ and β_1 represent possibly time-varying covariates and their corresponding regression coefficients, respectively. For the survival model, they considered a semi-parametric multiplicative model:

$$\lambda_i(t) = H_i(t)\alpha_0(t)\exp\{\mathbf{x}_{2i}(t)^T\beta_2 + W_{2i}(t)\},$$

with the form of $\alpha_0(t)$ left unspecified. In contrast with Tsiatis, DeGruttola, and Wulfson (1995)[94], Faucett and Thomas (1996)[18], and Wulfsohn and Tsiatis (1997)[100] worked with similar models but assumed $W_{1i}(t) = U_{1i} + U_{2i}t$ and $W_{2i}(t)$ proportional to $W_{1i}(t)$, Henderson et al.[42] defined $W_{1i}(t)$ and $W_{2i}(t)$ respectively as

$$W_{1i}(t) = Z_{1i}(t)^T U_{1i} + V_{1i}(t) \text{ and } W_{2i}(t) = \gamma_1 U_{1i} + \gamma_2 U_{2i} t + \gamma_3(U_{1i} + U_{2i}t) + U_{3i},$$

where $Z_{1i}(t)$ is a vector of covariates, U_{1i} is a corresponding vector of random effects that follow a multivariate normal distribution with mean zero and variance-covariance matrix Σ_1, $V_{1i}(t)$ is a stationary Gaussian process with mean zero, variance σ_{v1}^2 and correlation function $r_1(s) = cov\{V_{1i}(t), V_{1i}(t-s)\}/\sigma_{v1}^2$, and the frailty term $U_{3i} \sim N(0, \sigma_3^2)$ is independent of the (U_{1i}, U_{2i}). The parameters γ_1, γ_2, and γ_3 measure the association between the longitudinal and survival models induced through the random intercepts, slopes, and the current value of W_{1i} at time t. They estimated the parameters using EM algorithm proposed by Wulfsohn and Tsiatis [100] in 1997 and noted the identifiability problems which can arise when $W_{2i}(t)$ is allowed to be time-varying in conjunction with a non-parametric specification of the baseline intensity $\lambda_0(t)$.

Other authors have used similar general methods but with different distributions for the survival and/or longitudinal process or estimation procedures. For instance, Faucett and Thomas (1996)[18] used a linear mixed model and Bayesian methods for the parameter estimation. Xu and Zeger (2001a)[101] used a latent variable approach and implemented a Markov Chain Monte Carlo Algorithm for the estimation, and De Gruttola and Tu (1994)[16] implemented a fully parametric joint model by assuming that the survival and longitudinal

processes follow a multivariate normal distribution by transformation of the survival times to follow a normal distribution and estimation was via the EM algorithm. Tsiatis et al.(1995)[94] used a two stage approach (Partial likelihood approach), where the true value of the covariate at the event time was estimated by a linear mixed effects model in the first stage and then substituted into the hazards model in the second stage. Guo and Carlin (2004)[34] used the flexible joint model proposed by Henderson et al. (2000)[42] but used Bayesian approach via MCMC for parameter estimation. While Rizopoulos and Ghosh (2010)[79] proposed a Bayesian semiparameteric multivariate joint model that relates multiple longitudinal outcomes (Continuous and binary) and time-to-event. They used a spline-based approach to model the subject specific longitudinal evolutions and the baseline risk function in the Cox model for time-to-event outcome was assumed piece-wise constant.

Most of the joint models in the literature above (Section 2.3.2) are for one continuous longitudinal outcome and time-to-event. To the best of our knowledge, no one has worked on joint modeling of continuous, ordinal, and time-to-event outcomes.

2.4 OUR CONTRIBUTION

In Section 2.3.1, we noted that all of the previously proposed joint models for continuous and ordinal outcomes employed the Probit link function to model the ordinal outcome because of its flexibility (underlying normal framework) to reduce the computational burden. However, there are several disadvantages of using a Probit link as compared to Logit link functions. In addition, EM or modified EM algorithms were employed for parameter estimation which are computationally very intensive. Furthermore, no work has been done on joint modeling of continuous, ordinal, and time-to-event outcomes (Section 2.3.2). Thus, we propose a joint model for unbalanced repeatedly measured continuous and ordinal outcomes and time-to-event data. In the first part we develop a model for continuous and ordinal outcomes (Aim 1). Here, we employ a Cumulative Logit link function for the ordinal outcome and identity link function for the continuous outcome and model the two outcomes jointly as a multivariate generalized linear mixed effects model linked through correlated and/or shared random effects.

Secondly, we extend the model in Aim 1 to include time-to-event (Aim 2). Time-to-event is modeled parametrically using a Weibull distribution with an unshared frailty model. The Bayesian approach (i.e. MCMC) is employed for parameter estimation in both parts because it has the capacity to handle complex models with ease. The Aim 1 of this work is described in Chapter 3, while Aim 2 work, is discussed in Chapter 4.

3.0 BAYESIAN HIERARCHICAL JOINT MODELING OF REPEATEDLY MEASURED CONTINUOUS AND ORDINAL MARKERS OF DISEASE SEVERITY

3.1 INTRODUCTION

In many clinical studies, more than one biomarker of disease severity is obtained and some may be easier and cheaper to obtain than others. Although these biomarkers are often meant to measure the same disease severity, they may differ due to the instruments/reagents used as well as the scale of measurements. They could show different patterns for treatment because clinicians prescribe medications based on the severity of disease. Moreover, if these markers are modeled separately to determine the factors that are associated with disease progression over time or to predict the event of interest (i.e. time to remission) given different treatments, they may yield different or misleading results. Modeling these markers jointly while accounting for correlation between the two markers is likely to provide more valid results.

The motivation for our study is based on data collected retrospectively from medical registries of diabetic patients in three Ugandan hospitals. These patients were recruited in the diabetic clinics between January 1992 and December 2004. Diabetes which is a progressive illness occurs when the pancreas does not produce enough insulin or when the body does not respond properly to varying levels of insulin. This results in increased concentrations of glucose in the blood, which in turn damages many of the body's systems, in particular the blood vessels and nerves. Thus, the amount of glucose in the blood determines the state of the disease at a point in time. In addition, the amount of glucose in the urine is used to detect if the individual's blood glucose level is above the renal threshold of 180 mg/dl. The amount

of glucose in the urine is interpreted using the + symbolic method or the actual amount in the urine(mg/dl) depending on the manufacturer of the urine glucose reagent strips. That is, Nil (no urine sugar),+($\approx$ 100 mg/dl), + + ($\approx$ 250 mg/dl), + + +($\approx$ 500 mg/dl), and ++++($\approx$ 1000 mg/dl), respectively. Of the two biomarkers, blood glucose which is measured by fasting plasma glucose test (FPG) or oral glucose tolerance test (OGTT) is more accurate and hence recommended by both the World Health Organization (WHO) and the American Diabetes Association (ADA). The urine glucose tests to detect Glycosuria/Glucosuria (glucose in urine) are used as an alternative to blood glucose tests especially in developing countries because they are fast, do not require many reagents, easy to carry out and generally economical (Carter and Lema, 2003)[10]. However, it is important to note that the urine glucose test for diabetes may be contaminated by drugs and individual variations in renal threshold for glucose. Thus, making a clinical decision based on a urine test alone may be invalid or misleading.

During the hospital visits (follow-up period), the patients in the Ugandan study were periodically tested for the amount of glucose in the blood or urine or both to determine the severity of the disease so as to prescribe appropriate medication. The data are highly unbalanced because this was an observational study where patients reported for checkup at irregular intervals with the number of hospital visits varying from patient to patient. Clinically, the two markers measure the same diabetes severity although blood glucose is continuous and urine glucose is ordinal with five levels. Thus, for analysis purposes, joint modeling of the two markers simultaneously will produce optimal results because the two markers are highly correlated. In addition, an appropriate way of handling the unbalanced data will result in efficient estimates.

Joint modeling of longitudinal outcomes with mixed types have been studied by many authors. Catalano and Ryan (1992)[12], Guergieva and Agresti (2001)[32], Fitzmaurice and Laird (1995)[19] have worked on the combination of binary and continuous responses. Guerguieva and Sanacora (2006)[33], and Catalano (1997)[11] have dealt with a combination of ordinal and continuous responses. In modeling the relationship between the two processes, they used random effects (Guerguieva and Sanacora, 2006)[33] or the product of marginal distribution and conditional distributions where the association is induced through the mean

responses (Fitzmaurice and Laird, 1995) or correlated residual errors (Catalano and Ryan, 1992[12]; Guergieva and Agresti, 2001[32]; Catalano, 1997[11]). Although, the random effects approach has the disadvantage of handling high-dimensional vectors of random effects, the product distribution approach can lead to very different results depending on whether the conditioning variable is discrete or continuous. It can also be difficult to obtain easy expressions for the association between both continuous and discrete outcomes and it does not directly lead to marginal inference. Fitzmaurice and Laird[19], Catalano and Ryan[12], and Catalano[11] used the generalized estimating equations (GEE) approach of Liang and Zeger (1986)[60] for parameter estimation whereas Guerguieva and Sanacora[33] parameterized their models to be fit using NLMIXED procedure in SAS (Wolfinger, 1999)[99] which employs Gaussian quadrature methods to obtain the maximum likelihood (ML) estimates. Gueorguieva and Agresti[32] employed a modified EM algorithm or Monte Carlo ECM algorithm for parameter estimation. Compared with ML approaches, the GEE method is computationally easy to implement and leads to consistent parameter estimates even when the working correlation structure is misspecified under mild regulatory conditions. However, GEE is not a likelihood-based method and hence it is difficult to determine the goodness of fit of a model, to compare models, and to draw statistical conclusions on the model parameters. In addition, it will only produce consistent estimates for unbalanced data or missing data when the data are missing completely at random (Little and Rubin, 1987)[62]. On the other hand, both the Gaussian quadrature and the Monte Carlo ECM algorithm methods are computationally very intensive. This computational burden grows exponentially with the number of random effects in the model. In this paper, we propose a hierarchical joint model to handle unbalanced repeatedly measured continuous and ordinal markers of disease severity. We employ the Markov Chain Monte Carlo (MCMC) method for parameter estimation because it has the capacity to handle high-dimensional data with ease.

The remainder of this chapter is organized as follows: Section 3.2 presents the formulation of the multivariate generalized linear mixed effects model, the associated joint likelihood, and the prior and posterior distributions. Sections 3.3 and 3.4 show the estimation procedures of the parameters from their full conditionals and the convergence diagnostics and model assessment tools, respectively. Section 3.5 shows the simulation study and Section 3.6 indicates the application of the proposed joint model to diabetes data.

3.2 MODEL SPECIFICATIONS

3.2.1 Model formulation

Let $\mathbf{y}_i = (\mathbf{y}'_{i1}, \mathbf{y}'_{i2}, \ldots, \mathbf{y}'_{iL})'$, denote the L-variate response vector for i^{th} subject ($i = 1, \ldots, n$), where $\mathbf{y}_{il}, l = 1, \ldots, L$, is an $n_{il} \times 1$ vector of longitudinal biomarker for a certain disease severity taken at time points, $j = 1, \ldots, n_{il}$. For instance, $\mathbf{y}_{i1}$ and $\mathbf{y}_{i2}$ can be a vector of blood glucose and urine glucose levels for the i^{th} patient, respectively. Because these responses are assumed to have different scales of measurements (i.e., continuous, ordinal), for each response, we adopt a generalized linear mixed effects model (GLMM) which is an extension of generalized linear model (GLM) (Nelder and Wedderburn, 1972[72]; McCullagh and Nelder, 1989[68]). In particular, marginally, the conditional distribution of $\mathbf{y}_{il}$ given a vector of random effects $\mathbf{b}_{il}$ is assumed to be a member of exponential family, with linear predictor given by

$$\mathsf{g}_l(\mu_{ij,l}) = \mathsf{g}_l(E[y_{ij,l}|\mathbf{b}_{il}]) = \eta_{ij,l}, \tag{3.2.11}$$

where $\mathsf{g}_l(\cdot)$ denotes a known one-to-one monotonic link function, and $y_{ij,l}$ denotes the value of the l^{th} longitudinal outcome for the i^{th} subject at j^{th} time point. The unknown function $\eta_{ij}(\cdot)$ is assumed to describe the true, presumably nonlinear, longitudinal profile for the l^{th} outcome (Rizopoulos and Ghosh, 2011)[79].

In general, the distribution of the j^{th} component of the l^{th} vector $\mathbf{y}_{il}$ is given by

$$f_l(y_{ij,l}|\mathbf{b}_{il}) = \exp\left[\frac{y_{ij,l}\theta_{ij,l} - b(\theta_{ij,l})}{\phi_l} + c(y_{ij,l}, \phi_l)\right], \tag{3.2.12}$$

where $\theta_{ij,l}$ is the canonical parameter, ϕ_l is the scale parameter for the l^{th} outcome. The conditional mean $\mu_{ij,l} = E(y_{ij,l}|\mathbf{b}_{il})$ is related to the canonical parameter $\theta_{ij,l}$ via $\mu_{ij,l} = b'(\theta_{ij,l})$ and to the regression coefficients via the link relation $\mathsf{g}_l(\mu_{ij,l}) = \eta_{ij,l} = x'_{ij,l}\beta_l + z'_{ij,l}\mathbf{b}_{il}$, where $x_{ij,l}$ is an $p \times 1$ covariate vector and $z_{ij,l}$ is an $r \times 1$ design vector for random effects. The conditional variance $v_{ij,l} = \mathrm{Var}(y_{ij,l}|\mathbf{b}_{il})$ is a function of the mean, that is, $v_{ij,l} = b''(\theta_{ij,l})\phi_l = \mathsf{v}_l(\mu_{ij,l})\phi_l$. The link and variance functions g_l and v_l, respectively, and the scale parameter ϕ_l are assumed to be known. Because we have repeatedly measured outcomes where observations are correlated, the linear predictor includes the fixed effects

and random effects and hence is modeled as a multivariate GLMM. Conditional on these random effects, the outcomes are assumed to be independent and the repeated measurements within an individual are assumed to be independent observations from a distribution $f_l(\cdot)$. This is referred to as the "conditional independence assumption" (Laird and Ware, 1982)[57]. Thus, the shared latent terms or the random effects $\mathbf{b}_{il}$ account for all dependencies among the observed data (Diggle et al., 2002, pp.129)[17]. The random effects $\mathbf{b}_{il}$ are mutually independent with a common underlying multivariate distribution $g_l(\mathbf{b}_{il}; \mathbf{\Gamma}_l)$.

Let θ denote a vector of parameters, which is a conglomerate of outcome-specific parameters $\theta_1, \theta_2, \ldots, \theta_L$, i.e., $\theta = (\theta_1, \theta_2, \ldots, \theta_L)$. Then we model the joint distribution of $\mathbf{y}_i$ based on full conditional independence assumption as

$$f(\mathbf{y}_i|\mathbf{b}_i, \theta) = \prod_{l=1}^{L} f_l(\mathbf{y}_{il}|\mathbf{b}_{il}, \theta_l). \tag{3.2.13}$$

To estimate the parameters of interest using Bayesian methods, we specify the priors for the parameters and then the posterior inference is obtained by using the likelihood to convert prior uncertainty into posterior probability statements. The joint posterior of the parameters based on the observed data $\mathbf{y}$ and random effects $\mathbf{b}_i$ is

$$\pi(\theta|\mathbf{y}, \mathbf{b}_i) \propto \prod_{i=1}^{n} \prod_{l=1}^{L} f_l(\mathbf{y}_{il}|\mathbf{b}_{il}, \theta_l) \times \pi(\theta_{\mathbf{1}}, \theta_{\mathbf{2}}, \ldots, \theta_{\mathbf{L}}), \tag{3.2.14}$$

where $\pi(\theta_{\mathbf{1}}, \theta_{\mathbf{2}}, \ldots, \theta_{\mathbf{L}})$ denotes the prior distribution of θ. Because random effects are unknown, they need to be included in the posterior distribution and then integrated over to obtain the marginal posterior distributions of the parameters of interest. Thus, the joint posterior distribution for $\mathbf{b}_i$ and other parameters of interest is given by the hierarchical model

$$\pi(\theta, \mathbf{b}_i|\mathbf{y}) \propto \prod_{i=1}^{n} \prod_{l=1}^{L} f_l(\mathbf{y}_{il}|\mathbf{b}_{il}, \theta_l) g(\mathbf{b}_i|\mathbf{\Gamma}) \times \pi(\theta_{\mathbf{1}}, \theta_{\mathbf{2}}, \ldots, \theta_{\mathbf{L}}) \pi(\mathbf{\Gamma}). \tag{3.2.15}$$

Motivated by our data set, we consider two response variables, continuous and ordinal and hence, $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2})$ with identity and logit links, respectively. Specifically, for the continuous outcome, we have

$$y_{ij,1} = x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i1} + \epsilon_{ij,1}, \tag{3.2.16}$$

where $\epsilon_{i1} \sim N(0, \sigma^2_{e_1})$ is the measurement or intra-subject error and the linear predictor or mean is given as

$$\mu_{ij,1} = E\left[y_{ij,1}|\mathbf{b}_{i1}\right] = \eta_{ij,1} = x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i1}.$$

For the ordinal response variable $\mathbf{y}_{i2}$ with K ordered categories coded as $k = 1, 2, \cdots, K$, we define the conditional cumulative probabilities for the K categories as

$$p_{ijk,2} = Pr(y_{ij,2} \leq k) = \sum_{m=1}^{k} p_{ijm,2}, \tag{3.2.17}$$

where $p_{ijk,2}$ represents the conditional probability of response being in category k, $k = 1, \ldots, K$, of the i^{th} subject at the j^{th} time point. Then the logistic GLMM for the conditional cumulative probabilities is given in terms of the cumulative logit as

$$\log\left[\frac{p_{ijk,2}}{1 - p_{ijk,2}}\right] = \eta_{ijk,2} = \alpha_k - \left[x'_{ij,2}\beta_2 + z'_{ij,2}\mathbf{b}_{i2}\right], \tag{3.2.18}$$

with $K - 1$ strictly increasing model thresholds α_k ($i.e., \alpha_1 < \alpha_2 \cdots < \alpha_{K-1}$). The thresholds allow the cumulative response probabilities to be different. For identifiability, either the first threshold α_1 or the model intercept $\beta_{20} \in \beta_{\mathbf{2}}$ is usually set to zero. In this formulation, we are assuming the proportional odds assumption (McCullagh, 1980 [67]) where the covariates do not vary across categories. The conditional probability of a response in category k is obtained as the difference of two conditional cumulative probabilities:

$$\pi_{ijk} = Pr(y_{ij,2} = k|\mathbf{b}_{i2}, x_{i2}, z_{i2}) = \Psi(\eta_{ijk,2}) - \Psi(\eta_{ijk-1,2}), \tag{3.2.19}$$

where $\Psi(\eta_{ijk,2})$ is the logistic cumulative distribution function (cdf) given as

$$\Psi(\eta_{ijk,2}) = \frac{\exp(\eta_{ijk,2})}{1 + \exp(\eta_{ijk,2})} = \frac{1}{1 + \exp(-\eta_{ijk,2})}.$$

Here, $\alpha_0 = -\infty$ and $\alpha_K = \infty$, and so $\Psi(\eta_{ij0,2}) = 0$ and $\Psi(\eta_{ijK,2}) = 1$. Thus, the proposed multivariate generalized linear mixed effects model assumes

$$\begin{aligned} y_{ij,1}|\mathbf{b}_{i1} &\sim N\left(x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i1}, \sigma^2_{e_1} I_{ni}\right), \\ (y_{ij1,2}, \ldots, y_{ijK-1,2})|\mathbf{b}_{i2} &\sim \text{multinomial}(\pi_{ij1}, \ldots, \pi_{ijK-1}). \end{aligned} \tag{3.2.110}$$

The random effects $\mathbf{b}_{i1}$ and $\mathbf{b}_{i2}$ are assumed to follow a Gaussian distribution with mean vectors of zeros and precision matrices ${\mathbf{\Gamma}_1}^{-1}$ and ${\mathbf{\Gamma}_2}^{-1}$, respectively. In this paper, we

investigate both the shared and correlated random effects. In the shared random effect model, $\mathbf{b}_{i2}$ is assumed proportional to $\mathbf{b}_{i1}$ (e.g. $\mathbf{b}_{i2} = \gamma \mathbf{b}_{i1}$) with a restrictive correlation structure between the two outcomes while in the correlated random effects model, $\mathbf{b}_{i1}$ and $\mathbf{b}_{i2}$ are assumed to follow a multivariate distribution with a nonrestrictive covariance structure, which can be unstructured, Toeplitz, exchangeable, etc. In this particular case, with just two processes we considered the simple correlation structure.

3.2.2 Likelihood, Prior, and Posterior distribution for the proposed model

Let $\theta_{\mathbf{1}} = \left\{\beta_1, \sigma^2_{e_1}\right\}$, $\theta_{\mathbf{2}} = \{\beta_2, \alpha\}$ where $\alpha = (\alpha_1, \ldots, \alpha_{K-1})$ are the ordered threshold parameters for ordinal process, $\theta = (\theta_{\mathbf{1}}, \theta_{\mathbf{2}})$, and $\mathbf{\Gamma} = (\mathbf{\Gamma_1}, \mathbf{\Gamma_2})$ denote the parameters associated with the continuous, ordinal, combined processes, and random effects, respectively. In addition, let $\mathbf{y_1}$ and $\mathbf{y_2}$ be the observed continuous and ordinal data, and $\mathbf{b}$ the combined random effects. Under the correlated normal random effect model, the joint likelihood of the two processes is then given as

$$\begin{aligned} L(\theta, \mathbf{\Gamma}|\mathbf{b}, \mathbf{y}) &= L_1(\theta_{\mathbf{1}}|\mathbf{b}, \mathbf{y_1}) L_2(\theta_{\mathbf{2}}|\mathbf{b}, \mathbf{y_2}) g(\mathbf{b_1}, \mathbf{b_2}|\mathbf{\Gamma}) \\ &= \prod_{i=1}^{n} \left[\prod_{j=1}^{n_{i1}} \frac{1}{(2\pi\sigma^2_{e_1})^{1/2}} \exp\left\{ -\frac{(y_{ij,1} - \mu_{ij,1})^2}{2\sigma^2_{e_1}} \right\} \right] \times \\ &\quad \left[\prod_{j=1}^{n_{i2}} \prod_{k=1}^{K} \{\Psi(\eta_{ijk,2}) - \Psi(\eta_{ijk-1,2})\}^{y_{ijk,2}} \right] \left[\frac{|\mathbf{\Gamma}|^{-1/2}}{2\pi} \exp\left\{ -\frac{\mathbf{b}_i' \mathbf{\Gamma}^{-1} \mathbf{b}_i}{2} \right\} \right], \end{aligned} \tag{3.2.21}$$

where

$\mu_{ij,1} = x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i1}$, $\Psi(\eta_{ijk,2}) = \frac{\exp(\alpha_k - \mu_{ij,2})}{1+\exp(\alpha_k - \mu_{ij,2})}$, $\Psi(\eta_{ijk-1,2}) = \frac{\exp(\alpha_{k-1} - \mu_{ij,2})}{1+\exp(\alpha_{k-1} - \mu_{ij,2})}$, and $\mu_{ij,2} = x'_{ij,2}\beta_{\mathbf{2}} + z'_{ij,2}\mathbf{b}_{i2}$.

Furthermore, let $\mathring{\beta}_{\mathbf{1}}$ and $\mathring{\beta}_{\mathbf{2}}$, $\mathring{\mathbf{\Sigma}}_{\mathbf{1}}$ and $\mathring{\mathbf{\Sigma}}_{\mathbf{2}}$ denote the mean vectors and variance-covariance matrices for $\beta_{\mathbf{1}}$ and $\beta_{\mathbf{2}}$, respectively. We assume non-informative multivariate normal priors for the $\beta' s$, $\beta_{\mathbf{1}} \sim MVN\left(\mathring{\beta}_{\mathbf{1}}, \mathring{\mathbf{\Sigma}}_{\mathbf{1}}\right), \beta_{\mathbf{2}} \sim MVN\left(\mathring{\beta}_{\mathbf{2}}, \mathring{\mathbf{\Sigma}}_{\mathbf{2}}\right)$ and truncated normal prior for the thresholds α, $\alpha_k \sim N\left(\mu_{\alpha_k}, \sigma^2_{\alpha}\right) I(\alpha_{k-1}, \alpha_{k+1}) \quad k = 1, \ldots, K-1$, where $I(\cdot, \cdot)$ denotes truncation to specified interval, by having large variances or small precision. Alternatively, a uniform prior could be used for the thresholds. Furthermore, an Inverse Wishart prior is

assumed for the variance-covariance matrix of the random effects ($\mathbf{\Gamma} \sim IW(\nu, \Lambda)$) and an Inverse Gamma for the error variance $\sigma^2_{e_1}$, that is, $\sigma^2_{e_1} \sim IG(\zeta, \omega)$, which are both conjugate priors for the variance-covariance matrix in the multivariate and univariate normal likelihoods, respectively (Carlin and Louis, 2009)[9]. We choose non-informative priors so that the priors will have little impact relative to the data on the inferences made. Then given the prior distributions of all unknowns, and the observed data, the full conditional assumption presented in Section 3.2 implies that the joint posterior distribution can be expressed as

$$\begin{aligned}
\pi(\theta, \mathbf{\Gamma}, \mathbf{b}|\mathbf{y}) &\propto L(\theta, \mathbf{\Gamma}|\mathbf{b}, \mathbf{y})\pi(\theta)\,\pi(\mathbf{\Gamma}) \\
&\propto L(\theta, \mathbf{\Gamma}|\mathbf{b}, \mathbf{y}) \times \left(\sigma^2_{e_1}\right)^{-(\zeta+1)} \exp\left\{-\frac{\omega}{\sigma^2_{e_1}}\right\} \\
&\times \exp\left\{-\frac{1}{2}\left(\beta_1 - \mathring{\beta}_1\right)' \mathring{\mathbf{\Sigma}}_1^{-1}\left(\beta_1 - \mathring{\beta}_1\right)\right\} \times \prod_{k=1}^{K-1} \exp\left\{-\frac{(\alpha_k - \mu_\alpha)^2}{2\sigma^2_\alpha}\right\} I_{[.,.]}(\alpha_k) \\
&\times \exp\left\{-\frac{1}{2}\left(\beta_2 - \mathring{\beta}_2\right)' \mathring{\Sigma}_2^{-1}\left(\beta_2 - \mathring{\beta}_2\right)\right\} \times |\mathbf{\Gamma}|^{-\nu/2} \exp\left\{-\frac{1}{2} tr\left(\mathbf{\Gamma}^{-1}\Lambda\right)\right\},
\end{aligned} \tag{3.2.22}$$

where $L(\theta, \mathbf{\Gamma}|\mathbf{b}, \mathbf{y})$ is given by Equation (3.2.21). For ease of sampling, the parameters are divided into blocks of correlated parameters based on their conditional independence. The full conditional posterior distributions for the blocks $\beta_{\mathbf{1}}$, $\sigma^2_{e_1}$, $(\alpha, \beta_{\mathbf{2}})$, $\mathbf{b}$, and $\mathbf{\Gamma}$ are then determined by averaging the joint posterior distribution (3.2.22) over or integrating out the remaining parameters. Let $\pi(\theta|.)$ represent the full conditional distribution of parameter θ given other parameters in the model. The full conditionals for the blocks $\beta_{\mathbf{1}}$ and $(\alpha, \beta_{\mathbf{2}})$ are given as:

$$\begin{aligned}
\pi\left(\beta_{\mathbf{1}}|\sigma^2_{e_1}, \mathbf{b}_{i1}\right) &\propto \exp\left\{-\tfrac{1}{2}\left(\beta_{\mathbf{1}} - \mathring{\beta}_{\mathbf{1}}\right)' \mathring{\mathbf{\Sigma}}_{\mathbf{1}}^{-1}\left(\beta_{\mathbf{1}} - \mathring{\beta}_{\mathbf{1}}\right)\right\} \times \exp\left\{-\tfrac{1}{2\sigma^2_{e_1}} \sum_{i=1}^{n}\sum_{j=1}^{n_{i1}} \left(y_{ij,1} - \mu_{ij,1}\right)^2\right\} \\
&\propto N\left(\beta_{\mathbf{1}}{}^*, \mathbf{\Sigma_1}^*\right),
\end{aligned} \tag{3.2.23}$$

where $\beta^*_{\mathbf{1}} = \mathbf{\Sigma_1}^* \times \left[\mathring{\mathbf{\Sigma}}_{\mathbf{1}}^{-1}\mathring{\beta}_{\mathbf{1}} + X_1'\sigma_{e_1}^{-2}\epsilon\right], \mathbf{\Sigma_1}^* = \left[\mathring{\mathbf{\Sigma}}_{\mathbf{1}}^{-1} + X_1'X_1\sigma_{e_1}^{-2}\right]^{-1}$,

$\epsilon = \sum_{i=1}^{n}\sum_{j=1}^{n_{i1}} (y_{ij,1} - z_{ij,1}\mathbf{b}_{i1})$.

$$\begin{aligned}
\pi\left(\alpha,\beta_{\mathbf{2}}|\mathbf{b}_{i2}\right) &= \pi\left(\alpha|\mathbf{b}_{i2}\right)\pi\left(\beta_2|\alpha,\mathbf{b}_{i2}\right) \\
&\propto \prod_{k=1}^{K-1}\exp\left\{-\frac{\left(\alpha_k-\mu_\alpha\right)^2}{2\sigma_\alpha^2}\right\} I_{[\alpha_{k-1},\alpha_{k+1}]}\left(\alpha_k\right)\times \\
&\quad \prod_{i=1}^{n}\prod_{j=1}^{n_{i2}}\prod_{k=1}^{K}\left\{\frac{\exp(\alpha_k-\mu_{ij,2})}{1+\exp(\alpha_k-\mu_{ij,2})}-\frac{\exp(\alpha_{k-1}-\mu_{ij,2})}{1+\exp(\alpha_{k-1}-\mu_{ij,2})}\right\}^{y_{ijk,2}} \\
&\quad \times\exp\left\{-\frac{1}{2}\left(\beta_2-\mathring{\beta}_2\right)'\mathring{\Sigma}_2^{-1}\left(\beta_2-\mathring{\beta}_2\right)\right\}
\end{aligned} \tag{3.2.24}$$

The full conditional posterior distributions for blocks $\sigma_{e_1}^2$, $\mathbf{b}$, and $\mathbf{\Gamma}$ are derived in the similar manner as:

$$\begin{aligned}
\pi\left(\sigma_{e_1}^2|\beta_1,\mathbf{b}_{i1}\right) &\propto \left(\sigma_{e_1}^2\right)^{-n(\zeta+1)}\exp\left\{-\frac{\omega}{\sigma_{e_1}^2}\right\}\times\prod_{i=1}^{n}\prod_{j=1}^{n_{i1}}\frac{1}{(\sigma_{e_1}^2)^{1/2}}\exp\left\{-\frac{(y_{ij,1}-\mu_{ij,1})^2}{2\sigma_{e_1}^2}\right\} \\
&\propto \left(\sigma_{e_1}^2\right)^{-n(\zeta+1)}\exp\left\{-\frac{\omega}{\sigma_{e_1}^2}\right\}\times(\sigma_{e_1}^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_{e_1}^2}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}\left(y_{ij,1}-\mu_{ij,1}\right)^2\right\} \\
&\propto \left(\sigma_{e_1}^2\right)^{-\left(\frac{2\zeta+n}{2}+1\right)}\exp\left\{-\frac{1}{\sigma_{e_1}^2}\left[\omega+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}\left(y_{ij,1}-\mu_{ij,1}\right)^2\right]\right\} \\
&\propto IG\left(\zeta+\frac{n}{2},\omega+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}\left(y_{ij,1}-\mu_{ij,1}\right)^2\right)
\end{aligned} \tag{3.2.25}$$

$$\begin{aligned}
\pi\left(\mathbf{b}|.\right) &\propto |\Gamma|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}b_i'\Gamma^{-1}b_i\right\}\times|\Gamma|^{-\nu/2}\exp\left(-\frac{1}{2}tr\left(\Gamma^{-1}\Lambda\right)\right) \\
&\quad\times\exp\left\{-\frac{1}{2\sigma_{e_1}^2}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}\left(y_{ij,1}-\mu_{ij,1}\right)^2\right\} \\
&\quad\times\prod_{i=1}^{n}\prod_{j=1}^{n_{i2}}\prod_{k=1}^{K}\left\{\frac{\exp(\alpha_k-\mu_{ij,2})}{1+\exp(\alpha_k-\mu_{ij,2})}-\frac{\exp(\alpha_{k-1}-\mu_{ij,2})}{1+\exp(\alpha_{k-1}-\mu_{ij,2})}\right\}^{y_{ijk,2}}
\end{aligned} \tag{3.2.26}$$

$$\begin{aligned}\pi\left(\Gamma|\mathbf{b}\right) &\propto |\Gamma|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n} b_i'\Gamma^{-1}b_i\right\}\times|\Gamma|^{-\nu/2}\exp\left(-\frac{1}{2}tr\left(\Gamma^{-1}\Lambda\right)\right)\\ &\propto |\Gamma|^{-(n+\nu)/2}\exp\left(-\frac{1}{2}tr\left[\Gamma^{-1}\left(\Lambda+\sum_{i=1}^{n} b_i b_i'\right)\right]\right)\\ &\propto IW\left(\nu+n,\Lambda+\sum_{i=1}^{n} b_i b_i'\right).\end{aligned} \tag{3.2.27}$$

3.3 ESTIMATION

The parameters of interest are estimated by drawing random variates from their full conditional posterior distributions. To estimate, the variance-covariance parameters $\sigma^{\mathbf{2}}_{\mathbf{e_1}}, \mathbf{\Gamma}$, Gibbs sampling is employed, while for the fixed and random effects parameters, $(\beta_{\mathbf{1}}, \alpha, \beta_{\mathbf{2}})$ and $\mathbf{b}_i$, Gamerman's one step Metropolis-Hasting (M-H) method is employed to sample from their respective conditional posterior distributions (Gamerman, 1997)[20]. In this one step M-H method, the full conditionals of the parameters of interest are approximated by a Gaussian distribution, which is obtained by accomplishing one Fisher scoring step in every iteration of the sampler. In essence, to estimate parameter φ using a single iterative method of Gamerman[20], the following steps are taken.

Step 1: Start with $\varphi = \varphi^{(0)}$ and set $t = 1$;

Step 2a: Sample φ^* from $N(\mathbf{m}^{(t)}, \mathbf{c}^{(t)})$ proposal density and

Step 2b: Accept it with probability $\lambda(\varphi^{(t-1)}, \varphi^*)$ and set $\varphi^{(1)} = \varphi^*$; Otherwise, stay at $\varphi^{(t)} = \varphi^{(t-1)}$;

Step 3: Increase t by 1 and return to Step 2.

The moments of the proposal density are given by

$$\begin{aligned}\mathbf{m}^{(t)} &= (\Sigma_\varphi^{-1} + X'W(\varphi^{(t-1)})X)^{-1}\times\left\{\Sigma_\varphi^{-1}\mu_\varphi + X'W(\varphi^{(t-1)})[\tilde{y}(\varphi^{(t-1)}) - \tilde{\eta}]\right\}\\ \mathbf{c}^{(t)} &= (\Sigma_\varphi^{-1} + X'W(\varphi^{(t-1)})X)^{-1}\end{aligned} \tag{3.3.01}$$

where μ_φ and Σ_φ are respectively, the mean and variance-covariance matrix of the prior distribution for φ, $W(\varphi^{(t-1)}) = diag(W_{11}, \cdots, W_{nn_i})$ is the usual weight matrix for iterative weighted least squares (IWLS) algorithm. The vector $\tilde{\eta}$ known as the offset in GLM is the part of the predictor associated with all the remaining effects in the model. The components of the weight matrix W_{ij} and the transformed observations $\tilde{y}_{ij}$ are defined as

$$\begin{aligned} \tilde{y}_{ij}(\varphi) &= \eta_{ij} + (y_{ij} - \mu_{ij})g'(\mu_{ij}) \text{ and} \\ W_{ij}^{-1}(\varphi) &= \mathbf{V}_{ij}\left\{g'(\mu_{ij})\right\}^2, \ i = 1, \ldots, n; j = 1, \ldots, n_i, \end{aligned} \tag{3.3.02}$$

where $\mathbf{V}_{ij}$ is the conditional variance function of the outcome variable, and $g'(\mu_{ij})$ is the derivative of the link function with respect to the the mean value function. The acceptance probability is defined as

$$\lambda(\varphi^{(t-1)}, \varphi^*) = \min\left(1, \frac{\pi(\varphi^*)q(\varphi^{(t-1)}, \varphi^*)}{\pi(\varphi^{(t-1)})q(\varphi^*, \varphi^{(t-1)})}\right), \tag{3.3.03}$$

where $\pi(\varphi^*)$ and $\pi(\varphi^{(t-1)})$ is the posterior density of φ evaluated at φ^* and $\varphi^{(t-1)}$, respectively; $q(\varphi^{(t-1)}, \varphi^*)$ is the density specified in Step 2a evaluated at φ^* and $q(\varphi^*, \varphi^{(t-1)})$ is a $N(\mathbf{m}^*, \mathbf{c}^*)$ density evaluated at $\varphi^{(t-1)}$. Thus, to draw samples from the full conditionals $\pi(\beta_1|.)$, $\pi(\alpha, \beta_2|.)$, and $\pi(b_i|.) = \pi(b_{i1}, b_{i2}|.)$, the steps above are followed.

For the β_1 block, the transformed observations are $\tilde{y}_{ij,1}(\beta_1) = x'_{ij,1}\beta_{\mathbf{1}} + (y_{ij,1} - x'_{ij,1}\beta_{\mathbf{1}})g'(x'_{ij,1}\beta_{\mathbf{1}})$ which gives the original observations, $y_{ij,1}$; the offset is the random effect part, $z'_{ij,1}b_{i1}$, and the weights are $W_{ij,1}(\beta_1) = \sigma^2_{e_1} I_{n_{i1}}, i = 1, \ldots, n; j = 1, \ldots, n_{i1}$. The proposal density $N(\mathbf{m_1}^{(t)}, \mathbf{c_1}^{(t)})$ has moments

$$\begin{aligned} \mathbf{m_1}^{(t)} &= (\mathring{\Sigma}_1^{-1} + X_1'W_1(\beta_1^{(t-1)})X_1)^{-1} \times \left\{\mathring{\Sigma}_1^{-1}\mathring{\beta}_1 + X_1'W_1(\beta_1^{(t-1)})\left[\tilde{\mathbf{y}}_{\mathbf{1}}(\beta_{\mathbf{1}}^{(\mathbf{t-1})}) - \mathbf{z}'_{\mathbf{1}}\mathbf{b_1}\right]\right\} \\ \mathbf{c_1}^{(t)} &= (\mathring{\Sigma}_1^{-1} + X_1'W_1(\beta_1^{(t-1)})X_1)^{-1} \end{aligned} \tag{3.3.04}$$

where $W_1 = diag(W_{11,1}, \ldots, W_{nn_{i1},1})$; X_1 is the design matrix of fixed effects for outcome $\mathbf{y_1}$.

For the $\theta_{\mathbf{2}} = \{\alpha, \beta_2\}$ block associated with the ordinal outcome $\mathbf{y_2}$ with response vector for the i^{th} subject defined as $\mathbf{y}_{i2} = (y_{i1,2}, \cdots, y_{ij,2}, \cdots, y_{in_{i2},2})'$, we define $y^*_{ij,2} = 1$ if $y_{ij,2} = k, 0$ otherwise, with its expectation $\pi_{ij,2} = E(y^*_{ij,2})$ defined as in Equation (3.2.19). Thus, the $n_{i2} \times 1$ dimensional ordinal response vector $\mathbf{y}_{i2}$ is transformed into a $n_{i2}(K-1)$ dimensional

binary vector $\mathbf{y}^*_{i2} = (y_{i11}, \cdots, y_{i1K-1}, y_{i21}, \cdots, y_{in_{i2}K-1})'$ with expectation $\pi_{i2} = E(\mathbf{y}^*_{i2})$. The variance-convariance matrix $\mathbf{V}_{i2}$ of the dichotomized binary response vector $\mathbf{y}^*_{i2}$ has typical elements

$$\text{cov}(y_{ijk}, y_{ij'k'}) = \begin{cases} \pi_{ijk}(1-\pi_{ijk}) & \text{if} \quad j = j', k = k', \\ -\pi_{ijk}\pi_{ijk'} & \text{if} \quad j = j', k \neq k', \\ \frac{\text{corr}(y_{ijk}, y_{ij'k'})}{\left[\pi_{ijk}(1-\pi_{ijk})\pi_{ij'k'}(1-\pi_{ij'k'})\right]^{-1/2}} & \text{if} \quad j \neq j', \text{any } k, k' \end{cases} \tag{3.3.05}$$

Let $\mu_{\theta_2} = (\mu_\alpha, \mathring{\beta}_2)$ and $\Sigma_{\theta_2} = \begin{pmatrix} \sigma^2_\alpha I_{K-1} & 0 \\ 0 & \mathring{\Sigma}_2 \end{pmatrix}$ be the mean vector and variance-covariance matrix of $\theta_{\mathbf{2}}$, respectively. Thus, the transformed observations used in estimating $\theta_{\mathbf{2}}$, are $\tilde{y}_{ij,2}(\theta_2) = \eta_{ij,2}(\theta_2) + (y^*_{ij,2} - \pi_{ij,2}(\theta_2))g'(\pi_{ij,2}(\theta_2))$, where $\eta_{ij,2}(\theta_2) = \alpha_k - x'_{ij,2}\beta_2$. The offset and weights are $z'_{ij,2}b_{i2}$ and $W_{ij,2}(\theta_2) = [n_{i2}V_{ij,2}\left\{g'(\pi_{ij,2}(\theta_2))\right\}^2]^{-1}$, respectively, where $[g'(\pi_{ij,2}(\theta_2))]^{-1}$ is the derivative of the mean function with respect to the linear predictor whose elements are given as follows:

$$[g'(\pi_{ij,2}(\theta_2))]^{-1} = \begin{cases} \frac{\exp(\alpha_k - x'_{ij,2}\beta_2)}{(1+\exp(\alpha_k - x'_{ij,2}\beta_2))^2} & k = 1, \\ \left[\frac{\exp(\alpha_k - x'_{ij,2}\beta_2)}{(1+\exp(\alpha_k - x'_{ij,2}\beta_2))^2} - \frac{\exp(\alpha_{k-1} - x'_{ij,2}\beta_2)}{(1+\exp(\alpha_{k-1} - x'_{ij,2}\beta_2))^2}\right] & k \geq 2. \end{cases} \tag{3.3.06}$$

The proposal density $N(\mathbf{m_2}^{(t)}, \mathbf{c_2}^{(t)})$ has moments

$$\begin{aligned} \mathbf{m_2}^{(t)} &= (\Sigma^{-1}_{\theta_2} + X'_2 W_2(\theta_2^{(t-1)})X_2)^{-1} \times \left\{\Sigma^{-1}_{\theta_2}\mu_{\theta_2} + X'_2 W_2(\theta_2^{(t-1)})[\tilde{\mathbf{y}}_{\mathbf{2}}(\theta_{\mathbf{2}}^{(\mathbf{t-1})}) - \mathbf{z'_2 b_2}]\right\} \\ \mathbf{c_2}^{(t)} &= (\Sigma^{-1}_{\theta_2} + X'_2 W_2(\theta_2^{(t-1)})X_2)^{-1} \end{aligned} \tag{3.3.07}$$

where $W_2 = diag(W_{11,2}, \ldots, W_{nn_{i2},2})$ and X_2 is the design matrix of fixed effects for the binary outcome associated with $\mathbf{y_2}$.

Following the same steps, for the $b_i = (b_{i1}, b_{i2})$ block, when estimated separately, then for the b_{i1} block, we draw samples from the full conditional $\pi(b_{i1}|.)$. The transformed observations and weights are $\tilde{y}_{ij,1}(b_{i1}) = z'_{ij,1}b_{i1} + (y_{ij,1} - z'_{ij,1}b_{i1})g'(z'_{ij,1}b_{i1}) = y_{ij,1}$ and $W_{ij,1}(b_{i1}) = \sigma^2_\epsilon I_{n_{i1}}$, respectively. The proposal density is $N(\mathbf{m}^{(\mathbf{t})}_{i1}, \mathbf{c}^{(\mathbf{t})}_{i1})$ with moments

$$\begin{aligned} \mathbf{m}^{(\mathbf{t})}_{i1} &= (\Gamma_1^{-1} + Z'_{i1}W_{i1}(b_{i1}^{(t-1)})Z_{i1})^{-1}Z_{i1}W_{i1}(b_{i1}^{(t-1)}) \times \left\{\tilde{y}_{i1}(b_{i1}^{(t-1)}) - X'_{i1}\beta_1\right\} \\ \mathbf{c}^{(\mathbf{t})}_{i1} &= (\Gamma_1^{-1} + Z'_{i1}W_{i1}(b_{i1}^{(t-1)})Z_{i1})^{-1} \end{aligned} \tag{3.3.08}$$

where $W_{i1} = diag(W_{i1,1}, \ldots, W_{in_{i1},1})$ and $Z_{i1} = (z_{i1,1}, \ldots, z_{in_{i1},1})'$.

And for the b_{i2} block, we draw samples from the full conditional $\pi(b_{i2}|.)$. The transformed observations and weights for b_{i2} are $\tilde{y}^*_{ij,2}(b_{i2}) = \eta_{ij,2}(b_{i2}) + (y^*_{ij,2} - \pi_{ij,2}(b_{i2}))g'(\pi_{ij,2}(b_{i2}))$ and $W_{ij,2}(b_{i2}) = [n_{i2}V_{ij,2}\{g'(\pi_{ij,2}(b_{i2}))\}^2]^{-1}$, respectively. The proposal density is $N(\mathbf{m}^{(\mathbf{t})}_{i2}, \mathbf{c}^{(\mathbf{t})}_{i2})$ with moments

$$\begin{aligned}
\mathbf{m}^{(\mathbf{t})}_{i2} &= (\Gamma_2^{-1} + Z'_{i2}W_{i2}(b^{(t-1)}_{i2})Z_{i2})^{-1}Z_{i2}W_{i2}(b^{(t-1)}_{i2}) \times \left\{\tilde{y}^*_{i2}(b^{(t-1)}_{i2}) - (\alpha_k - X'_{i2}\beta_2)\right\} \\
\mathbf{c}^{(\mathbf{t})}_{i2} &= (\Gamma_2^{-1} + Z'_{i2}W_{i2}(b^{(t-1)}_{i2})Z_{i2})^{-1}
\end{aligned} \tag{3.3.09}$$

where $W_{i2} = diag(W_{i1,2}, \ldots, W_{in_{i2},2})$ and $Z_{i2} = (z_{i1,2}, \ldots, z_{in_{i2},2})'$.

Our goal is to estimate the random effects from a multivariate distribution. Thus, following the same steps above, we draw samples from the full conditional $\pi(b_i|.)$. The proposal density is

$$q_{b_i} \sim MVN\left(\begin{pmatrix} \mathbf{m}^{(\mathbf{t})}_{i1} \\ \mathbf{m}^{(\mathbf{t})}_{i2} \end{pmatrix}, \begin{pmatrix} \mathbf{c}^{(\mathbf{t})}_{i1} & \rho\sqrt{\mathbf{c}^{(\mathbf{t})}_{i1}}\sqrt{\mathbf{c}^{(\mathbf{t})}_{i2}} \\ \rho\sqrt{\mathbf{c}^{(\mathbf{t})}_{i1}}\sqrt{\mathbf{c}^{(\mathbf{t})}_{i2}} & \mathbf{c}^{(\mathbf{t})}_{i2} \end{pmatrix}\right), \tag{3.3.010}$$

where ρ is the correlation between the continuous and ordinal processes, which is estimated from the data.

3.4 CONVERGENCE DIAGNOSTICS AND MODEL ASSESSMENT

3.4.1 Convergence Diagnostics

From the theory of Markov chains governing the MCMC methods of estimation, the chains are expected to converge to the stationary distribution, which is also the target distribution, in the long run. In addition, as noted earlier, the first samples are discarded as burn-in, and inferences are made from the remaining samples. Thus, determining how much burn-in is optimal and whether the chains are mixing well or converged to the distribution of interest are of paramount interest in Bayesian analysis. Several methods that include visual inspection and statistical tests have been developed to diagnose convergence. These methods are full implemented in CODA (Convergence Diagnosis and Output Analysis) package (Best et al., 1996; Plummer et al., 2006)[3, 74] available in R software[75].

3.4.1.1 Visual Inspection One way to see if the chain has converged is to see how well that chain is mixing, or moving around the parameter space. If the chain is taking a long time to move around the parameter space, then it will take longer to converge. In this study, we employed trace and density plots, and autocorrelations to visually examine the mixing of the chains for each of the parameters, and to determine the optimal burn-in (Gilks et al., 1996)[31].

Trace and Density plots

A trace plot is a plot of the iteration number against the value of the draw of the parameter at each iteration. Proper mixing of the chains, hence convergence is exhibited if the chains remain stable for a longer period of time. A density plot on the other hand shows a smoothed probability density curve of the draws or the distribution of the parameters. A multimodal density may indicate non-convergence of the chain.

Autocorrelation

Another way to assess convergence is to assess the autocorrelations between the draws of the Markov chain. The lag k autocorrelation ρ_k is the correlation between every draw and its k^{th} lag:

$$\rho_k = \frac{\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The k^{th} lag autocorrelation is expected to be smaller as k increases (e.g., the 2^{nd} and 50^{th} draws should be less correlated than the 2^{nd} and 4^{th} draws). If autocorrelation is still relatively high for higher values of k, it is an indication of high degree of correlation between the draws and slow mixing.

3.4.1.2 Statistical Diagnostic Tests To substantiate the visual inspection results, we carried out statistical diagnostic tests that included Gelman and Rubin Multiple Sequence (Gelman and Rubin, 1992)[26], Geweke (Geweke, 1992)[30], and Heidelberg and Welch (Heidelberger and Welch, 1983) [41] diagnostic tests.

Gelman and Rubin Multiple Sequence Diagnostic

The Gelman and Rubin Multiple Sequence Diagnostic is based on comparing two or more parrallel chains drawn from different starting points and checking to see if they are not different

Comparison of the within and between chain variances for each parameter is carried out, where convergence is assumed to be reached when the within variance is equal or not less than the between chain variance. To implement Gelman and Rubin test, the following steps are carried out for each parameter:

1. Run $m \geq 2$ chains of length $2n$ from overdispersed starting values.
2. Discard the first n draws in each chain.
3. Calculate the within-chain and between-chain variance.
4. Calculate the estimated variance of the parameter as a weighted sum of the within-chain and between-chain variance.
5. Calculate the potential scale reduction factor.

The within-chain variance is defined as

$$W = \frac{1}{m} \sum_{j=1}^{m} S_j^2$$

where

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\theta_{ij} - \bar{\theta_j})^2,$$

is the variance of the j^{th} chain. Thus, W is the mean of the variance of each chain. To some extent, the within-chain variance underestimates the true variance of the stationary distribution because the chains may have not reached all the points of the stationary distribution.

The between-chain variance B is

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\theta_j} - \bar{\bar{\theta}})^2$$

where

$$\bar{\bar{\theta}} = \frac{1}{m} \sum_{j=1}^{m} \bar{\theta_j}.$$

Thus, B is the variance of the chain means multiplied by n because each chain is based on n draws.

The estimated variance of the stationary distribution is by

$$\hat{Var}(\theta) = \left(1 - \frac{1}{n}\right) W + \frac{1}{n} B$$

Because of overdispersion of the starting values, the estimated variance overestimates the true variance, but is unbiased if the starting distribution equals the stationary distribution (if starting values were not overdispersed).

The potential scale reduction factor is

$$\hat{R} = \sqrt{\frac{\hat{Var}(\theta)}{W}}$$

When $\hat{R}$ is high (i.e., greater than 1.1 or 1.2), the chains should be run longer to improve convergence to the stationary distribution. The potential reduction factor is calculated for each parameter of interest. Brooks and Gelman (1997) proposed a multivariate potential scale reduction factor which is computed for all parameters [5].

Geweke Diagnostic

The Geweke diagnostic takes two nonoverlapping parts (usually the first 10% and last 50%) of the Markov chain and compares the means of both parts, using equality of the means test. If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution. The test statistic is a standard Z-score: the difference between the two sample means divided by its estimated standard error. The standard error is estimated from the spectral density at zero and so takes into account any autocorrelation. The Z-score is calculated under the assumption that the two parts of the chain are asymptotically independent, which requires that the sum of first proportion and last proportion be strictly less than 1.

Heidelberg and Welch Diagnostic

The Heidelberg and Welch diagnostic calculates a test statistic (based on the Cramer-von Mises test statistic) to accept or reject the null hypothesis that the Markov chain is from a stationary distribution. The diagnostic consists of two parts.

- First Part:
    1. Generate a chain of N iterations and define an α level.

2. Calculate the test statistic on the whole chain. Accept or reject null hypothesis that the chain is from a stationary distribution.
3. If null hypothesis is rejected, discard the first 10% of the chain. Calculate the test statistic and accept or reject null.
4. If null hypothesis is rejected, discard the next 10% and calculate the test statistic.
5. Repeat until null hypothesis is accepted or 50% of the chain is discarded. If test still rejects null hypothesis, then the chain fails the test and needs to be run longer.

- Second Part:
  - If the chain passes the first part of the diagnostic, then it takes the part of the chain not discarded from the first part to test the second part.
  - The halfwidth test calculates half the width of the $(1-\alpha)\%$ credible interval around the mean.
  - If the ratio of the halfwidth and the mean is lower than some ϵ, then the chain passes the test. Otherwise, the chain must be run out longer.

3.4.2 Model Assessment

To asses the model fit and compare different models, we employed the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002)[87]. This model assessment tool was chosen because it is readily available in BUGS software and can be used with informative priors, noninformative priors, or improper priors. Let θ denote the vector of the model parameters and $\mathbf{y}$ denote the observed data, then the deviance $D(\theta)$ is defined as

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

where $f(\mathbf{y}|\theta)$ is the likelihood function and $h(\mathbf{y})$ is a standardizing function of the data alone (Carlin and Louis, 2009)[9]. The DIC is then computed with the formula:

$$DIC = \overline{\theta} + 2P_D,$$

where $P_D = \overline{D(\theta)} - D(\overline{\theta})$ is the effective number of parameters that measure model complexity. Smaller values of DIC are better as with other known model selection tools like Akaike

Information Criterion (AIC) and the Bayesian Information Criterion (BIC), and a difference of 10 is said to be meaningful. DIC is a generalization of AIC and BIC is more suitable for assessing goodness of fit for hierarchical models[9].

Other appropriate model assessment tools partly explored in this study include the Conditional Predictive Ordinate (CPO) statistic and the logarithm of the Pseudomarginal likelihood (LPML) statistic or its average value (ALPML). For the i^{th} observation, the CPO statistic is defined as

$$\text{CPO}_i = f(\mathbf{y}_i|D^{(-i)}) = \int f(\mathbf{y}_i|\theta, \mathbf{x}_i)\pi(\theta|D^{(-i)})d\theta,$$

where $\mathbf{y}_i$ denotes the response variable and $\mathbf{x}_i$ is the vector of covariates for case i, $D^{(-i)}$ denotes the data with the i^{th} case deleted, and $\pi(\theta|D^{(-i)})$ is the posterior density of θ based on the data $D^{(-i)}$ (see Ibrahim et al.(2001), Chapter 6.3 for more details)[51]. The LPML statistic (Geisser and Eddy, 1979)[21] is derived from the CPO as

$$LPML = \sum_{i=1}^{n} \log(CPO_i).$$

To compare LPML's from two different studies for a given model, the average LPML, given by

$$ALPML = \frac{LPML}{n},$$

where n is the sample size is preferred. In contrast with DIC, larger values of CPO or LPML or ALPML imply a better fitting model.

3.5 SIMULATION STUDY

To examine the performance of the regression estimators from the proposed joint model and to compare them to the estimators from the separate regression models, we performed a series of simulation studies. The data were simulated from the proposed joint model of continuous and ordinal outcomes correlated through correlated and/or shared random effects. From each of the joint models, we simulated 500 data sets of sample sizes $n = 100$ and $n = 50$.

In all simulations, number of repeated measures per subject ranged randomly from 1 to 10. For each subject, time t between successive visits were generated uniformly between 0.5 and 2.0 to mimic the motivating dataset. In addition, we generated one baseline covariate (treatment variable indicator) x from a Bernoulli (0.5) distribution. Motivated by the dataset, the models included only random intercepts. For the correlated random effects models, the random effects were generated from a multivariate normal distribution with mean vector zero and variance-covariance matrix $\Gamma = \begin{pmatrix} \sigma^2_{b_1} = 5.87 & \rho\sigma_{b_1}\sigma_{b_2} \\ \rho\sigma_{b_1}\sigma_{b_2} & \sigma^2_{b_2} = 4.89 \end{pmatrix}$. We considered different correlation values, $\rho = (0.9, 0.6, 0.0)$ which formed Part I of our simulations. For the shared random effects models ($b_{i2} = \gamma b_{i1}$), we considered $\gamma = (0.9, 0.6)$ in Part II of our simulations. The models for constructing the continuous and ordinal outcomes are as in Equation (3.2.110). The error $\epsilon_{ij,1}$ for the continuous outcome was simulated from $N\left(0, \sigma^2_{\epsilon_1} = 7.4\right)$. For the ordinal outcome, we considered three categories; the first threshold value was set to zero to guard against identifiability issues. Let the $\beta's$, $(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})$ and $(\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})$ denote the regression coefficients for the fixed effects (intercept, time, treatment, and time by treatment interaction) for the continuous and ordinal outcomes, respectively, and α_2 the threshold parameter for the ordinal outcome. The true values for all the variance and regression parameters were chosen based on the results of a joint correlated random effects model fit to the motivating dataset, namely, $\beta_{10} = 15.34, \beta_{11} = -0.56, \beta_{12} = -0.50, \beta_{13} = 0.3$, and $\alpha_2 = 1.25, \beta_{20} = 1.8, \beta_{21} = -0.35, \beta_{22} = -0.50, \beta_{23} = 0.1$. Once the latent parameters $(\mathbf{b}_{i1}, \mathbf{b}_{i2})$ were generated from their respective distributions, we generated $y_{ij,1}|b_{i1}$ from a normal distribution with mean $\mu_{ij,1} = \beta_{10} + \beta_{11} \times t_{ij} + \beta_{12} \times x_i + \beta_{13} \times t_{ij} \times x_i + \mathbf{b}_{i1}$ and standard deviation σ_{ϵ_1}. For the ordinal outcome, we simulated data from a multinomial distribution. The multinomial probabilities were the marginal probabilities constructed from the cumulative logit model as specified in Equation (3.2.19), where $\eta_{ijk,2} = \alpha_k - [\beta_{20} + \beta_{21} \times t_{ij} + \beta_{22} \times x_i + \beta_{23} \times t_{ij} \times x_i + \mathbf{b}_{i2}]$.

After generating the data, we fitted the joint (correlated through correlated and shared random effects) and separate models to each data set. For the MCMC sampling we ran two chains of 10,000 iterations with 2,000 iterations of each chain used as burn-in period. The initial values for MCMC sampling were taken from a linear mixed model fit to the continuous data and

a generalized linear mixed model fit to the ordinal data. The following priors were considered for the different parameters: $\beta_1 \sim N_4(0, 100I_4)$, $\beta_2 \sim N_4(0, 100I_4)$, $\alpha_2 \sim N(0, 10^6)I(0,)$, $\sigma^2_{e_1} \sim IG(0.001, 0.001)$, $\sigma^2_{b_1} \sim IG(0.001, 0.001)$, $\sigma^2_{b_2} \sim IG(0.001, 0.001)$, $\gamma \sim N(0, 100)$, and $\Gamma \sim IW(3, 1I_2)$, where I_q indicates an $q \times q$ identity matrix, and N, IG, and IW, respectively stand for Normal, Inverse Gamma, and Inverse Wishart. The MCMC sampling was done using OpenBUGS (version 3.2.2) software and its R interface BRugs Version 0.4-1.

The simulation results for Part I are shown in Tables 1-3. In each of the tables, the estimated Bias, Posterior Standard Deviation (SD), Coverage Probabilities (CP) of the 95% highest posterior density (HPD) intervals, and the Relative Efficiency (RE) are shown. RE is calculated as the ratio of the mean squared error (MSE) of estimates from the fitted models to the mean squared error (MSE) of estimates for the same parameters from the true model. All estimates were calculated based on 500 replicates.

The results in Table 1 ($\rho = 0.9$), indicate that when the true processes were correlated through correlated random effects both joint model and separate model fits provided unbiased estimates but the estimated posterior means were more biased for separate models with larger SD. These biases were larger for the ordinal outcome which may be due to the less informative nature of ordinal data as compared to continuous data. The gain in efficiency using joint model relative to the separate model was as high as 15% and was more pronounced in the ordinal outcome. However, when a shared random effects model was fitted to this correlated random effects data, the estimates from the shared random effects were smaller with similar SD with virtually no gain in efficiency when compared to estimates from correlated random effects model. In addition, the results in Table 1 indicate that nominal coverage of 95% HPD intervals was maintained for all the fitted models. The coverage probabilities were robust to the sample size as seen from the bottom panel of Table 1.

The results for moderate correlation of $\rho = 0.6$ are shown in Table 2. The results showed a similar trend as in Table 1. Specifically, when we fitted the correct model (JC) the biases were smaller than those when fitted the JS model. The standard errors (SD) were smaller for the JC (true) model compared to the JS model. In most cases there were considerable gain in efficiency as high as 2.35 for the threshold parameter α_2 with sample size $n = 100$ and extremely high for the variance parameters (highlighted in red). Coverage of 95% HPD

intervals were adequate except some subject–specific covariates for the ordinal outcome modeled through shared random effects.

Table 3 also indicates that when the true processes were uncorrelated ($\rho = 0$) the estimates from the separate and joint correlated random effects models were quite similar, though there was still some gain in efficiency for ordinal outcome estimates. Meanwhile, if a shared random effects model was fitted to this uncorrelated data, most of estimates were more biased and the trend of results was similar to those when $\rho = 0.6$.

Similar results were seen when separate models were fitted to data that were correlated through shared random effects (see Appendix A: Tables 19 & 20). Although fitting joint correlated random effects model to the shared random effects data resulted in more biased estimates, less efficient estimates, and slightly larger standard errors (SD) for some of the parameters for the ordinal outcome, the coverage of 95% HPD intervals were quite similar in all scenarios. In general, a joint correlated random effects model did not perform as poorly as a shared random effects model did in Part I simulations.

In essence, the estimates from fitting true models became less biased with smaller standard errors as the sample size increased in all scenarios. The coverage probabilities were quite robust to the sample size. The gain in efficiency, which was more pronounced in the ordinal outcome, reduced with the increase in sample size keeping the correlation constant, and decreased with the decrease in correlation between the two outcomes.

Table 1: Results when data were correlated under a correlated random effects model with strong correlation ($\rho = 0.9$): SD and CP, stand for posterior standard deviation and coverage probabilities of the 95% HPD intervals, respectively.

			Fitted Model										
			Joint-Correlated (JC)			Joint-Shared (JS)				Separate (SP)			
True Model (n)	Parameter	Truth	Bias	SD	CP	Bias	SD	CP	$RE_1 = \frac{MSE_{JS}}{MSE_{JC}}$	Bias	SD	CP	$RE_2 = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=50)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.032	0.581	0.94	-0.096	0.595	0.94	1.02	-0.051	0.601	0.94	0.98
	β_{11}: time	-0.56	0.001	0.070	0.94	0.001	0.070	0.94	1.02	0.001	0.072	0.94	1.07
	β_{12}: treatment	-0.50	0.017	0.841	0.95	0.132	0.851	0.94	1.00	0.060	0.850	0.94	1.00
	β_{13}: time×treatment	0.30	-0.003	0.099	0.95	-0.003	0.100	0.95	1.01	-0.002	0.102	0.95	1.10
	Ordinal Process												
	α_2: threshold	1.25	0.038	0.164	0.94	0.007	0.164	0.95	0.93	0.045	0.165	0.94	1.05
	β_{20}: intercept	1.80	0.069	0.561	0.95	-0.062	0.559	0.95	0.98	0.096	0.614	0.95	1.15
	β_{21}: time	-0.35	-0.013	0.073	0.95	-0.003	0.069	0.96	0.99	-0.012	0.076	0.94	1.13
	β_{22}: treatment	-0.50	-0.076	0.788	0.95	0.079	0.789	0.95	0.99	-0.102	0.841	0.94	1.13
	β_{23}: time×treatment	0.10	0.003	0.099	0.95	-0.003	0.092	0.95	1.01	0.001	0.103	0.95	1.11
	Association Parameters & Variances												
	ρ	0.90	0.012	0.050	0.97	-	-	-	-	-	-	-	-
	γ	0.91	-	-	-	0.064	0.148	0.93	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.136	1.464	0.95	-0.202	1.520	0.95	1.20	0.202	1.586	0.95	1.20
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	0.379	1.730	0.94	-	-	-	-	0.984	2.081	0.89	1.94
	$\sigma^2_{e_1}$: error	7.40	0.121	0.651	0.95	0.414	0.680	0.91	1.45	0.079	0.649	0.95	0.98
Correlated (n=100)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.009	0.417	0.95	-0.003	0.410	0.95	0.98	-0.036	0.425	0.94	1.01
	β_{11}: time	-0.56	0.000	0.048	0.96	0.001	0.048	0.96	1.01	-0.001	0.050	0.96	1.06
	β_{12}: treatment	-0.50	0.024	0.598	0.94	0.039	0.583	0.95	1.00	0.047	0.605	0.94	1.01
	β_{13}: time×treatment	0.30	-0.001	0.069	0.94	-0.001	0.069	0.94	1.02	-0.001	0.071	0.94	1.04
	Ordinal Process												
	α_2: threshold	1.25	0.016	0.115	0.94	-0.015	0.114	0.96	0.99	0.028	0.119	0.94	1.06
	β_{20}: intercept	1.80	0.030	0.400	0.95	-0.023	0.376	0.95	0.96	0.011	0.404	0.96	1.07
	β_{21}: time	-0.35	-0.004	0.050	0.95	0.005	0.049	0.96	1.00	-0.004	0.053	0.95	1.10
	β_{22}: treatment	-0.50	-0.013	0.556	0.95	0.031	0.517	0.95	0.97	0.042	0.553	0.94	1.05
	β_{23}: time×treatment	0.10	0.003	0.066	0.96	-0.001	0.065	0.96	1.01	-0.002	0.071	0.94	1.14
	Association Parameters & Variances												
	ρ	0.90	0.010	0.038	0.96	-	-	-	-	-	-	-	-
	γ	0.91	-	-	-	0.017	0.096	0.95	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.084	1.026	0.95	-0.291	1.012	0.93	1.29	0.103	1.076	0.95	1.12
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	0.142	1.139	0.95	-	-	-	-	0.443	1.283	0.93	1.56
	$\sigma^2_{e_1}$: error	7.40	0.057	0.449	0.95	0.369	0.476	0.88	1.81	0.020	0.446	0.95	0.98

Table 2: Results when data were correlated under a correlated random effects model with moderate correlation ($\rho = 0.6$): SD and CP, stand for posterior standard deviation and coverage probabilities of the 95% HPD intervals, respectively.

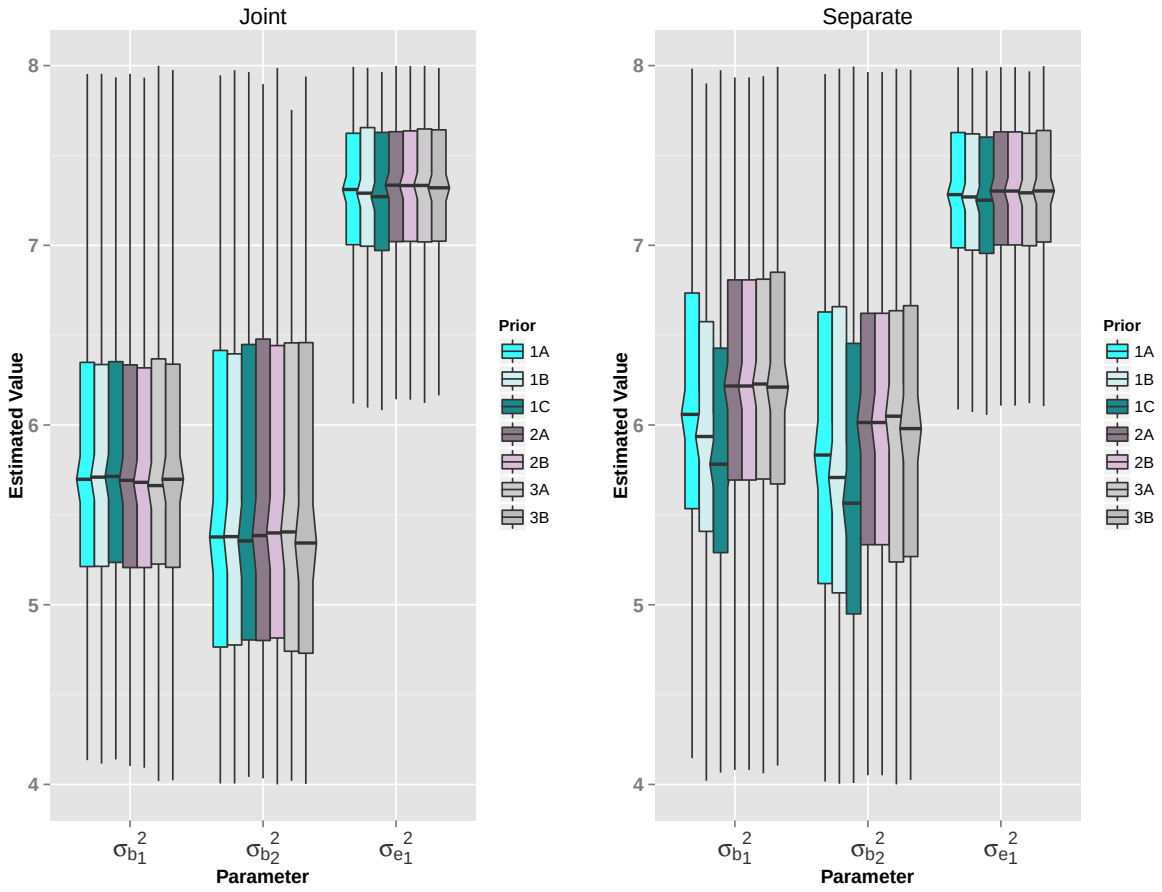
True Model (n)	Parameter	Truth	Fitted Model: Joint-Correlated (JC) Bias	JC SD	JC CP	Joint-Shared (JS) Bias	JS SD	JS CP	$RE_1 = \frac{MSE_{JS}}{MSE_{JC}}$	Separate (SP) Bias	SP SD	SP CP	$RE_2 = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=50)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.066	0.589	0.94	-0.107	0.568	0.93	1.03	-0.075	0.603	0.94	0.98
	β_{11}: time	-0.56	-0.004	0.072	0.95	-0.001	0.075	0.95	1.23	-0.004	0.072	0.95	1.01
	β_{12}: treatment	-0.50	0.048	0.838	0.94	0.138	0.809	0.95	1.02	0.082	0.850	0.95	0.99
	β_{13}: time×treatment	0.30	0.005	0.101	0.95	0.002	0.106	0.96	1.21	0.005	0.102	0.95	1.01
	Ordinal Process												
	α_2: threshold	1.25	0.029	0.165	0.94	-0.079	0.159	0.93	1.29	0.048	0.166	0.94	1.09
	β_{20}: intercept	1.80	0.031	0.575	0.95	-0.193	0.524	0.93	1.02	0.085	0.624	0.95	1.05
	β_{21}: time	-0.35	-0.006	0.075	0.94	0.024	0.068	0.94	1.07	-0.011	0.076	0.94	1.11
	β_{22}: treatment	-0.50	-0.051	0.790	0.95	0.104	0.733	0.95	0.87	-0.091	0.852	0.96	1.06
	β_{23}: time×treatment	0.10	0.000	0.101	0.94	-0.008	0.090	0.95	0.97	0.002	0.103	0.96	1.12
	Association Parameters & Variances												
	ρ	0.60	0.033	0.122	0.93	-	-	-	-	-	-	-	-
	γ	0.91	-	-	-	0.133	0.232	0.93	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.229	1.466	0.94	-1.511	1.334	0.73	4.39	0.189	1.579	0.94	1.05
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	0.292	1.782	0.94	-	-	-	-	1.092	2.119	0.90	1.88
	$\sigma^2_{e_1}$: error	7.40	0.153	0.661	0.95	1.585	0.858	0.68	7.55	0.093	0.649	0.96	0.94
Correlated (n=100)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.029	0.420	0.95	-0.013	0.394	0.95	1.01	-0.046	0.426	0.95	1.01
	β_{11}: time	-0.56	0.001	0.050	0.95	0.001	0.052	0.96	1.12	0.001	0.050	0.94	1.02
	β_{12}: treatment	-0.50	0.037	0.597	0.95	0.037	0.558	0.95	1.02	0.047	0.605	0.95	1.01
	β_{13}: time×treatment	0.30	-0.002	0.070	0.95	-0.001	0.073	0.94	1.16	-0.003	0.071	0.96	1.01
	Ordinal Process												
	α_2: threshold	1.25	0.004	0.115	0.94	-0.124	0.110	0.81	2.35	0.019	0.119	0.94	1.06
	β_{20}: intercept	1.80	0.023	0.407	0.94	-0.163	0.349	0.91	1.12	0.017	0.405	0.95	1.05
	β_{21}: time	-0.35	-0.001	0.052	0.95	0.036	0.048	0.91	1.45	-0.004	0.053	0.94	1.06
	β_{22}: treatment	-0.50	-0.035	0.567	0.95	0.045	0.472	0.95	0.84	0.021	0.553	0.95	1.04
	β_{23}: time×treatment	0.10	-0.003	0.069	0.94	-0.016	0.063	0.94	1.02	-0.006	0.071	0.95	1.06
	Association Parameters & Variances												
	ρ	0.60	0.008	0.088	0.94	-	-	-	-	-	-	-	-
	γ	0.91	-	-	-	0.033	0.142	0.95	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.085	1.032	0.96	-1.513	0.906	0.57	8.11	0.133	1.081	0.95	1.08
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	0.162	1.192	0.95	-	-	-	-	0.510	1.296	0.92	1.42
	$\sigma^2_{e_1}$: error	7.40	0.049	0.451	0.96	1.536	0.612	0.45	13.94	0.020	0.446	0.96	0.98

Table 3: Results when data were uncorrelated ($\rho = 0.0$): SD and CP, stand for posterior standard deviation and coverage probabilities of the 95% HPD intervals, respectively.

True Model (n)	Parameter	Truth	Fitted Model: Joint-Correlated (JC) Bias	JC SD	JC CP	Joint-Shared (JS) Bias	JS SD	JS CP	$RE_1 = \frac{MSE_{JS}}{MSE_{JC}}$	Separate (SP) Bias	SP SD	SP CP	$RE_2 = \frac{MSE_{SP}}{MSE_{JC}}$
Uncorrelated (n=50)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.041	0.586	0.94	-0.032	0.530	0.96	1.04	-0.049	0.599	0.94	0.99
	β_{11}: time	-0.56	0.001	0.071	0.96	0.004	0.079	0.95	1.51	0.002	0.071	0.96	1.00
	β_{12}: treatment	-0.50	0.006	0.831	0.94	0.013	0.749	0.95	1.04	0.037	0.847	0.94	0.99
	β_{13}: time×treatment	0.30	0.002	0.101	0.95	0.001	0.111	0.95	1.51	0.001	0.101	0.95	1.00
	Ordinal Process												
	α_2: threshold	1.25	0.021	0.164	0.96	-0.183	0.164	0.92	3.62	0.038	0.165	0.95	1.06
	β_{20}: intercept	1.80	0.079	0.579	0.95	-0.248	0.487	0.94	1.24	0.115	0.628	0.95	1.13
	β_{21}: time	-0.35	-0.019	0.077	0.94	0.038	0.071	0.95	1.67	-0.023	0.077	0.93	1.07
	β_{22}: treatment	-0.50	-0.066	0.786	0.95	0.063	0.654	0.95	0.72	-0.113	0.858	0.95	1.11
	β_{23}: time×treatment	0.10	0.013	0.102	0.95	0.000	0.088	0.93	0.88	0.015	0.103	0.95	1.06
	Association Parameters & Variances												
	ρ	0.00	-0.002	0.175	0.95	-	-	-	-	-	-	-	-
	γ	0.91	-	-	-	-0.668	3.144	0.96	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.225	1.468	0.93	-2.941	1.278	0.64	19.04	0.176	1.575	0.94	1.04
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	0.388	1.827	0.93	-	-	-	-	1.199	2.162	0.90	1.91
	$\sigma^2_{e_1}$: error	7.40	0.096	0.650	0.95	2.858	1.204	0.73	32.26	0.061	0.643	0.96	0.97
Uncorrelated (n=100)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.010	0.419	0.95	0.007	0.398	0.96	1.01	-0.018	0.426	0.95	1.00
	β_{11}: time	-0.56	-0.001	0.050	0.96	-0.002	0.053	0.95	1.24	-0.001	0.050	0.96	1.00
	β_{12}: treatment	-0.50	-0.004	0.595	0.96	-0.018	0.568	0.96	1.00	-0.005	0.606	0.95	0.99
	β_{13}: time×treatment	0.30	0.003	0.071	0.95	0.004	0.075	0.94	1.17	0.002	0.072	0.94	1.00
	Ordinal Process												
	α_2: threshold	1.25	0.005	0.116	0.94	-0.377	0.096	0.44	15.17	0.018	0.119	0.95	1.04
	β_{20}: intercept	1.80	0.039	0.416	0.94	-0.520	0.262	0.77	3.57	0.022	0.406	0.95	1.00
	β_{21}: time	-0.35	-0.007	0.053	0.95	0.101	0.045	0.76	5.77	-0.009	0.053	0.96	1.03
	β_{22}: treatment	-0.50	-0.049	0.568	0.94	0.121	0.335	0.93	0.60	0.021	0.555	0.94	1.01
	β_{23}: time×treatment	0.10	0.006	0.071	0.94	-0.025	0.056	0.93	1.03	0.002	0.071	0.94	1.01
	Association Parameters & Variances												
	ρ	0.00	-0.001	0.124	0.95	-	-	-	-	-	-	-	-
	γ	0.91	-	-	-	-1.130	0.980	0.89	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.097	1.036	0.95	-1.258	1.072	0.80	16.68	0.121	1.080	0.94	1.05
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	0.200	1.214	0.94	-	-	-	-	0.548	1.304	0.91	1.41
	$\sigma^2_{e_1}$: error	7.40	0.073	0.452	0.95	1.322	0.717	0.81	28.80	0.053	0.449	0.95	0.98

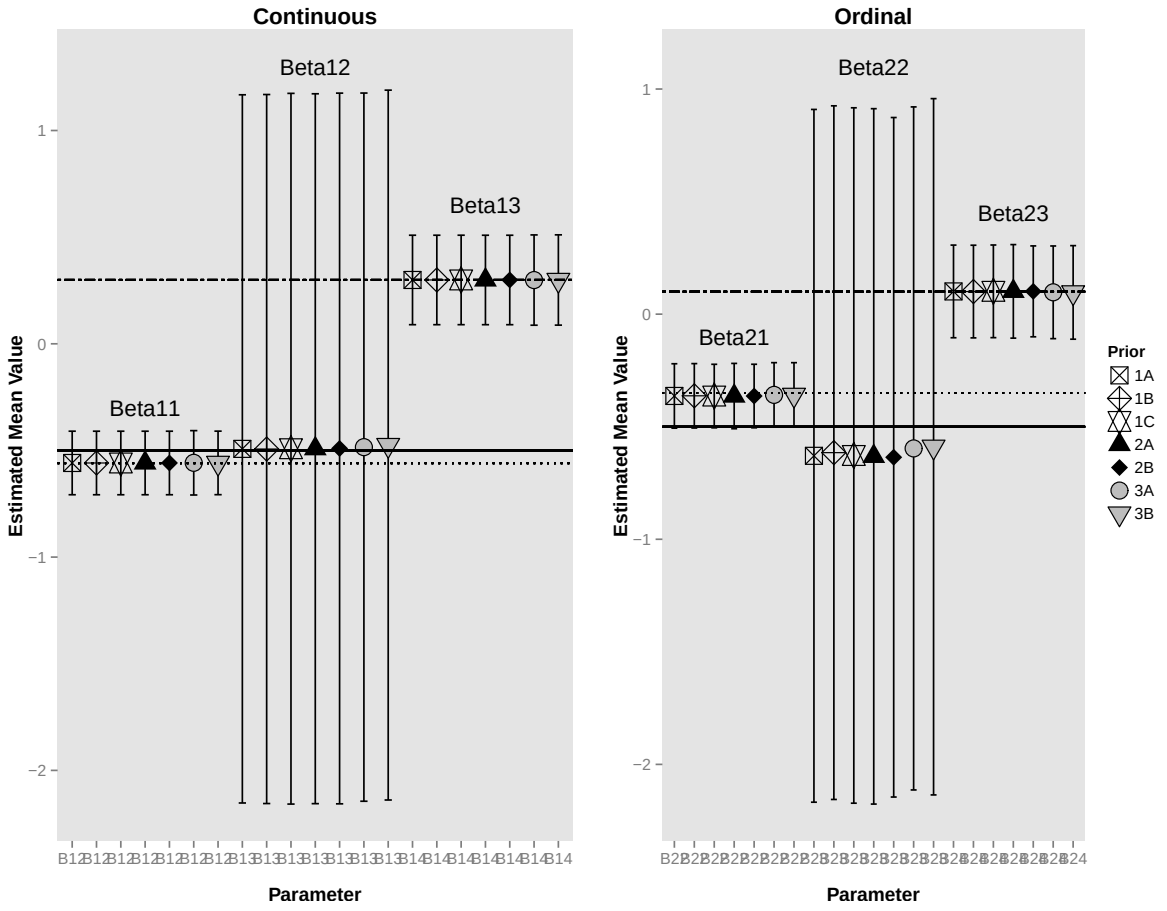
The results in Tables 1-3 further indicated very high efficiency gain in variances, which could have been due to the choice of prior distributions assigned to them. To determine their (i.e. the prior distribution for variance parameters) effect on the the main effects regression parameters, we performed a sensitivity analysis. The priors considered were Gamma(.001,.001), Gamma(0.5,0.5), Gamma(1,1), Pareto(0.5,.0001), and Pareto(0.5,0.01) for the precision, and half-Cauchy(s=25) and half-Cauchy(s=20) for the standard deviations. We simulated 200 data sets of size $n = 50$ from a joint model correlated through correlated random effects but with $\rho = 0.0$ and then fitted a joint and separate models to the data generated. We chose $\rho = 0.0$ for better comparison of the performance of the priors through relative efficiency. Evidence of no gain in efficiency of the estimates from the joint and separate models would imply a more appropriate prior distribution. We employed the models used in the main simulation study to simulate the data. In fitting the models, we also maintained same prior distributions for the regression parameters β. For the joint models, Inverse Wishart prior was employed for the variance-covariance matrix Γ of the random effects while we varied the priors for the error variance. For the separate models, we varied the priors for the error variance and random effects variances.

The results are summarized in Figures 1, 2, and 3. The variance estimates differed when we varied the hyper-parameters for Gamma and half-Cauchy but not Pareto (Figure 1). The effect was more pronounced in the estimates of random effects variances and when we varied hyper-parameters of Gamma prior (right panel of Figure 1). However, the regression coefficient estimates remained similar under all the above prior distributions (Figures 2 & 3).



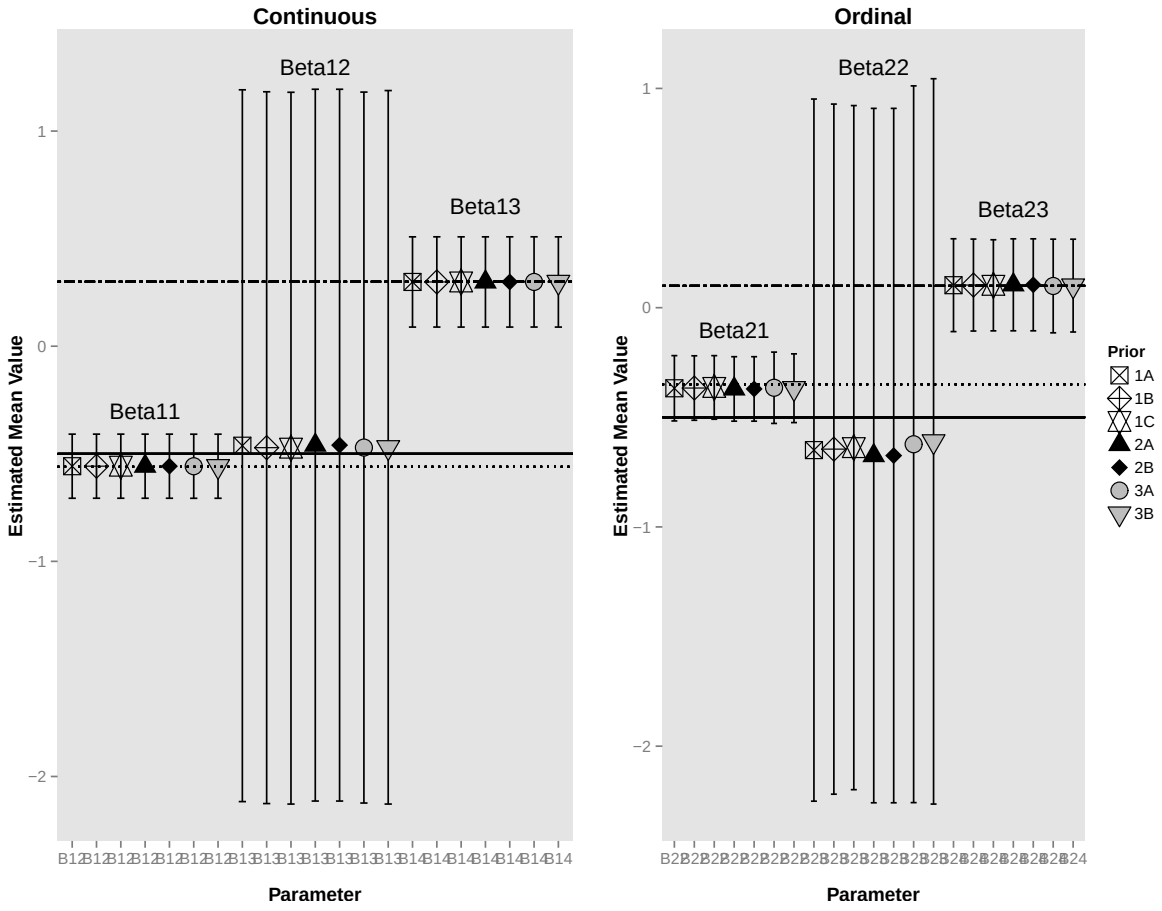
The priors 1A, 1B, 1C, 2A, 2B, 3A, 3B, stand for Gamma(.001,.001), Gamma(0.5,0.5), Gamma(1,1), Pareto(0.5,.0001), Pareto(0.5,0.01), half-Cauchy(s=25), half-Cauchy(s=20), respectively. In the joint model (left panel) where we varied the prior for the error variance keeping the prior for the precision matrix of random effects as Wishart$(3, 1I_2)$, the estimates are quite similar to each other. Meanwhile, in the separate models (right panel) where we varied the priors for the error and random effects variances, the variance estimates for random effects varied when we varied the hyper-parameters for Gamma prior (1A, 1B, 1C).

Figure 1: Box plots of estimates for the error variance, $\sigma^2_{e_1}$, and random effects variances, $\sigma^2_{b_1}$ and $\sigma^2_{b_2}$, under different priors.



The effect of error variance prior on the regression parameter estimates. The horizontal dotted, solid, and dashed lines represent the true values for Beta11/Beta21, Beta12/Beta22, and Beta13/Beta23, respectively. The priors 1A, 1B, 1C, 2A, 2B, 3A, 3B, stand for Gamma(.001,.001), Gamma(0.5,0.5), Gamma(1,1), Pareto(0.5,.0001), Pareto(0.5,0.01), half-Cauchy(s=25), half-Cauchy(s=20), respectively. The regression parameter estimates are similar across different priors.

Figure 2: Mean estimates and 95% credible intervals for the Joint Model.



The effect of error and random effects variance priors on the regression parameter estimates. The horizontal dotted, solid, and dashed lines represent the true values for Beta11/Beta21, Beta12/Beta22, and Beta13/Beta23, respectively. The priors 1A, 1B, 1C, 2A, 2B, 3A, 3B, stand for Gamma(.001,.001), Gamma(0.5,0.5), Gamma(1,1), Pareto(0.5,.0001), Pareto(0.5,0.01), half-Cauchy(s=25), half-Cauchy(s=20), respectively. The regression parameter estimates are similar across different priors.

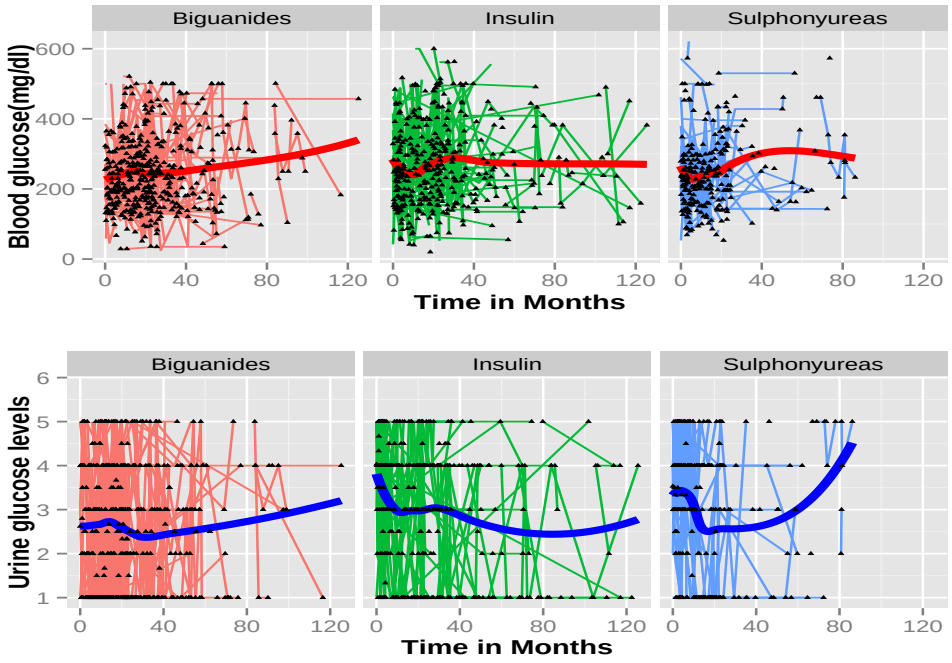
Figure 3: Mean estimates and 95% credible intervals for the Separate Models.

3.6 ANALYSIS OF UGANDAN DIABETES DATA

In this section, we present the analysis of the diabetes data introduced in Section 3.1. Our interest is to jointly model the continuous and ordinal measures of disease severity. We considered diabetes data that were collected retrospectively from three hospitals (Mulago, Nsambya, and Rubaga all in Kampala, Uganda). The data were monthly records of 321 diabetic patients from the medical registries in the three hospitals (for details see Buhule et al., 2007 [8]) who had at least two measurements of both blood and urine glucose and taken on the same occasions. The covariates of interest included treatment (Biguanides, Sulphonyureas, and Insulin (baseline)), baseline age in years, gender (male=1, female=0), time of hospital visits in months, and time and treatment interaction. The profile plots of the 321 individuals grouped by treatment are shown in Figure 4. We observed that both blood glucose and urine glucose levels tended to increase over time for individuals who were treated with Biguanides at baseline. For those on Insulin, blood glucose levels were fairly constant over time while the urine glucose showed slight decline. Lastly for individuals on Sulphonyureas, blood glucose seemed to remain constant but urine glucose showed a U-shape trend based on the smooth lowess plot. The U-shape trend might be due to outliers or sparsity of data at large time values. The summary statistics for baseline characteristics of the 321 patients are also indicated in Table 4. Most of the patients were Type 2 diabetics because it is the main form of diabetes. They were all adults although Type 1 diabetics were younger on average. Their BMI levels were about 29 on average indicating majority of these patients were overweight at baseline. The majority of the patients were female, and were mostly treated with Insulin at baseline; however, considering type of diabetes, Sulphonyureas treatment was only given to Type 2 diabetics.

The two biomakers (blood glucose and urine glucose) were modeled jointly through correlated random effects and shared random effects and then compared to separate models. To improve normality, we used square root transformation of blood glucose levels which were linked to the linear predictor (covariates) with an identity link. That is,

$$y_{ij,1} = \beta_{10} + \beta_{11} \times time_{ij} + \beta_{12} \times Biguanides_{ij} + \beta_{13} \times Sulphonyureas_{ij} + \beta_{14} \times age_i + \beta_{15} \times gender_i + \beta_{16} \times Biguanides_{ij} \times time_{ij} + \beta_{17} \times Sulphonyureas_{ij} \times time_{ij} + \mathbf{b}_{i1} + \epsilon_{ij,1},$$



The distribution of blood glucose (upper panel) and urine glucose (lower panel) by treatment. The thick lines represent the lowess smooth curves.

Figure 4: Distribution of Blood/Urine glucose by treatment over time.

where, $y_{ij,1}$ is the square root of blood glucose (mg/dl) for the i^{th} subject measured at the j^{th} occasion/hospital visit, $\mathbf{b}_{i1}$ is the random intercept and $\epsilon_{ij,1} \sim N(0, \sigma^2_{e_1})$ is the measurement error independent of $\mathbf{b}_{i1}$.

The marginal probabilities of urine glucose levels were linked to the covariates through a cumulative logit link as follows:

$$\log\left(\frac{Pr(y_{ij,2} \leq k)}{1 - Pr(y_{ij,2} \leq k)}\right) = \alpha_k - [\beta_{21} \times time_{ij} + \beta_{22} \times Biguanides_{ij} + \beta_{23} \times Sulphonyureas_{ij} + \beta_{24} \times age_i + \beta_{25} \times gender_i + \beta_{26} \times Biguanides_{ij} \times time_{ij} + \beta_{27} \times Sulphonyureas_{ij} \times time_{ij} + \mathbf{b}_{i2}],$$

where $k = 1, \ldots, 5$ and $\mathbf{b}_{i2}$ is the random intercept. In this cumulative logistic model, a

Table 4: Descriptive statistics for baseline characteristics

Variable	All $n = 321$	Type 1 $n = 69$	Type 2 $n = 252$
Age (years), $\bar{x}$ (s)	49.7 (13.1)	37.4 (12.6)	53.1 (11.1)
BMI (kg/m^2), $\bar{x}$ (s)	28.6 (5.9)	28.9 (5.2)	28.5 (6.1)
Gender			
Male, n (%)	70 (21.8)	16 (23.2)	54 (21.4)
Female, n (%)	251 (78.2)	53 (76.8)	198 (78.6)
Treatment			
Biguanides, n (%)	116 (36.1)	20 (29.0)	96 (38.1)
Sulphonyureas, n (%)	85 (26.5)	0 (0.0)	85 (33.7)
Insulin, n (%)	120 (37.4)	49 (71.0)	71 (28.2)

positive regression coefficient indicates a higher probability of being in higher category of urine glucose level.

For all parameters, we assumed similar priors as those used in the simulation study. Time was standardized while age was centered to improve estimation. The MCMC was run for $30,000$ iterations with the first $5,000$ discarded as burn-in. The models were fitted in OpenBUGS (version 3.2.2) and its R interface BRugs Version 0.8.3. The standard MCMC diagnostic tests (Table 5) indicated that all parameters estimated converged according to the Gelman and Rubin diagnostics test [26] and its multivariate test proposed by Brook and Gelman[5]. The few parameters that marginally failed the Geweke[30] and the Heidelberg and Welch[41] diagnostic tests could be improved by running longer chains. The diagnostic plots further indicated proper mixing of the chains and convergence to the stationary distributions (see Appendix D Figures 5-7).

Table 5: Convergence Diagnostic tests results for analysis of Diabetes data

	Joint-Correlated (JC)				Joint-Shared (JS)				Separate (SP)			
	Gelman	Geweke	Heidelberg		Gelman	Geweke	Heidelberg		Gelman	Geweke	Heidelberg	
Parameter	Estimate	Z-score	Stationarity	Halfwidth	Estimate	Z-score	Stationarity	Halfwidth	Estimate	Z-score	Stationarity	Halfwidth
Continuous Process: blood glucose levels												
Intercept	1.00	1.137	passed	passed	1.00	2.286	passed	passed	1.00	-1.440	passed	passed
Time	1.00	1.348	passed	passed	1.00	-1.333	passed	passed	1.00	-0.315	passed	passed
Treatment												
Biguanides	1.00	-1.925	passed	passed	1.00	-1.265	passed	passed	1.00	-0.416	passed	passed
Sulphonyureas	1.00	-1.211	passed	passed	1.00	-1.697	passed	passed	1.00	0.668	passed	passed
Age	1.00	1.247	passed	failed	1.00	-0.144	passed	failed	1.00	-1.778	passed	failed
Male	1.00	-1.188	passed	passed	1.00	-1.386	passed	passed	1.00	0.699	passed	passed
Treatment×Time												
Biguanides×Time	1.00	0.186	passed	failed	1.00	1.960	passed	failed	1.00	-0.924	passed	passed
Sulphonyureas×Time	1.00	-0.294	passed	passed	1.00	-0.031	passed	passed	1.00	0.247	passed	passed
Ordinal Process: urine glucose categories												
Threshold-1	1.01	-0.971	passed	passed	1.00	-1.799	passed	passed	1.00	-0.494	passed	passed
Threshold-2	1.01	-1.016	passed	passed	1.00	-1.702	passed	passed	1.00	-0.531	passed	passed
Threshold-3	1.01	-1.088	passed	passed	1.00	-1.606	passed	passed	1.00	-0.565	passed	passed
Threshold-4	1.00	-1.319	passed	passed	1.00	-1.396	passed	passed	1.00	-0.673	passed	passed
Time	1.00	-0.465	passed	passed	1.00	-0.622	passed	passed	1.00	1.774	passed	passed
Treatment												
Biguanides	1.00	-1.814	passed	passed	1.00	-0.712	passed	passed	1.00	-1.798	passed	passed
Sulphonyureas	1.00	-0.977	passed	passed	1.00	-1.282	passed	failed	1.00	-0.932	passed	passed
Age	1.00	1.677	passed	passed	1.00	0.134	passed	passed	1.00	0.976	passed	passed
Male	1.00	-2.381	passed	passed	1.00	-1.471	passed	passed	1.01	0.958	passed	passed
Treatment×Time												
Biguanides×Time	1.00	1.787	passed	passed	1.00	-0.042	passed	passed	1.00	-1.174	passed	passed
Sulphonyureas×Time	1.00	0.575	passed	passed	1.01	0.315	passed	passed	1.00	-2.740	passed	passed
Association Parameters & Variances												
ρ	1.00	1.124	passed	passed	-	-	-	-	-	-	-	-
γ	-	-	-	-	1.00	1.016	passed	passed	-	-	-	-
$\sigma^2_{b_1}$	1.00	-1.264	passed	passed	1.00	-1.287	passed	passed	1.00	0.290	passed	passed
$\sigma^2_{b_2}$	1.00	-0.297	passed	passed	-	-	-	-	1.00	-0.973	passed	passed
$\sigma^2_{e_1}$	1.00	1.472	passed	passed	1.00	1.063	passed	passed	1.00	-1.138	passed	passed
Multivariate Test	1.01	-	-	-	1.01	-	-	-	1.01	-	-	-

The posterior estimates of the regression coefficients and their 95% credible intervals (CI) for the joint (correlated random effects or shared random effects) versus separate analyses are summarized in Table 6. The point estimates from the separate and joint analyses were quite similar but the CIs somewhat differed. Time and gender were found to be significantly associated with urine glucose levels in all the three models, such that the urine glucose levels declined over time, and the males tended to have higher urine glucose levels compared to females. In addition, gender was significantly associated with blood glucose levels in the shared random effects model, implying, male patients had lower blood glucose levels as compared to female patients. Although Biguanides treatment was found to be significantly associated with urine glucose levels in the joint correlated model and separate models, the interaction between Biguanides and time was not statistically significant in these two models. However, there was a statistically significant interaction effect between Biguanides and time in the shared random effects model; indicating patients who were treated with Biguanides at baseline had higher urine glucose levels over time as compared to those treated with Insulin, which is also evident in the profile plot (Figure 4). Moreover, the posterior estimates of the association parameters ρ and γ in the joint analyses were positive and significantly different from zero, providing strong evidence of association between the blood glucose and urine glucose sub-models and indicating that the initial level of blood glucose was positively associated with the urine glucose levels. The credible intervals shrunk in the joint models with the shared random effects model having more shrinkage. All the measures of fit , that is, DIC, LPML, and ALPML showed that the joint correlated random effects model fitted the data better than the shared random effects model. Although DIC indicated that the joint correlated random effects model fitted the data better than the separate models, the LPML and ALPML measures showed no significant difference between the joint correlated and separate models.

We also analyzed only type 2 diabetics data and the results (see Appendix A: Table 22) were not very different those from the combined (type 1 and type 2 diabetics) data in Table 6. Moreover, the convergence diagnostic results (see Appendix A: Table 21 and Appendix D: Figures 8-10) indicated proper mixing of the chains and/or convergence of parameters to their target distributions. Because age was not significantly associated with any of the two

Table 6: Analysis of Diabetes Data

Parameter	Joint-Correlated Random Effects		Joint-Shared Random Effects		Separate	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
Continuous Process: blood glucose levels						
Intercept	15.68	(15.20, 16.17)	15.86	(15.43, 16.28)	15.66	(15.15, 16.16)
Time	-0.10	(-0.42, 0.22)	0.18	(-0.12, 0.48)	-0.20	(-0.53, 0.12)
Treatment						
Biguanides	-0.15	(-0.71, 0.41)	-0.35	(-0.86, 0.16)	-0.15	(-0.74, 0.43)
Sulphonyureas	-0.33	(-0.96, 0.29)	-0.48	(-1.05, 0.09)	-0.23	(-0.88, 0.42)
Age in years	0.00	(-0.02, 0.02)	0.00	(-0.02, 0.02)	0.00	(-0.03, 0.02)
Male	-0.75	(-1.53, 0.03)	-1.11	(-1.76, -0.45)	-0.64	(-1.43, 0.15)
Treatment×Time						
Biguanides×Time	-0.01	(-0.44, 0.44)	-0.02	(-0.46, 0.41)	0.07	(-0.39, 0.52)
Sulphonyureas×Time	0.22	(-0.35, 0.79)	0.19	(-0.35, 0.73)	0.23	(-0.36, 0.81)
Ordinal Process: urine glucose categories						
Threshold-1	-0.90	(-1.28, -0.54)	-0.81	(-1.16, -0.46)	-1.02	(-1.41, -0.66)
Threshold-2	-0.42	(-0.79, -0.07)	-0.35	(-0.69, -0.00)	-0.54	(-0.93, -0.18)
Threshold-3	0.46	(0.09, 0.82)	0.49	(0.15, 0.84)	0.34	(-0.05, 0.70)
Threshold-4	2.64	(2.24, 3.04)	2.57	(2.20, 2.95)	2.53	(2.13, 2.93)
Time	-0.36	(-0.59, -0.14)	-0.40	(-0.62, -0.19)	-0.25	(-0.49, -0.01)
Treatment						
Biguanides	-0.43	(-0.83, -0.04)	-0.37	(-0.74, 0.01)	-0.59	(-0.98, -0.19)
Sulphonyureas	-0.20	(-0.65, 0.23)	-0.18	(-0.58, 0.26)	-0.26	(-0.73, 0.19)
Age in years	-0.01	(-0.03, 0.01)	-0.01	(-0.03, 0.01)	-0.01	(-0.03,0.01)
Male	0.90	(0.33, 1.48)	1.00	(0.45, 1.56)	0.78	(0.18, 1.36)
Treatment×Time						
Biguanides×Time	0.16	(-0.16, 0.48)	0.16	(0.14, 0.46)	0.15	(-0.18, 0.49)
Sulphonyureas×Time	0.30	(-0.10, 0.68)	0.28	(-0.10, 0.65)	0.29	(-0.13, 0.71)
Association Parameters & Variances						
ρ	0.61	(0.49, 0.72)	-	-	-	-
γ	-	-	0.98	(0.75, 1.23)	-	-
$\sigma^2_{b_1}$	5.86	(4.60, 7.34)	3.05	(2.05, 4.25)	5.99	(4.68, 7.52)
$\sigma^2_{b_2}$	3.17	(2.35, 4.17)	-	-	3.11	(2.30, 4.09)
$\sigma^2_{e_1}$	8.02	(7.34, 8.74)	10.15	(9.23, 11.13)	7.98	(7.30, 8.70)
Goodness of Fit						
Outcome	DIC	LPML (ALPML)	DIC	LPML (ALPML)	DIC	LPML (ALPML)
Continuous	6913	-2098 (-1.539)	7109	-2259 (-1.657)	6930	-2096 (-1.538)
Ordinal	3432	-1604 (-1.177)	3482	-1649 (-1.210)	3455	-1605 (-1.178)
Total	10350	-3702 (-2.716)	10590	-3908 (-2.867)	-	-

outcomes in Table 6 results, we excluded it from this second analysis but instead included BMI. Sulphonyureas treatment and gender were significantly associated with blood glucose levels in the shared random effects model. Patients who were treated with Sulphonyureas at baseline had their blood glucose levels decrease as compared to those treated with Insulin. Male patients had lower blood glucose levels compared to their female counterparts. On the other hand, Biguanides treatment was found to be significantly associated with urine glucose levels in all the three models, and gender and time were significant in the joint models (correlated and shared). The patients who were treated with Biguanides had lower urine glucose levels compared to those treated with Insulin, and males tended to have higher urine glucose levels compared to females. Overall, these patients had lower urine glucose levels over time. Again,the posterior estimates of the association parameters ρ and γ in the joint analyses were positive and significantly different from zero, providing strong evidence of association between the blood glucose and urine glucose sub-models and indicating that the initial level of blood glucose was positively associated with the urine glucose levels. The DIC also indicated that the joint correlated random effects model fitted the data better than the shared random effects model and the separate models.

3.7 DISCUSSION

In this dissertation we have proposed a full Bayesian hierarchical multivariate generalized linear mixed effects model for multiple repeatedly measured mixed outcomes (continuous and ordinal) that are measures of disease severity. The key features of our model are the Bayesian hierarchical formulation for modeling the subject-specific random effects when data are highly unbalanced. We employed the nested indexing approach of handling unbalanced data (Lunn, Thomas, and Spiegelhalter, 2000) [65] in formulating the proposed model. Furthermore, the use of Bayesian approach that is fully implemented in freely available software like WinBUGS (Lunn et al., 2000) [65] and OpenBUGS (Lunn et al., 2009)[64] avoids the difficulties of routinely implementing maximum likelihood-based methods to these complex and useful models. Currently, probit models are very widely used in complex

applications involving categorical ordinal data due to ease of modeling and computation using the underlying normal framework. Our approach uses the cumulative logit model which is more stable especially in handling outlying data and easier to interpret than the probit models, particularly for biomedical researchers who routinely use logistic regression in analyses of binary outcomes. We have considered the cumulative logit model (proportional odds model) but this proportional odds assumption can be relaxed by fitting a non-proportional or partial proportional cumulative logit model as proposed by Hedeker and Mermelstein (1998)[40]. This extension allows the covariates to vary across the cumulative logits or cut points. As it is straight forward to generalize our approach to essentially any data structure, the proposed methodology would be a useful toolkit in the statistician's toolbox.

The simulation study results demonstrated that joint modeling leads to efficient estimates and adequate 95% coverage probabilities for the population parameters. The efficiency gain was larger for the ordinal outcome estimates compared to that for the continuous outcome. Overall, ordinal outcome regression coefficient estimates gained more efficiency when the the joint model was correlated through correlated random effects than through shared random effects. The efficiency gain was justified as modeling of mixed outcome types using shared random effects is disadvantages to outcomes having non-normal distributions especially those that have a natural tie between the mean and variance (McCulloch, 2008)[69]. In general, our results agreed with those of Guerguieva and Sanacora (2006)[33] who found more efficient gain in the ordinal outcome than the continuous outcome when they jointly modeled balanced longitudinal data. Moreover, our real data example indicated improved efficiency when blood glucose and urine glucose were modeled jointly. Finally, varying priors for variance parameters was found to have no effect on the parameters of interest, although differences in the estimates of the variance parameters were observed especially when we varied the hyper-parameters for Gamma priors as indicated by Gelman (2006) [25].

In the next chapter (Chapter 4) we extend the proposed joint model to include time-to-event data.

4.0 BAYESIAN HIERARCHICAL JOINT MODELING OF REPEATEDLY MEASURED MIXED BIOMARKERS OF DISEASE SEVERITY AND TIME-TO-EVENT

4.1 INTRODUCTION

Joint modeling has been widely used in simultaneous modeling of longitudinal outcomes and time-to-event data. When repeatedly measured markers over time are used as time-dependent covariates in survival analysis, they are likely to lead to biased estimates. Bias occurs because these markers are prone to measurement errors and have increased within patient variability due to biological fluctuations. Modeling these markers and time-to-event data jointly leads to unbiased and more efficient estimates. The goal of joint modeling in this context may include: modeling the distribution of the time to a terminal event conditional on longitudinal measurement sequence, adjusting inference about a longitudinal measurement sequence to allow for informative dropout or joint evaluation of a measurement and an event-time process. Most of the work in this field has focused on joint modeling of a single longitudinal outcome and time-to-event data. However, in clinical studies where associations between the event process and more than one biomarker are of interest, joint modeling of all the markers and the event process is likely to increase the efficiency of the estimates.

This study is motivated by data for diabetic patients that were collected retrospectively from three Ugandan hospitals. These patients attended the diabetic clinics between January 1992 and December 2004 during which several clinical measurements were taken. Blood glucose in mg/dl was taken as the main biomaker reflecting disease severity. In addition, urine glucose levels were taken as a compliment or substitute to blood glucose because the urine glucose test is cheaper and faster. Other variables that are known to be associated with

Type 2 diabetes like body mass index (BMI) and blood pressure in mm Hg (both systolic and diastolic) were also collected often during hospital visits. These patients were mostly treated with Sulphonyureas or Biguanides or Insulin at baseline (the first time they visited the diabetes clinic). Because diabetes is a chronic illness with no known cure, these treatments are only used to control and prevent the development of diabetes complications. Thus, they slow down diabetes progression by reducing the rate of further injury to the biological system without necessarily improving the current level of functioning.

The normal range for blood glucose is between 70 to 130 mg/dl or 70 to 180 mg/dl depending on whether someone is fasting or not at the time of testing. The clinical interest is to ensure the blood glucose is in the normal range. In addition, maintaining normal blood pressure and BMI for diabetics is known to reduce the risk of mortality from diabetes compilations such as heart attack and stroke. Of the three biomarkers (i.e., blood glucose, blood pressure, and BMI), blood glucose is expensive to measure and yet it is the main biomarker of disease severity for diabetes. Normalization of the blood glucose levels is one important objective for all diabetic patients. However, since blood glucose levels are more expensive to measure it may be reasonable to establish the association between normalization of it and other biomarkers such as blood pressure levels and BMI.

Joint models that combine the longitudinal and time-to-event processes have been widely studied by many authors. Hogan and Laird (1997a)[44], Tsiatis and Davidian (2004)[93], and Ibrahim, Chen, and Sinha (2001, Chapter 7)[51] give a detailed discussion of joint modeling. Pawitan and Self (1993)[73], DeGruttola and Tu (1994)[16], Tsiatis, DeGruttola, and Wulfsohn (1995)[94], Faucett and Thomas (1996)[18], Lavalley and De Gruttola (1996)[58], Wulfsohn and Tsiatis (1997)[100], Henderson, Diggle, and Dobson (2000)[42], Xu and Zeger (2001a)[101], Tsiatis and Davidian (2001)[92], Wang and Taylor (2001)[97], Guo and Carlin (2004)[34], Brown and Ibrahim (2003) [6, 7], Ibrahim, Chu, and Chen(2010)[52], Wang, Shen, and Boye (2012)[96], Huang, Hu, and Dagne (2014)[49] all have worked on one longitudinal outcome and time-to-event. While Rizopoulos and Ghosh (2010)[79] and Hatfield, Boye, and Carlin (2011)[36] extended the longitudinal outcome to multivariate case.

In this literature, the models employed for each outcome included mainly a proportional hazards model (semi-parametric or parametric) for the survival times and a linear mixed

effects model for the longitudinal measurements. The two outcomes were jointly modeled through partial likelihood models or joint likelihood models. In the partial likelihood models, the longitudinal measurements are taken to be time-dependent covariates in the hazard function of survival times. This method also known as two-stage approach was used by Tsiatis, DeGruttola, and Wulfsohn (1995) [94], who first estimated the true longitudinal measurements at each event time by method of moments and then plugged the fitted values into the Cox's partial likelihood before maximizing it to obtain the estimates of the regression parameters. This method does not make use of all the information and is known to lead to biased estimates because covariate values are often measured with error (Prentice, 1982)[76]. In the joint likelihood models approach, on the other hand, the model for survival is conditioned on the observed longitudinal covariate or the other way around, depending on whether the interest is in the survival or longitudinal outcome, respectively. The random effects model (Laird and Ware, 1982 [57]) is often used to model the longitudinal outcome and the individual random effects are included in the survival model. This leads to more efficient estimates than the two-stage approach because it uses the full likelihood in estimation and hence makes more use of the data. Parameter estimation is carried out using maximum likelihood (ML) methods like EM algorithms or Bayesian Markov Chain Monte Carlo (MCMC) methods; however, ML methods are computationally very intensive.

Faucett and Thomas (1996) [18], Xu and Zeger (2001a) [101], DeGruttola and Tu (1994) [16], Henderson, Diggle, and Dobson (2000) [42], Wang and Taylor (2001)[97], Guo and Carlin (2004) used the likelihood approach to jointly model the longitudinal marker and time-to-event; however, Faucett and Thomas (1996) [18], Xu and Zeger (2001a) [101], Wang and Taylor (2001)[97], Guo and Carlin (2004) [34] employed the Bayesian MCMC methods for parameter estimation while others implemented EM algorithms. Henderson, Diggle, and Dobson (2000) [42] and Xu and Zeger (2001a) [101] introduced a stationary Gaussian process (as part of random effects) to allow the longitudinal trajectory to vary with time, while Wang and Taylor (2001)[97] incorporated an integrated Ornstein–Uhlenbeck (IOU) process to monitor the biological fluctuations in the longitudinal process about a smooth trend. Furthermore, Rizopoulos and Ghosh (2010) [79] proposed a Bayesian semiparameteric multivariate joint model that relates multiple longitudinal outcomes (continuous and binary) and time-to-event.

They used a spline-based approach to model the subject-specific longitudinal evolution, and the baseline risk function in the Cox model for time-to-event outcome was assumed piece-wise constant.

Although these joint models are very complex and computationally intensive, ignoring the association between processes leads to inefficient if not biased estimation of the parameters involved. Irrespective of the methods employed, joint modeling in most cases results in efficient and unbiased parameter estimates when compared to separate modeling. As noted earlier, most of the joint models have focused on one longitudinal outcome and time-to-event, and to the best of our knowledge, no one has worked on joint modeling of mixed longitudinal outcomes (e.g. continuous, ordinal) and time-to-event. Therefore, we propose a hierarchical joint model to handle unbalanced repeatedly measured continuous and ordinal markers of disease severity, and time-to-event. We use Bayesian methods to construct the posterior distribution of the parameters of interest. Markov Chain Monte Carlo (MCMC) methods are employed for parameter estimation because they avoid the difficulties of dealing with high-dimensional integrals by sampling from the posterior distribution.

The remainder of this chapter is organized as follows: Section 4.2 presents the formulation of the multivariate generalized linear mixed effects model, the associated joint likelihood, and the prior and posterior distributions. Section 4.3 shows the derivation of the full conditionals and estimation procedures of the parameters from these full conditionals, Section 4.4 shows the simulation study, and Section 4.5 indicates the application of the proposed joint model to diabetes data.

4.2 MODEL SPECIFICATIONS

4.2.1 Model formulation

Let $\mathbf{y}_i = (\mathbf{y}'_{i1}, \mathbf{y}'_{i2}, \ldots, \mathbf{y}'_{iL})'$, denote the L-variate response vector for i^{th} subject $(i = 1, \ldots, n)$, where $\mathbf{y}_{il},\ l = 1, \ldots, L$, is an $n_i \times 1$ vector of longitudinal biomarker for a certain disease severity taken at time points, $j = 1, \ldots, n_i$. In addition, let $T_i = \min(T_i^*, C_i)$ be the failure

time and $\delta_i = I(T_i^* \leq C_i)$ an event indicator which indicates whether the observed failure time is a true failure time, T_i^*, or a censoring time C_i for the i^{th} subject. For example, in the Ugandan Diabetes data set, $\mathbf{y}_{i1}$ is a vector of systolic blood pressure levels, $\mathbf{y}_{i2}$ is a vector of BMI levels, and T_i is time to normalization of blood glucose levels, for the i^{th} patient. Because these responses are assumed to be of mixed types, we employ generalized linear mixed effects models (GLMM) to unify them. Under this framework (GLMM), the conditional distribution of each response is assumed to be a member of exponential family. In particular, the conditional mean μ is linked to the linear predictor η (including fixed effects and random effects) through a known one-to-one monotonic link function $\mathbf{g}(\cdot)$.

Motivated by our data set, we consider two repeatedly measured response variables (continuous and ordinal) and one time-to-event outcome. Thus, $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2})$ for the continuous and ordinal outcomes, respectively. For the continuous outcome, the j^{th} component is linked to the linear predictor through an identity link as in 4.2.11

$$y_{ij,1} = x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i1} + \epsilon_{ij,1}, \tag{4.2.11}$$

where $\epsilon_{i1} \sim N(0, \sigma^2_{e_1})$ is the measurement error, $x_{ij,1}$ and $z_{ij,1}$ are vectors of fixed covariates and random effects, respectively, $\beta_{\mathbf{1}}$ are the regression parameters for the fixed part and $\mathbf{b}_{i1}$ is random effect of the i^{th} subject. The linear predictor or mean is given as

$$\mu_{ij,1} = E\left[y_{ij,1}|\mathbf{b}_{i1}\right] = \eta_{ij,1} = x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i1}.$$

For the ordinal response variable $\mathbf{y}_{i2}$ with K ordered categories $k = 1, 2, \cdots, K$, the conditional cumulative probabilities $p_{ijk,2}$ for the K categories defined as

$$p_{ijk,2} = Pr(y_{ij,2} \leq k) = \sum_{m=1}^{k} p_{ijm,2}, \tag{4.2.12}$$

are linked to the linear predictor $\eta_{ijk,2}$ through a cumulative logit 4.2.13

$$\log\left[\frac{p_{ijk,2}}{1 - p_{ijk,2}}\right] = \eta_{ijk,2} = \alpha_k - \left[x'_{ij,2}\beta_2 + z'_{ij,2}\mathbf{b}_{i2}\right], \tag{4.2.13}$$

with $K - 1$ strictly increasing model thresholds α_k $(i.e., \alpha_1 < \alpha_2 \cdots < \alpha_{K-1})$. In this cumulative logit, we are assuming the proportional odds assumption (McCullagh, 1980) [67], and a positive regression coefficient implies a higher probability of being in higher category.

The first threshold α_1 or the model intercept β_{20} is set to zero to guard against identifiability problems. The conditional marginal probability π_{ijk} for category k is given by a difference between two conditional cumulative probabilities as

$$\pi_{ijk} = Pr(y_{ij,2} = k|\mathbf{b}_{i2}, x_{i2}, z_{i2}) = \Psi(\eta_{ijk,2}) - \Psi(\eta_{ijk-1,2}), \tag{4.2.14}$$

where $\Psi(\eta_{ijk,2})$ is the logistic cumulative distribution function (cdf) given as

$$\Psi(\eta_{ijk,2}) = \frac{\exp(\eta_{ijk,2})}{1 + \exp(\eta_{ijk,2})} = \frac{1}{1 + \exp(-\eta_{ijk,2})}.$$

The thresholds α_0 and α_K are respectively set to $-\infty$ and ∞, such that $\Psi(\eta_{ij0,2}) = 0$ and $\Psi(\eta_{ijK,2}) = 1$.

For the time-to-event T_i, we define the hazard for the i^{th} individual by the proportional hazards model

$$h_i(t) = h_0(t)\mathbf{w}_i \exp\left(\mathbf{x}'_{i\mathfrak{z}}\beta_{\mathbf{3}}\right), \tag{4.2.15}$$

where $\mathbf{x}_{i\mathfrak{z}}$ is a vector of baseline covariates, β_3 is a vector of regression coefficients of covariates, and $h_0(t)$ is baseline hazard function, which can be assumed to be of parametric form or left unspecified. The latent parameter $\mathbf{w}_i$ is the unshared (individual-specific) frailty accounting for unobservable heterogeneity and is assumed to have a log-normal distribution. By letting $\mathbf{b}_{i\mathfrak{z}} = \exp(\mathbf{w}_i)$, Equation 4.2.15 is rewritten as

$$h_i(t) = h_0(t) \exp\left(\mathbf{x}'_{i\mathfrak{z}}\beta_{\mathbf{3}} + \mathbf{b}_{i\mathfrak{z}}\right), \tag{4.2.16}$$

where $\mathbf{b}_{i\mathfrak{z}}$ is now assumed to be normally distributed. We assume a parametric Weibull distribution describes the errors in a proportional hazards model (4.2.16), such that the hazard for the i^{th} individual at time t, $h_i(t)$, is a product of baseline hazard function $h_0(t) = \lambda t^{\lambda-1}$ and $\mu_{i3}(t) = \exp\left(\mathbf{x}'_{i\mathfrak{z}}\beta_{\mathbf{3}} + \mathbf{b}_{i\mathfrak{z}}\right)$. Thus, every individual has a survival time that is Weibull with a fixed shape parameter (λ) and a scale parameter ($\mu_{i\mathfrak{z}}$), which depends on the covariates. A Weibull distribution is chosen because of its flexibility. It allows a monotonous increasing and decreasing hazard rate and by setting $\lambda = 1$ we have an Exponential distribution, which

assumes a constant hazard rate. Putting together, our proposed multivariate generalized linear mixed model for the three processes is expressed as

$$\begin{aligned} y_{ij,1}|\mathbf{b}_{i1} &\sim N\left(x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i1}, \sigma^2_{e_1} I_{ni}\right), \\ (y_{ij1,2}, \ldots, y_{ijK-1,2})|\mathbf{b}_{i2} &\sim \text{multinomial}(\pi_{ij1}, \ldots, \pi_{ijK-1}), \\ T_i|\mathbf{b}_{i3} &\sim Weibull(\lambda, \mu_{i3}), \\ \log(\mu_{i3}) &= \mathbf{x}_{i3}{}'\beta_{\mathbf{3}} + \mathbf{b}_{i3}. \end{aligned} \tag{4.2.17}$$

The random effects $\mathbf{b}_{i1}$, $\mathbf{b}_{i2}$, and $\mathbf{b}_{i3}$ are assumed to follow a Gaussian distribution with mean vectors of zeros and variance-covariance matrices $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, and $\mathbf{\Gamma}_3$, respectively. Given these random effects, the three processes are assumed to be independent of each other and the repeated measures within an individual are assumed to be independent observations from a distribution $f_l(\cdot)$. Thus, by having the latent parameters $\mathbf{b}_{i1}$, $\mathbf{b}_{i2}$, and $\mathbf{b}_{i3}$ correlated induces the association among the three processes. In this dissertation, both the shared and correlated random effects are explored. In the shared random effects model, the random effects $\mathbf{b}_{i2}$ and $\mathbf{b}_{i3}$ are assumed proportional to $\mathbf{b}_{i1}$, that is, $\mathbf{b}_{i2} = \gamma_{\mathbf{1}}\mathbf{b}_{i1}$ and $\mathbf{b}_{i3} = \gamma_{\mathbf{2}}\mathbf{b}_{i1}$. This formulation assumes a restrictive correlation structure among the outcomes. Meanwhile, in the correlated random effects model, the random effects $\mathbf{b}_{i1}$, $\mathbf{b}_{i2}$, and $\mathbf{b}_{i3}$ are assumed to follow a multivariate normal distribution with a nonrestrictive covariance structure, which could be exchangeable or any other correlation structure.

4.2.2 Likelihood for the proposed model

Let $\theta_{\mathbf{1}} = \left\{\beta_1, \sigma^2_{e_1}\right\}$, $\theta_{\mathbf{2}} = \{\beta_2, \alpha\}$ where $\alpha = (\alpha_1, \ldots, \alpha_{K-1})$ are the ordered threshold parameters for ordinal process, denote the parameters associated with the continuous and ordinal processes, respectively. In addition, let $\theta_{\mathbf{3}} = \{\beta_3, \lambda\}$, $\mathbf{\Theta} = (\theta_{\mathbf{1}}, \theta_{\mathbf{2}}, \theta_{\mathbf{3}})$, and $\mathbf{\Gamma} = (\mathbf{\Gamma_1}, \mathbf{\Gamma_2}, \mathbf{\Gamma_3})$ denote parameters associated with survival process, combined three processes, and random effects, respectively. Furthermore, let $\mathbf{y_1}$, $\mathbf{y_2}$, and $\mathbf{y_3}$ be the observed continuous, ordinal, and survival data, respectively, and $\mathbf{b}$ be the combined random effects. The survival data are right censored and we assume censoring is noninformative. The full conditional independence

assumption (Laird and Ware, 1982)[57] is also assumed such that, under the correlated normal random effects model, the joint likelihood of the three processes is given as

$$L(\boldsymbol{\Theta},\boldsymbol{\Gamma}|\mathbf{b},\mathbf{y},\mathbf{t}) = L_1(\theta_{\mathbf{1}}|\mathbf{b},\mathbf{y_1})L_2(\theta_{\mathbf{2}}|\mathbf{b},\mathbf{y_2})L_3(\theta_{\mathbf{3}}|\mathbf{b},\mathbf{y_3})g(\mathbf{b_1},\mathbf{b_2},\mathbf{b_3}|\boldsymbol{\Gamma}),$$

where

$$\begin{aligned}
L_1(\theta_{\mathbf{1}}|\mathbf{b},\mathbf{y}_{\mathit{1}}) &= \prod_{i=1}^{n} f_1(\mathbf{y}_{i\mathit{1}}|\mathbf{b}_{i\mathit{1}};\theta_{\mathbf{1}}) = \prod_{i=1}^{n}\prod_{j=1}^{n_{i1}} \frac{1}{(2\pi\sigma_{e_1}^2)^{1/2}} \exp\left\{-\frac{(y_{ij,1}-\mu_{ij,1})^2}{2\sigma_{e_1}^2}\right\},\\
\mu_{ij,1} &= x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i\mathit{1}}.\\
L_2(\theta_{\mathbf{2}}|\mathbf{b},\mathbf{y}_{\mathit{2}}) &= \prod_{i=1}^{n} f_2(\mathbf{y}_{i\mathit{2}}|\mathbf{b}_{i\mathit{2}};\theta_{\mathbf{2}}) = \prod_{i=1}^{n}\prod_{j=1}^{n_{i2}}\prod_{k=1}^{K}\{\Psi(\eta_{ijk,2}) - \Psi(\eta_{ijk-1,2})\}^{y_{ijk,2}}\\
\Psi(\eta_{ijk,2}) &= \frac{\exp(\alpha_k - \mu_{ij,2})}{1+\exp(\alpha_k-\mu_{ij,2})},\ \ \Psi(\eta_{ijk-1,2}) = \frac{\exp(\alpha_{k-1} - \mu_{ij,2})}{1+\exp(\alpha_{k-1}-\mu_{ij,2})},\\
\mu_{ij,2} &= x'_{ij,2}\beta_{\mathbf{2}} + z'_{ij,2}\mathbf{b}_{i\mathit{2}}.\\
L_3(\theta_{\mathbf{3}}|\mathbf{b},\mathbf{y_3}) &= \prod_{i=1}^{n} f_3(\mathbf{t}_i,\delta_i|\mathbf{b}_{i\mathit{3}},\theta_{\mathit{3}})^{\delta_i} S_3(\mathbf{t}_i,\delta_i|\mathbf{b}_{i\mathit{3}},\theta_{\mathit{3}})^{1-\delta_i}\\
&= \prod_{i=1}^{n}[h_0(t_i)\exp(\mathbf{x}'_{i\mathit{3}}\beta_{\mathit{3}}+\mathbf{b}_{i\mathit{3}})]^{\delta_i} \times \exp\{-\int_0^{t_i} h_0(s)\exp(\mathbf{x}'_{i\mathit{3}}\beta_{\mathit{3}}+\mathbf{b}_{i\mathit{3}})ds\}
\end{aligned} \tag{4.2.21}$$

Likelihood (4.2.21) can be written as

$$L_3(\theta_{\mathbf{3}}|\mathbf{b},\mathbf{y_3}) = \prod_{i=1}^{n}\left[\mu_i^{\delta_i}\exp(-\mu_i)\right]\left[\frac{h_0(t_i)}{H_0(t_i)}\right]^{\delta_i},$$

where

$$\mu_i = H_0(t_i)\mu_{i3},\ \ \mu_{i3} = \exp(\mathbf{x}'_{i\mathit{3}}\beta_{\mathit{3}}+\mathbf{b}_{i\mathit{3}}),\ \text{and}\ H_0(t_i) = \int_0^{t_i} h_0(s)ds. \tag{4.2.22}$$

By assuming Weibull distribution, $h_0(t_i) = \lambda t_i^{(\lambda-1)}$, $H_0(t_i) = t_i^{\lambda}$, $\dfrac{h_0(t_i)}{H_0(t_i)} = \dfrac{\lambda}{t_i}$.

Therefore,

$$L_3(\theta_{\mathbf{3}}|\mathbf{b},\mathbf{t}) = \prod_{i=1}^{n}\underbrace{\left[\mu_i^{\delta_i}\exp(-\mu_i)\right]}_{A}\underbrace{\left[\frac{\lambda}{t_i}\right]^{\delta_i}}_{B},\ \ \mu_i = t_i^{\lambda}\exp(\mathbf{x}'_{i\mathit{3}}\beta_{\mathit{3}}+\mathbf{b}_{i\mathit{3}}).$$

Part A in Equation (4.2.22) was first shown by Aitkin and Clayton (1980) [1] to be the kernel of the likelihood function for n independent "Poisson variates" δ_i with mean μ_i . Thus, the log-linear model for the hazard function implies a log-linear model for the Poisson mean: $\log \mu_i = \log(t_i^{\lambda}) + \mathbf{x}'_{i3}\beta_3 + \mathbf{b}_{i3}$, where $\log(t_i^{\lambda})$ is the offset. This formulation allows us to model the three processes as multivariate GLMM. The joint distribution of $\mathbf{b} = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, \mathbf{b}_{i3})$ is assumed to be a multivariate normal with mean vector of zero and variance-covariance matrix $\mathbf{\Gamma}$. That is,

$$\begin{pmatrix} \mathbf{b}_{i1} \\ \mathbf{b}_{i2} \\ \mathbf{b}_{i3} \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_1 & \mathbf{\Gamma}_{12} & \mathbf{\Gamma}_{13} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_2 & \mathbf{\Gamma}_{23} \\ \mathbf{\Gamma}_{31} & \mathbf{\Gamma}_{32} & \mathbf{\Gamma}_3 \end{pmatrix} \right).$$

Combining the above, the joint likelihood for the three processes is given as

$$\begin{aligned} L(\mathbf{\Theta}, \mathbf{\Gamma}|\mathbf{b}, \mathbf{y}) = & \prod_{i=1}^{n} \left[\prod_{j=1}^{n_{i1}} \frac{1}{(2\pi\sigma_{e_1}^2)^{1/2}} \exp\left\{ -\frac{(y_{ij,1} - \mu_{ij,1})^2}{2\sigma_{e_1}^2} \right\} \right] \\ & \times \left[\prod_{j=1}^{n_{i2}} \prod_{k=1}^{K} \{\Psi(\eta_{ijk,2}) - \Psi(\eta_{ijk-1,2})\}^{y_{ijk,2}} \right] \\ & \times \left[\left[\mu_i^{\delta_i} \exp(-\mu_i)\right] \left[\frac{\lambda}{t_i}\right]^{\delta_i} \right] \times \left[\frac{|\mathbf{\Gamma}|^{-1/2}}{2\pi} \exp\left\{ -\frac{\mathbf{b}_i' \mathbf{\Gamma}^{-1} \mathbf{b}_i}{2} \right\} \right], \end{aligned} \tag{4.2.23}$$

where,

$$\begin{aligned} \mathbf{b}_i &= (\mathbf{b}_{i1}, \mathbf{b}_{i2}, \mathbf{b}_{i3}),\ \mu_{ij,1} = x'_{ij,1}\beta_{\mathbf{1}} + z'_{ij,1}\mathbf{b}_{i1},\ \mu_{ij,2} = x'_{ij,2}\beta_{\mathbf{2}} + z'_{ij,2}\mathbf{b}_{i2}, \\ \Psi(\eta_{ijk,2}) &= \frac{\exp(\alpha_k - \mu_{ij,2})}{1 + \exp(\alpha_k - \mu_{ij,2})},\ \Psi(\eta_{ijk-1,2}) = \frac{\exp(\alpha_{k-1} - \mu_{ij,2})}{1 + \exp(\alpha_{k-1} - \mu_{ij,2})},\ \text{and} \\ \mu_i &= t_i^{\lambda} \exp(\mathbf{x}'_{i3}\beta_3 + \mathbf{b}_{i3}). \end{aligned}$$

4.2.3 Prior Specifications and Posterior distribution

Bayesian inferences are based on the posterior distribution, which is a function of the likelihood and prior distribution. Prior distributions are chosen to have less influence on the inferences made and conjugate (i.e., prior whose kernel has same form as that of the likelihood) where possible. Let $\mathring{\beta}_{\mathbf{1}}$ $\mathring{\beta}_{\mathbf{2}}$, and $\mathring{\beta}_{\mathbf{3}}$ denote the mean vectors, and $\mathring{\mathbf{\Sigma}}_{\mathbf{1}}$, $\mathring{\mathbf{\Sigma}}_{\mathbf{2}}$, and $\mathring{\mathbf{\Sigma}}_{\mathbf{3}}$ the variance-covariance matrices for $\beta_{\mathbf{1}}$, $\beta_{\mathbf{2}}$, and $\beta_{\mathbf{3}}$, respectively. Non-informative multivariate normal priors were

assumed for the $\beta' s$, $\beta_{\mathbf{1}} \sim MVN\left(\mathring{\beta}_{\mathbf{1}}, \mathring{\mathbf{\Sigma}}_{\mathbf{1}}\right), \beta_{\mathbf{2}} \sim MVN\left(\mathring{\beta}_{\mathbf{2}}, \mathring{\mathbf{\Sigma}}_{\mathbf{2}}\right), \beta_{\mathbf{3}} \sim MVN\left(\mathring{\beta}_{\mathbf{3}}, \mathring{\mathbf{\Sigma}}_{\mathbf{3}}\right)$ and truncated normal prior for the thresholds α, $\alpha_k \sim N\left(\mu_{\alpha_k}, \sigma_{\alpha}^2\right) I(\alpha_{k-1}, \alpha_{k+1}) \quad k = 1, \ldots, K-1$, where $I(\cdot, \cdot)$ denotes truncation to specified interval. For the variance parameters, an Inverse Wishart prior was assumed for the variance-covariance matrix of the random effects ($\mathbf{\Gamma} \sim IW\left(\nu, \Lambda\right)$) and an Inverse Gamma prior for the error variance $\left(\sigma_{e_1}^2 \sim IG(\zeta, \omega)\right)$, which are both conjugate priors in the multivariate and univariate normal likelihoods, respectively (Carlin and Louis, 2009)[9]. In separate models, we assumed Inverse Gamma for the variance of each random effects. Alternatively, a half-Cauchy prior is assumed for the standard deviation of random effects (Gelman, 2006)[27]. Lastly, for the shape parameter λ if $\neq 1$ a Gamma prior was assumed (i.e., $\lambda \sim G(\varrho, \xi)$) which is also a conjugate prior. By using vague or non-informative priors we allow the likelihood or the data to dominate the inferences made. Because random effects are unknown they are included as parameters in the posterior distribution and hence estimated together with other parameters. Given the prior distributions of all the unknowns and the observed data, the joint posterior distribution can be expressed as

$$
\begin{aligned}
\pi\left(\mathbf{\Theta}, \mathbf{\Gamma}, \mathbf{b} | \mathbf{y}\right) &\propto L(\mathbf{\Theta}, \mathbf{\Gamma} | \mathbf{b}, \mathbf{y}) \pi\left(\mathbf{\Theta}\right) \pi\left(\mathbf{\Gamma}\right) \\
&\propto \prod_{i=1}^{n} \left[\prod_{j=1}^{n_{i1}} \frac{1}{(2\pi\sigma_{e_1}^2)^{1/2}} \exp\left\{ -\frac{(y_{ij,1} - \mu_{ij,1})^2}{2\sigma_{e_1}^2} \right\} \times \frac{\omega^{\zeta}}{\Gamma(\zeta)} \left(\sigma_{e_1}^2\right)^{-(\zeta+1)} \exp\left\{ -\frac{\omega}{\sigma_{e_1}^2} \right\} \right. \\
&\quad \times \left| \mathring{\Sigma}_1 \right|^{-1/2} \exp\left\{ -\frac{1}{2} \left(\beta_1 - \mathring{\beta}_1\right)' \mathring{\Sigma}_1^{-1} \left(\beta_1 - \mathring{\beta}_1\right) \right\} \\
&\quad \times \prod_{j=1}^{n_{i2}} \prod_{k=1}^{K} \left\{ \frac{\exp(\alpha_k - \mu_{ij,2})}{1 + \exp(\alpha_k - \mu_{ij,2})} - \frac{\exp(\alpha_{k-1} - \mu_{ij,2})}{1 + \exp(\alpha_{k-1} - \mu_{ij,2})} \right\}^{y_{ijk,2}} \\
&\quad \times \prod_{k=1}^{K-1} \left(2\pi\sigma_{\alpha}^2\right)^{-1/2} \exp\left\{ -\frac{\left(\alpha_k - \mu_{\alpha}\right)^2}{2\sigma_{\alpha}^2} \right\} I_{[\alpha_{k-1}, \alpha_{k+1}]}\left(\alpha_k\right) \\
&\quad \times \left| \mathring{\Sigma}_2 \right|^{-1/2} \exp\left\{ -\frac{1}{2} \left(\beta_2 - \mathring{\beta}_2\right)' \mathring{\Sigma}_2^{-1} \left(\beta_2 - \mathring{\beta}_2\right) \right\} \\
&\quad \times \left[\mu_i^{\delta_i} \exp(-\mu_i) \right] \left[\frac{\lambda}{t_i} \right]^{\delta_i} \times \frac{\xi^{\varrho}}{\Gamma(\varrho)} \lambda^{(\varrho-1)} \exp\left\{ -\xi\lambda \right\} \\
&\quad \times \left| \mathring{\Sigma}_3 \right|^{-1/2} \exp\left\{ -\frac{1}{2} \left(\beta_3 - \mathring{\beta}_3\right)' \mathring{\Sigma}_3^{-1} \left(\beta_3 - \mathring{\beta}_3\right) \right\} \\
&\quad \left. \times |\Gamma|^{-1/2} \exp\left\{ -\frac{b_i' \Gamma^{-1} b_i}{2} \right\} \times |\Gamma|^{-\nu/2} \exp\left(-\frac{1}{2} tr\left(\Gamma^{-1}\Lambda\right) \right) \right].
\end{aligned}
\tag{4.2.31}
$$

Ignoring constants in Equation 4.2.31 gives the joint posterior as

$$\begin{aligned}
\pi\left(\boldsymbol{\Theta},\boldsymbol{\Gamma},\mathbf{b}|\mathbf{y}\right) &\propto L(\boldsymbol{\Theta},\boldsymbol{\Gamma}|\mathbf{b},\mathbf{y})\pi\left(\boldsymbol{\Theta}\right)\pi\left(\boldsymbol{\Gamma}\right) \\
&\propto L(\boldsymbol{\Theta},\boldsymbol{\Gamma}|\mathbf{b},\mathbf{y}) \times \left(\sigma_{e_1}^2\right)^{-(\zeta+1)} \exp\left\{-\frac{\omega}{\sigma_{e_1}^2}\right\} \\
&\times \exp\left\{-\frac{1}{2}\left(\beta_1-\mathring{\beta}_1\right)' \mathring{\boldsymbol{\Sigma}}_1^{-1}\left(\beta_1-\mathring{\beta}_1\right)\right\} \\
&\times \prod_{k=1}^{K-1} \exp\left\{-\frac{\left(\alpha_k-\mu_\alpha\right)^2}{2\sigma_\alpha^2}\right\} I_{[.,.]}\left(\alpha_k\right) \\
&\times \exp\left\{-\frac{1}{2}\left(\beta_2-\mathring{\beta}_2\right)' \mathring{\Sigma}_2^{-1}\left(\beta_2-\mathring{\beta}_2\right)\right\} \\
&\times \exp\left\{-\frac{1}{2}\left(\beta_3-\mathring{\beta}_3\right)' \mathring{\Sigma}_3^{-1}\left(\beta_3-\mathring{\beta}_3\right)\right\} \times (\lambda)^{(\varrho-1)} \exp\left\{-\xi\lambda\right\} \\
&\times |\boldsymbol{\Gamma}|^{-\nu/2} \exp\left\{-\frac{1}{2} tr\left(\boldsymbol{\Gamma}^{-1}\Lambda\right)\right\},
\end{aligned} \tag{4.2.32}$$

where, $L(\boldsymbol{\Theta},\boldsymbol{\Gamma}|\mathbf{b},\mathbf{y})$ is given by Equation (4.2.23). From the joint posterior distribution in Equation (4.2.32), we can draw inferences about the parameters of interest; however, to determine the appropriate MCMC sampling method or specifically to implement Gibbs sampling, the full conditionals or conditional marginal distributions for each parameter need to be constructed. In the next section (Section 4.3), we show the derivation of the full conditionals and estimation procedures for these parameters.

4.3 FULL CONDITIONAL DISTRIBUTIONS AND ESTIMATION

The conditional marginal posterior distributions or full conditionals are determined by averaging the joint posterior distribution (4.2.32) over or integrating out the remaining parameters. If a parameter distribution is proportional to some known, standard distribution, then sampling can be done using standard Gibbs sampling method. However, if a distribution is not standard and also to improve the convergence, we use the one iteration Metropolis-Hasting algorithm (Gamerman, 1997) [20]. This method uses Metropolis-Hasting algorithm with weighted least squares (WLS) proposal to draw samples from full conditional distributions. Because some of these parameters are correlated, which can lead to slow convergence, the

full conditional distributions are formed as blocks of correlated parameters according to their conditional independence [57][20]. Thus, in our case we have blocks $\beta_{\mathbf{1}}$, $\sigma^2_{e_1}$, $(\alpha, \beta_{\mathbf{2}})$, λ, $\beta_{\mathbf{3}}$, $\mathbf{b}$, and $\mathbf{\Gamma}$ and their full conditional distributions are derived below.

Let $\pi(\theta|.)$ represent the full conditional distribution of parameter θ given as parameters, then:

$$\pi\left(\beta_1|\sigma^2_{e_1}, \mathbf{b}_{i1}\right) \propto \exp\left\{-\frac{1}{2}\left(\beta_1 - \mathring{\beta}_1\right)' \mathring{\Sigma}_1^{-1}\left(\beta_1 - \mathring{\beta}_1\right)\right\} \times \prod_{i=1}^{n}\prod_{j=1}^{n_{i1}} \exp\left\{-\frac{(y_{ij,1} - \mu_{ij,1})^2}{2\sigma^2_{e_1}}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\beta_1 - \mathring{\beta}_1\right)' \mathring{\Sigma}_1^{-1}\left(\beta_1 - \mathring{\beta}_1\right)\right\} \times \exp\left\{-\frac{1}{2\sigma^2_{e_1}}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}(y_{ij,1} - \mu_{ij,1})^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\beta_1 - \mathring{\beta}_1\right)' \mathring{\Sigma}_1^{-1}\left(\beta_1 - \mathring{\beta}_1\right) + \frac{1}{\sigma^2_{e_1}}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}(y_{ij,1} - \mu_{ij,1})^2\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\beta_1 - \mathring{\beta}_1\right)' \mathring{\Sigma}_1^{-1}\left(\beta_1 - \mathring{\beta}_1\right) + \frac{1}{\sigma^2_{e_1}}(\mathbf{y}_1 - X_1\beta_1 - Z_1 b_1)'(\mathbf{y}_1 - X_1\beta_1 - Z_1 b_1)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\beta_1'\mathring{\Sigma}_1^{-1}\beta_1 - 2\beta_1'\mathring{\Sigma}_1^{-1}\mathring{\beta}_1 + \mathring{\beta}_1{}'\mathring{\Sigma}_1^{-1}\mathring{\beta}_1 \; + \sigma_{e_1}^{-2}(\mathbf{y}_1'\mathbf{y}_1 - 2\mathbf{y}_1'X_1\beta_1 - \mathbf{y}_1'Z_1b_1\right.\right.$$

$$\left.\left. + \beta_1'X_1'X_1\beta_1 + 2\beta_1'X_1'Z_1b_1 - b_1'Z_1'\mathbf{y}_1 + b_1'Z_1'Z_1b_1)\right]\right\}$$

Ignoring the terms that do not involve β_1 in the exponent above leads to

$$\propto \exp\left\{-\frac{1}{2}\left[\beta_1'\mathring{\Sigma}_1^{-1}\beta_1 - 2\beta_1'\mathring{\Sigma}_1^{-1}\mathring{\beta}_1 + \sigma_{e_1}^{-2}(-2\mathbf{y}_1'X_1\beta_1 + \beta_1'X_1'X_1\beta_1 + 2\beta_1'X_1'Z_1b_1)\right]\right\} \tag{4.3.01}$$

Simplifying 4.3.01 further gives

$$\pi\left(\beta_1|\sigma^2_{e_1}, \mathbf{b}_{i1}\right) \propto \exp\left\{-\frac{1}{2}\left[\beta_1'\mathring{\Sigma}_1^{-1}\beta_1 + \beta_1'X_1'X_1\sigma_{e_1}^{-2}\beta_1 - 2\beta_1'\left(\mathring{\Sigma}_1^{-1}\mathring{\beta}_1 + X_1'\sigma_{e_1}^{-2}(\mathbf{y}_1 - Z_1b_1)\right)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\beta_1'[\mathring{\Sigma}_1^{-1} + X_1'X_1\sigma_{e_1}^{-2}]\beta_1 - 2\beta_1'\left(\mathring{\Sigma}_1^{-1}\mathring{\beta}_1 + X_1'\sigma_{e_1}^{-2}(\mathbf{y}_1 - Z_1b_1)\right)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\beta_1'[\mathring{\Sigma}_1^{-1} + X_1'X_1\sigma_{e_1}^{-2}]\beta_1 - 2\beta_1'\left([\mathring{\Sigma}_1^{-1} + X_1'X_1\sigma_{e_1}^{-2}]^{-1} \times [\mathring{\Sigma}_1^{-1} + X_1'X_1\sigma_{e_1}^{-2}]\right.\right.\right.$$

$$\left.\left.\left. \times \left[\mathring{\Sigma}_1^{-1}\mathring{\beta}_1 + X_1'\sigma_{e_1}^{-2}\epsilon\right]\right)\right]\right\}$$

$$\propto N\left(\beta_{\mathbf{1}}{}^*, \mathbf{\Sigma_1}^*\right),$$

where,

$$\beta^*_{\mathbf{1}} = \mathbf{\Sigma_1}^* \times \left[\mathring{\mathbf{\Sigma}}_{\mathbf{1}}^{-\mathbf{1}}\mathring{\beta}_{\mathbf{1}} + X_1'\sigma_{e_1}^{-2}\epsilon\right], \; \mathbf{\Sigma_1}^* = \left[\mathring{\mathbf{\Sigma}}_{\mathbf{1}}^{-\mathbf{1}} + X_1'X_1\sigma_{e_1}^{-2}\right]^{-1},$$

$$\epsilon = \sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}\left(y_{ij,1} - z'_{ij,1}\mathbf{b}_{i1}\right).$$

(4.3.02)

$$\begin{aligned}
\pi\left(\sigma_{e_1}^2|\beta_1,\mathbf{b}_{i1}\right) &\propto \left(\sigma_{e_1}^2\right)^{-n(\zeta+1)}\exp\left\{-\frac{\omega}{\sigma_{e_1}^2}\right\}\times\prod_{i=1}^{n}\prod_{j=1}^{n_{i1}}\frac{1}{(\sigma_{e_1}^2)^{1/2}}\exp\left\{-\frac{(y_{ij,1}-\mu_{ij,1})^2}{2\sigma_{e_1}^2}\right\}\\
&\propto \left(\sigma_{e_1}^2\right)^{-n(\zeta+1)}\exp\left\{-\frac{\omega}{\sigma_{e_1}^2}\right\}\times(\sigma_{e_1}^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_{e_1}^2}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}(y_{ij,1}-\mu_{ij,1})^2\right\}\\
&\propto \left(\sigma_{e_1}^2\right)^{-\left(\frac{2\zeta+n}{2}+1\right)}\exp\left\{-\frac{1}{\sigma_{e_1}^2}\left[\omega+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}(y_{ij,1}-\mu_{ij,1})^2\right]\right\}\\
&\propto IG\left(\zeta+\frac{n}{2},\omega+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}(y_{ij,1}-\mu_{ij,1})^2\right)
\end{aligned}\tag{4.3.03}$$

$$\begin{aligned}
\pi\left(\alpha,\beta_{\mathbf{2}}|\mathbf{b}_{i2}\right) &= \pi\left(\alpha|\mathbf{b}_{i2}\right)\pi\left(\beta_2|\alpha,\mathbf{b}_{i2}\right)\\
&\propto \prod_{k=1}^{K-1}\exp\left\{-\frac{(\alpha_k-\mu_\alpha)^2}{2\sigma_\alpha^2}\right\}I_{[\alpha_{k-1},\alpha_{k+1}]}\left(\alpha_k\right)\times\\
&\quad\prod_{i=1}^{n}\prod_{j=1}^{n_{i2}}\prod_{k=1}^{K}\left\{\frac{\exp(\alpha_k-\mu_{ij,2})}{1+\exp(\alpha_k-\mu_{ij,2})}-\frac{\exp(\alpha_{k-1}-\mu_{ij,2})}{1+\exp(\alpha_{k-1}-\mu_{ij,2})}\right\}^{y_{ijk,2}}\\
&\quad\times\exp\left\{-\frac{1}{2}\left(\beta_2-\mathring{\beta}_2\right)'\mathring{\Sigma}_2^{-1}\left(\beta_2-\mathring{\beta}_2\right)\right\}
\end{aligned}\tag{4.3.04}$$

$$\begin{aligned}
\pi\left(\beta_3|\lambda,\mathbf{b}_{i3}\right) &\propto \exp\left\{-\frac{1}{2}\left(\beta_3-\mathring{\beta}_3\right)'\mathring{\Sigma}_3^{-1}\left(\beta_3-\mathring{\beta}_3\right)\right\}\times\prod_{i=1}^{n}\mu_i^{\delta_i}\exp(-\mu_i)\\
&\propto \exp\left\{-\frac{1}{2}\left(\beta_3-\mathring{\beta}_3\right)'\mathring{\Sigma}_3^{-1}\left(\beta_3-\mathring{\beta}_3\right)\right\}\times\\
&\quad\prod_{i=1}^{n}(t_i^\lambda\exp(\mathbf{x}'_{i\mathfrak{Z}}\beta_{\mathfrak{Z}}+\mathbf{b}_{i\mathfrak{Z}}))^{\delta_i}\exp(-t_i^\lambda\exp(\mathbf{x}'_{i\mathfrak{Z}}\beta_{\mathfrak{Z}}+\mathbf{b}_{i\mathfrak{Z}}))\\
&\propto \exp\left\{-\frac{1}{2}\left(\beta_3-\mathring{\beta}_3\right)'\mathring{\Sigma}_3^{-1}\left(\beta_3-\mathring{\beta}_3\right)\right\}\times\\
&\quad\prod_{i=1}^{n}\exp(\lambda\delta_i\log(t_i))[\exp(\mathbf{x}'_{i\mathfrak{Z}}\beta_{\mathfrak{Z}}+\mathbf{b}_{i\mathfrak{Z}})]^{\delta_i}\exp(-t_i^\lambda\exp(\mathbf{x}'_{i\mathfrak{Z}}\beta_{\mathfrak{Z}}+\mathbf{b}_{i\mathfrak{Z}}))
\end{aligned}\tag{4.3.05}$$

$$
\begin{aligned}
\pi\left(\beta_3|\lambda,\mathbf{b}_{i\beta}\right) &\propto \exp\left\{-\frac{1}{2}\left(\beta_3-\mathring{\beta}_3\right)'\mathring{\Sigma}_3^{-1}\left(\beta_3-\mathring{\beta}_3\right)\right\}\times \\
&\qquad \exp\sum_{i=1}^{n}\left[\delta_i\left(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta}\right)+\lambda\delta_i\log(t_i)-t_i^{\lambda}\exp(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta})\right] \\
&\propto \exp\left\{\sum_{i=1}^{n}\left[\delta_i\left(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta}\right)+\lambda\delta_i\log(t_i)-t_i^{\lambda}\exp(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta})\right]\right. \\
&\qquad \left.-\frac{1}{2}\left(\beta_3-\mathring{\beta}_3\right)'\mathring{\Sigma}_3^{-1}\left(\beta_3-\mathring{\beta}_3\right)\right\} \\
&\propto \exp\left\{\sum_{i=1}^{n}\left[\delta_i\left(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta}\right)-t_i^{\lambda}\exp(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta})\right]\right. \\
&\qquad \left.-\frac{1}{2}\left(\beta_3-\mathring{\beta}_3\right)'\mathring{\Sigma}_3^{-1}\left(\beta_3-\mathring{\beta}_3\right)\right\},
\end{aligned}
\tag{4.3.06}
$$

which is a log-concave function.

Similarly, the conditional marginal distribution of $\pi\left(\lambda|\beta_3,\mathbf{b}_{i\beta}\right)$, and for the random effects and their associated variance-covariance matrix are derived as

$$
\begin{aligned}
\pi\left(\lambda|\beta_3,\mathbf{b}_{i\beta}\right) &\propto (\lambda)^{n(\varrho-1)}\exp\left\{-\xi\lambda\right\}\times\prod_{i=1}^{n}\left[\mu_i^{\delta_i}\exp(-\mu_i)\right]\left[\frac{\lambda}{t_i}\right]^{\delta_i} \\
&\propto (\lambda)^{n(\varrho-1)}\exp\left\{-\xi\lambda\right\}\times(\lambda)^{n\delta_i} \\
&\qquad \times\prod_{i=1}^{n}t_i^{-\delta_i}(t_i^{\lambda}\exp(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta}))^{\delta_i}\exp(-t_i^{\lambda}\exp(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta})) \\
&\propto (\lambda)^{(\varrho+\delta_i-1)}\exp\left\{-\xi\lambda\right\}\times\exp\sum_{i=1}^{n}\left[-\delta_i\log(t_i)\right. \\
&\qquad \left.+\delta_i\left(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta}\right)+\lambda\delta_i\log(t_i)-t_i^{\lambda}\exp(\mathbf{x}'_{i\beta}\beta_\beta+\mathbf{b}_{i\beta})\right] \\
&\propto (\lambda)^{(\varrho+\delta_i-1)}\times\exp\left\{\sum_{i=1}^{n}\lambda\delta_i\log(t_i)-\xi\lambda\right\}\text{(By ignoring constants)} \\
&\propto (\lambda)^{(\varrho+\delta_i-1)}\times\exp\left\{n\lambda\sum_{i=1}^{n}\delta_i\log(t_i)-\xi\lambda\right\} \\
&\propto (\lambda)^{(\varrho+\delta_i-1)}\times\exp\left\{-(\xi-n\sum_{i=1}^{n}\delta_i\log(t_i))\lambda\right\} \\
&\propto \text{Gamma}\left(\varrho+\delta_i,\xi-n\sum_{i=1}^{n}\delta_i\log(t_i)\right).
\end{aligned}
\tag{4.3.07}
$$

$$
\begin{aligned}
\pi\left(\mathbf{b}|.\right) \propto & \left|\Gamma\right|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} b_i'\Gamma^{-1}b_i\right\} \times \left|\Gamma\right|^{-\nu/2} \exp\left(-\frac{1}{2}tr\left(\Gamma^{-1}\Lambda\right)\right) \\
& \times \exp\left\{-\frac{1}{2\sigma_{e_1}^2}\sum_{i=1}^{n}\sum_{j=1}^{n_{i1}}\left(y_{ij,1}-\mu_{ij,1}\right)^2\right\} \\
& \times \prod_{i=1}^{n}\prod_{j=1}^{n_{i2}}\prod_{k=1}^{K}\left\{\frac{\exp(\alpha_k-\mu_{ij,2})}{1+\exp(\alpha_k-\mu_{ij,2})}-\frac{\exp(\alpha_{k-1}-\mu_{ij,2})}{1+\exp(\alpha_{k-1}-\mu_{ij,2})}\right\}^{y_{ijk,2}} \\
& \times \prod_{i=1}^{n}(t_i^{\lambda}\exp(\mathbf{x}_{i\beta}'\beta_{\beta}+\mathbf{b}_{i\beta}))^{\delta_i}\exp(-t_i^{\lambda}\exp(\mathbf{x}_{i\beta}'\beta_{\beta}+\mathbf{b}_{i\beta}))
\end{aligned}
\tag{4.3.08}
$$

$$
\begin{aligned}
\pi\left(\Gamma|\mathbf{b}\right) & \propto \left|\Gamma\right|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} b_i'\Gamma^{-1}b_i\right\} \times \left|\Gamma\right|^{-\nu/2} \exp\left(-\frac{1}{2}tr\left(\Gamma^{-1}\Lambda\right)\right) \\
& \propto \left|\Gamma\right|^{-(n+\nu)/2} \exp\left(-\frac{1}{2}tr\left[\Gamma^{-1}\left(\Lambda+\sum_{i=1}^{n} b_i b_i'\right)\right]\right) \\
& \propto IW\left(\nu+n, \Lambda+\sum_{i=1}^{n} b_i b_i'\right).
\end{aligned}
\tag{4.3.09}
$$

Following the same procedures as in Section 3.3 Chapter 3, the parameters of interest are estimated by drawing random variates from their full conditional posterior distributions. Because the the variance-covariance parameters $\sigma_{\mathbf{e_1}}^{\mathbf{2}}, \mathbf{\Gamma}$ and the shape parameter λ have standard distributions, Gibbs sampling is employed to estimate these parameters. While for the fixed and random effects parameters, $(\beta_{\mathbf{1}}, \alpha, \beta_{\mathbf{2}}, \beta_{\mathbf{3}})$ and $\mathbf{b}_i$, Gamerman's one step Metropolis-Hasting (M-H) method is employed to sample from their respective conditional posterior distributions [20]. The following steps are taken to estimate for instance, parameter φ using a single iterative method of Gamerman[20].

Step 1: Start with $\varphi = \varphi^{(0)}$ and set $t = 1$;

Step 2a: Sample φ^* from $N(\mathbf{m}^{(t)}, \mathbf{c}^{(t)})$ proposal density and

Step 2b: Accept it with probability $\lambda(\varphi^{(t-1)}, \varphi^*)$ and set $\varphi^{(1)} = \varphi^*$; Otherwise, stay at $\varphi^{(t)} = \varphi^{(t-1)}$;

Step 3: Increase t by 1 and return to Step 2.

The moments of the proposal density are given by

$$\begin{aligned}\mathbf{m}^{(t)} &= (\Sigma_{\varphi}^{-1} + X'W(\varphi^{(t-1)})X)^{-1} \times \left\{\Sigma_{\varphi}^{-1}\mu_{\varphi} + X'W(\varphi^{(t-1)})[\tilde{y}(\varphi^{(t-1)}) - \tilde{\eta}]\right\} \\ \mathbf{c}^{(t)} &= (\Sigma_{\varphi}^{-1} + X'W(\varphi^{(t-1)})X)^{-1} \end{aligned} \tag{4.3.010}$$

where μ_{φ} and Σ_{φ} are respectively, the mean and variance-covariance matrix of the prior distribution for φ, $W(\varphi^{(t-1)}) = diag(W_{11}, \cdots, W_{nn_i})$ is the usual weight matrix for iterative weighted least squares (IWLS) algorithm. The components of the weight matrix W_{ij} and the transformed observations $\tilde{y}_{ij}$ are defined as

$$\begin{aligned}\tilde{y}_{ij}(\varphi) &= \eta_{ij} + (y_{ij} - \mu_{ij})g'(\mu_{ij}) \text{ and} \\ W_{ij}^{-1}(\varphi) &= \mathbf{V}_{ij}\left\{g'(\mu_{ij})\right\}^2, \ i = 1, \ldots, n; j = 1, \ldots, n_i, \end{aligned} \tag{4.3.011}$$

where $\mathbf{V}_{ij}$ is the conditional variance function of the outcome variable, and $g'(\mu_{ij})$ is the derivative of the link function with respect to the mean value function. The vector $\tilde{\eta}$ known as the offset in GLM is the part of the predictor associated with all the remaining effects in the model. The acceptance probability is defined as

$$\Delta(\varphi^{(t-1)}, \varphi^*) = \min\left(1, \frac{\pi(\varphi^*)q(\varphi^{(t-1)}, \varphi^*)}{\pi(\varphi^{(t-1)})q(\varphi^*, \varphi^{(t-1)})}\right), \tag{4.3.012}$$

where $\pi(\varphi^*)$ and $\pi(\varphi^{(t-1)})$ is the posterior density of φ evaluated at φ^* and $\varphi^{(t-1)}$, respectively; $q(\varphi^{(t-1)}, \varphi^*)$ is the density specified in Step 2a evaluated at φ^* and $q(\varphi^*, \varphi^{(t-1)})$ is a $N(\mathbf{m}^*, \mathbf{c}^*)$ density evaluated at $\varphi^{(t-1)}$. Thus, to draw samples from the full conditionals $\pi(\beta_1|.)$, $\pi(\alpha, \beta_2|.)$, and $\pi(b_i|.) = \pi(b_{i1}, b_{i2}|.)$, the steps above are followed.

For the β_1 block, the transformed observations are $\tilde{y}_{ij,1}(\beta_1) = x'_{ij,1}\beta_{\mathbf{1}} + (y_{ij,1} - x'_{ij,1}\beta_{\mathbf{1}})g'(x'_{ij,1}\beta_{\mathbf{1}})$ which gives the original observations, $y_{ij,1}$; the offset is the random effect part, $z'_{ij,1}b_{i1}$, and the weights are $W_{ij,1}(\beta_1) = \sigma^2_{e_1}I_{n_{i1}}, i = 1, \ldots, n; j = 1, \ldots, n_{i1}$. The proposal density $N(\mathbf{m_1}^{(t)}, \mathbf{c_1}^{(t)})$ has moments

$$\begin{aligned}\mathbf{m_1}^{(t)} &= (\mathring{\Sigma}_1^{-1} + X_1'W_1(\beta_1^{(t-1)})X_1)^{-1} \times \left\{\mathring{\Sigma}_1^{-1}\mathring{\beta}_1 + X_1'W_1(\beta_1^{(t-1)})\left[\tilde{\mathbf{y}}_{\mathbf{1}}(\beta_{\mathbf{1}}^{(\mathbf{t-1})}) - \mathbf{z}'_{\mathbf{1}}\mathbf{b}_{\mathbf{1}}\right]\right\} \\ \mathbf{c_1}^{(t)} &= (\mathring{\Sigma}_1^{-1} + X_1'W_1(\beta_1^{(t-1)})X_1)^{-1} \end{aligned} \tag{4.3.013}$$

where $W_1 = diag(W_{11,1}, \ldots, W_{nn_{i1},1})$; X_1 is the design matrix of fixed effects for outcome $\mathbf{y_1}$.

For the $\theta_{\mathbf{2}} = \{\alpha, \beta_2\}$ block associated with the ordinal outcome $\mathbf{y_2}$ with response vector for the i^{th} subject defined as $\mathbf{y}_{i2} = (y_{i1,2}, \cdots, y_{ij,2}, \cdots, y_{in_{i2},2})'$, we define $y^*_{ij,2} = 1$ if $y_{ij,2} = k, 0$ otherwise, with its expectation $\pi_{ij,2} = E(y^*_{ij,2})$ defined as in Equation (4.2.14). Thus, the $n_{i2} \times 1$ dimensional ordinal response vector $\mathbf{y}_{i2}$ is transformed into a $n_{i2}(K-1)$ dimensional binary vector $\mathbf{y}^*_{i2} = (y_{i11}, \cdots, y_{i1K-1}, y_{i21}, \cdots, y_{in_{i2}K-1})'$ with expectation $\pi_{i2} = E(\mathbf{y}^*_{i2})$. The variance-convariance matrix $\mathbf{V}_{i2}$ of the dichotomized binary response vector $\mathbf{y}^*_{i2}$ has typical elements

$$\text{cov}(y_{ijk}, y_{ij'k'}) = \begin{cases} \pi_{ijk}(1-\pi_{ijk}) & \text{if} \quad j = j', k = k', \\ -\pi_{ijk}\pi_{ijk'} & \text{if} \quad j = j', k \neq k', \\ \frac{\text{corr}(y_{ijk}, y_{ij'k'})}{\left[\pi_{ijk}(1-\pi_{ijk})\pi_{ij'k'}(1-\pi_{ij'k'})\right]^{-1/2}} & \text{if} \quad j \neq j', \text{any } k, k' \end{cases} \tag{4.3.014}$$

Let $\mu_{\theta_2} = (\mu_\alpha, \mathring{\beta}_2)$ and $\Sigma_{\theta_2} = \begin{pmatrix} \sigma^2_\alpha I_{K-1} & 0 \\ 0 & \mathring{\Sigma}_2 \end{pmatrix}$ be the mean vector and variance-covariance matrix of $\theta_{\mathbf{2}}$, respectively. Thus, the transformed observations used in estimating $\theta_{\mathbf{2}}$, are $\tilde{y}_{ij,2}(\theta_2) = \eta_{ij,2}(\theta_2) + (y^*_{ij,2} - \pi_{ij,2}(\theta_2))g'(\pi_{ij,2}(\theta_2))$, where $\eta_{ij,2}(\theta_2) = \alpha_k - x'_{ij,2}\beta_2$. The offset and weights are $z'_{ij,2}b_{i2}$ and $W_{ij,2}(\theta_2) = [n_{i2}V_{ij,2}\{g'(\pi_{ij,2}(\theta_2))\}^2]^{-1}$, respectively, where $[g'(\pi_{ij,2}(\theta_2))]^{-1}$ is the derivative of the mean function with respect to the linear predictor whose elements are given as follows:

$$[g'(\pi_{ij,2}(\theta_2))]^{-1} = \begin{cases} \frac{\exp(\alpha_k - x'_{ij,2}\beta_2)}{(1+\exp(\alpha_k - x'_{ij,2}\beta_2))^2} & k = 1, \\ \left[\frac{\exp(\alpha_k - x'_{ij,2}\beta_2)}{(1+\exp(\alpha_k - x'_{ij,2}\beta_2))^2} - \frac{\exp(\alpha_{k-1} - x'_{ij,2}\beta_2)}{(1+\exp(\alpha_{k-1} - x'_{ij,2}\beta_2))^2}\right] & k \geq 2. \end{cases} \tag{4.3.015}$$

The proposal density $N(\mathbf{m_2}^{(t)}, \mathbf{c_2}^{(t)})$ has moments

$$\begin{aligned} \mathbf{m_2}^{(t)} &= (\Sigma^{-1}_{\theta_2} + X'_2 W_2(\theta_2^{(t-1)})X_2)^{-1} \times \left\{\Sigma^{-1}_{\theta_2}\mu_{\theta_2} + X'_2 W_2(\theta_2^{(t-1)})[\tilde{\mathbf{y}}_{\mathbf{2}}(\theta_{\mathbf{2}}^{(\mathbf{t-1})}) - \mathbf{z'_2 b_2}]\right\} \\ \mathbf{c_2}^{(t)} &= (\Sigma^{-1}_{\theta_2} + X'_2 W_2(\theta_2^{(t-1)})X_2)^{-1} \end{aligned} \tag{4.3.016}$$

where $W_2 = diag(W_{11,2}, \ldots, W_{nn_{i2},2})$ and X_2 is the design matrix of fixed effects for the binary outcome associated with $\mathbf{y_2}$.

Given the scale parameter λ, the transformed observations for β_3 block are $\tilde{\delta}_i(\beta_3) = x'_{i3}\beta_{\mathbf{3}} + (\delta_i - \mu_i(\beta_3))g'(\mu_i(\beta_3))$. The offset is $\tilde{\eta}(\beta_3) = \log(t_i^\lambda) + b_{i3}$ and the weights are $W_{i3}(\beta_3) =$

$[\mu_i \{g'(\mu_i(\beta_3))\}^2]^{-1}$, where $\mu_i = t_i^{\lambda} \exp(\mathbf{x}'_{i\beta}\beta_\beta + \mathbf{b}_{i\beta})$ and $[g'(\mu_i(\beta_3))]^{-1}$ is the derivative of the mean function $\mu_i(\beta_3)$ with respect to the linear predictor $x'_{i3}\beta_{\mathbf{3}}$. The proposal density $N(\mathbf{m_3}^{(t)}, \mathbf{c_3}^{(t)})$ has moments

$$\begin{aligned}
\mathbf{m_3}^{(t)} &= (\overset{\circ}{\Sigma}_3^{-1} + X_3' W_3(\beta_3^{(t-1)}) X_3)^{-1} \times \left\{ \overset{\circ}{\Sigma}_3^{-1} \overset{\circ}{\beta}_3 + X_3' W_3(\beta_3^{(t-1)}) \left[\tilde{\delta}_{\mathbf{i}}(\beta_{\mathbf{3}}^{(\mathbf{t-1})}) - \tilde{\eta}(\beta_{\mathbf{3}}) \right] \right\} \\
\mathbf{c_3}^{(t)} &= (\overset{\circ}{\Sigma}_3^{-1} + X_1' W_3(\beta_3^{(t-1)}) X_3)^{-1}
\end{aligned} \tag{4.3.017}$$

where $W_3 = diag(W_{13}, \ldots, W_{n3})$; X_3 is the design matrix of fixed effects for time-to-event.

Following the same steps, for the $b_i = (b_{i1}, b_{i2}, b_{i3})$ block, when estimated separately, then for the b_{i1} block, we draw samples from the full conditional $\pi(b_{i1}|.)$. The transformed observations and weights are $\tilde{y}_{ij,1}(b_{i1}) = z'_{ij,1} b_{i1} + (y_{ij,1} - z'_{ij,1} b_{i1}) g'(z'_{ij,1} b_{i1}) = y_{ij,1}$ and $W_{ij,1}(b_{i1}) = \sigma_\epsilon^2 I_{n_{i1}}$, respectively. The proposal density is $N(\mathbf{m}_{i1}^{(\mathbf{t})}, \mathbf{c}_{i1}^{(\mathbf{t})})$ with moments

$$\begin{aligned}
\mathbf{m}_{i1}^{(\mathbf{t})} &= (\Gamma_1^{-1} + Z_{i1}' W_{i1}(b_{i1}^{(t-1)}) Z_{i1})^{-1} Z_{i1} W_{i1}(b_{i1}^{(t-1)}) \times \left\{ \tilde{y}_{i1}(b_{i1}^{(t-1)}) - X_{i1}' \beta_1 \right\} \\
\mathbf{c}_{i1}^{(\mathbf{t})} &= (\Gamma_1^{-1} + Z_{i1}' W_{i1}(b_{i1}^{(t-1)}) Z_{i1})^{-1}
\end{aligned} \tag{4.3.018}$$

where $W_{i1} = diag(W_{i1,1}, \ldots, W_{in_{i1},1})$ and $Z_{i1} = (z_{i1,1}, \ldots, z_{in_{i1},1})'$.

And for the b_{i2} block, we draw samples from the full conditional $\pi(b_{i2}|.)$. The transformed observations and weights for b_{i2} are $\tilde{y}^*_{ij,2}(b_{i2}) = \eta_{ij,2}(b_{i2}) + (y^*_{ij,2} - \pi_{ij,2}(b_{i2})) g'(\pi_{ij,2}(b_{i2}))$ and $W_{ij,2}(b_{i2}) = [n_{i2} V_{ij,2} \{g'(\pi_{ij,2}(b_{i2}))\}^2]^{-1}$, respectively. The proposal density is $N(\mathbf{m}_{i2}^{(\mathbf{t})}, \mathbf{c}_{i2}^{(\mathbf{t})})$ with moments

$$\begin{aligned}
\mathbf{m}_{i2}^{(\mathbf{t})} &= (\Gamma_2^{-1} + Z_{i2}' W_{i2}(b_{i2}^{(t-1)}) Z_{i2})^{-1} Z_{i2} W_{i2}(b_{i2}^{(t-1)}) \times \left\{ \tilde{y}^*_{i2}(b_{i2}^{(t-1)}) - (\alpha_k - X_{i2}' \beta_2) \right\} \\
\mathbf{c}_{i2}^{(\mathbf{t})} &= (\Gamma_2^{-1} + Z_{i2}' W_{i2}(b_{i2}^{(t-1)}) Z_{i2})^{-1}
\end{aligned} \tag{4.3.019}$$

where $W_{i2} = diag(W_{i1,2}, \ldots, W_{in_{i2},2})$ and $Z_{i2} = (z_{i1,2}, \ldots, z_{in_{i2},2})'$.

Similarly, for the b_{i3} block we draw samples from the full conditional $\pi(b_{i3}|.)$. The transformed observations and weights for b_{i3} are $\tilde{\delta}_i(b_{i3}) = \eta_{i3}(b_{i3}) + (\delta_i - \mu_i(b_{i3})) g'(\mu_i(b_{i3}))$ and $W_{i3}(b_{i3}) = [\mu_i \{g'(\mu_i(b_{i3}))\}^2]^{-1}$, respectively. The offset is $\tilde{\eta}(b_{i3}) = \log(t_i^{\lambda}) + \mathbf{x}'_{i\beta}\beta_\beta$. The proposal density is $N(\mathbf{m}_{i\beta}^{(\mathbf{t})}, \mathbf{c}_{i\beta}^{(\mathbf{t})})$ with moments

$$\begin{aligned}
\mathbf{m}_{i\beta}^{(\mathbf{t})} &= (\Gamma_3^{-1} + I_{i3}' W_{i3}(b_{i3}^{(t-1)}) I_{i3})^{-1} I_{i3} W_{i3}(b_{i3}^{(t-1)}) \times \left\{ \tilde{\delta}_i(b_{i3}^{(t-1)}) - \tilde{\eta}(b_{i3}) \right\} \\
\mathbf{c}_{i\beta}^{(\mathbf{t})} &= (\Gamma_3^{-1} + I_{i3}' W_{i3}(b_{i3}^{(t-1)}) I_{i3})^{-1}
\end{aligned} \tag{4.3.020}$$

where $W_{i3} = diag(W_{i1}, \ldots, W_{in})$ and $I_{\cdot}$ is an identity matrix.

Because the random effects are assumed to be correlated, we estimate them jointly from a multivariate distribution. Following the same steps above, we draw samples from the full conditional $\pi(b_i|.)$. The proposal density is

$$q_{b_i} \sim MVN\left(\begin{pmatrix} \mathbf{m}_{i1}^{(\mathbf{t})} \\ \mathbf{m}_{i2}^{(\mathbf{t})} \\ \mathbf{m}_{i3}^{(\mathbf{t})} \end{pmatrix}, \begin{pmatrix} \mathbf{c}_{i1}^{(\mathbf{t})} & \rho_{12}\sqrt{\mathbf{c}_{i1}^{(\mathbf{t})}}\sqrt{\mathbf{c}_{i2}^{(\mathbf{t})}} & \rho_{13}\sqrt{\mathbf{c}_{i1}^{(\mathbf{t})}}\sqrt{\mathbf{c}_{i3}^{(\mathbf{t})}} \\ \rho_{12}\sqrt{\mathbf{c}_{i2}^{(\mathbf{t})}}\sqrt{\mathbf{c}_{i1}^{(\mathbf{t})}} & \mathbf{c}_{i2}^{(\mathbf{t})} & \rho_{23}\sqrt{\mathbf{c}_{i2}^{(\mathbf{t})}}\sqrt{\mathbf{c}_{i3}^{(\mathbf{t})}} \\ \rho_{13}\sqrt{\mathbf{c}_{i3}^{(\mathbf{t})}}\sqrt{\mathbf{c}_{i1}^{(\mathbf{t})}} & \rho_{23}\sqrt{\mathbf{c}_{i3}^{(\mathbf{t})}}\sqrt{\mathbf{c}_{i2}^{(\mathbf{t})}} & \mathbf{c}_{i3}^{(\mathbf{t})} \end{pmatrix}\right),$$

where ρ_{12}, ρ_{13}, and ρ_{23} are the correlations between the continuous and ordinal, continuous and survival, and ordinal and survival processes, respectively. All correlations are estimated from the data.

4.4 SIMULATION STUDY

In order to examine the performance of the proposed joint model, we performed a series of simulation studies. In particular, we compared the regression estimates from the proposed joint model to the estimates from the separate regression models. The data were simulated from the proposed joint model of continuous and ordinal outcomes and time-to-event correlated through correlated and/or shared random effects. From each of the joint models, we simulated 500 data sets of sample sizes $n = 100$ and $n = 50$. In all simulations, number of repeated measures per subject were randomly generated from a Poisson$(1, 10)$ and time t between successive visits from a Uniform$(0.2, 2.0)$ distribution. In addition, one baseline treatment variable indicator x was generated from a Bernoulli (0.5) distribution. For the correlated random effects models, the random effects were generated from a multivariate normal distribution with mean vector zero and variance-covariance matrix

$$\Gamma = \begin{pmatrix} \sigma_{b_1}^2 = 5.87 & \rho_{12}\sigma_{b_1}\sigma_{b_2} & \rho_{13}\sigma_{b_1}\sigma_{b_3} \\ \rho_{12}\sigma_{b_1}\sigma_{b_2} & \sigma_{b_2}^2 = 4.89 & \rho_{23}\sigma_{b_2}\sigma_{b_3} \\ \rho_{13}\sigma_{b_1}\sigma_{b_3} & \rho_{23}\sigma_{b_2}\sigma_{b_3} & \sigma_{b_3}^2 = 3.2 \end{pmatrix}.$$

To determine the effect of correlation on the estimates from the joint model, we considered exchangeable with strong, moderate, and zero correlation values and unstructured correlation structures, which formed Part I of our simulations. Specifically, we considered exchangeable of $(\rho_{12} = \rho_{13} = \rho_{23} = 0.9)$, $(\rho_{12} = \rho_{13} = \rho_{23} = 0.6)$, and $(\rho_{12} = \rho_{13} = \rho_{23} = 0.0)$, respectively and unstructured of $(\rho_{12} = 0.9, \rho_{13} = 0.6, \rho_{23} = 0.3)$. Similarly, for the shared random effects models $(b_{i2} = \gamma_1 b_{i1},\ b_{i3} = \gamma_2 b_{i1})$, we considered $(\gamma_1 = \gamma_2 = 0.9)$, $(\gamma_1 = \gamma_2 = 0.6)$, and $(\gamma_1 = \gamma_2 = -0.5)$, respectively as Part II of our simulations. The error $\epsilon_{ij,1}$ for the continuous outcome was simulated from $N\left(0, \sigma^2_{\epsilon_1} = 7.4\right)$. The true values for the variances (error and random effects) were obtained by fitting a joint model to the motivating data set of Chapter 3 of this dissertation because they looked more reasonable. Meanwhile, the true values for the regression parameters were chosen based on the results of a joint correlated random effects model fit to the current motivating data set. The $\beta' s$, $(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})$ and $(\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})$ denote the regression coefficients for the fixed effects (intercept, time, treatment, and time by treatment interaction) for the continuous and ordinal outcomes, respectively, (β_{30}, β_{31}) the regression coefficients for the baseline covariates (intercept, treatment) for the survival outcome, and α_2 the threshold parameter for the ordinal outcome. The true values for the $\beta' s$ and the threshold α_2 were: $(\beta_{10} = 13.50, \beta_{11} = -0.56, \beta_{12} = -3.35, \beta_{13} = -0.57)$, $(\alpha_2 = 0.45, \beta_{20} = 1.7, \beta_{21} = -0.27, \beta_{22} = 0.65, \beta_{23} = -0.1)$, and $(\beta_{30} = -3.31, \beta_{31} = 0.37)$ for continuous, ordinal, and survival outcomes, respectively. After generating the latent parameters from their respective distributions and specifying the true values, the three outcomes were then constructed as in Equation (4.2.17). For the continuous outcome, we generated $y_{ij,1}|b_{i1}$ from a normal distribution with mean $\mu_{ij,1} = \beta_{10} + \beta_{11} \times t_{ij} + \beta_{12} \times x_i + \beta_{13} \times t_{ij} \times x_i + \mathbf{b}_{i1}$ and standard deviation σ_{ϵ_1}. For the ordinal outcome, we considered three categories and the first threshold value α_1 was set to zero to ensure identification. We then generated $y_{ijk,2}|b_{i2}$ from a multinational distribution with probabilities given by the marginal probabilities constructed from the cumulative logit model as in Equation (4.2.14). The linear predictor, $\eta_{ijk,2} = \alpha_k - [\beta_{20} + \beta_{21} \times t_{ij} + \beta_{22} \times x_i + \beta_{23} \times t_{ij} \times x_i + \mathbf{b}_{i2}]$.

For time-to-event data, the survival time, T^*_i, for the i^{th} subject was generated from a Weibul(λ, μ_{i3}) distribution, where $\log(\mu_{i3}) = \beta_{30} + \beta_{31} \times x_i + \mathbf{b}_{i3}$ and $\lambda = 1$, which is essentially an exponential distribution with rate parameter μ_{i3}. The censoring time, C_i, was

generated from a Uniform$(0, 50)$ which accounted for about 40% censoring. The failure time, T_i was taken as the minimum of survival time and censoring time, that is, $T_i = \min(T_i^*, C_i)$ and the event indicator, δ_i, was defined as

$$\delta_i = \begin{cases} 1 & T_i^* \leq C_i \\ 0 & \text{otherwise} \end{cases}$$

In summary, the following steps were taken for each simulation:

- For simulations in Part I, we generated data assuming the three processes are correlated through correlated random effects b_{i1}, b_{i2}, and b_{i3} and then fitted the correlated and separate models to the data sets generated.
- For simulations in Part II, we generated data assuming the processes are correlated through shared random effects i.e., $b_{i2} = \gamma_1 b_{i1}$ and $b_{i3} = \gamma_2 b_{i1}$ and then fitted the shared and separate models to the data sets generated.

The MCMC sampling was done using OpenBUGS (version 3.2.2) software and its R interface BRugs Version 0.8.3. We ran two chains of 10,000 iterations with 2,000 iterations of each chain used as burn-in period. The initial values for MCMC sampling were taken from a linear mixed model fit to the continuous data and a generalized linear mixed model fit to the ordinal data and time-to-event data. Let $\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})$, $\beta_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})$, $\beta_3 = (\beta_{30}, \beta_{31})$. The following priors were considered for the different parameters: $\beta_1 \sim N_4(0, 100I_4)$, $\beta_2 \sim N_4(0, 100I_4)$, $\alpha_2 \sim N(0, 10^6)\, I(0,)$, $\sigma^2_{e_1} \sim IG(1.0, 1.0)$, $\sigma^2_{b_1} \sim IG(1.0, 1.0)$, $\sigma^2_{b_2} \sim IG(1.0, 1.0)$, $\gamma \sim N(0, 100)$, and $\Gamma \sim IW(3, 1I_2)$, where I_q indicates an $q \times q$ identity matrix.

The simulation results for Part I are shown in Tables 7-12 (exchangeable correlation structure) and Tables 13-14 (unstructured correlation structure) . In each of the tables, the estimated Bias, Monte Carlo Standard Deviation (MCSD), Posterior Standard Deviation (SD), Coverage Probabilities (CP) of the 95% highest posterior density (HPD) intervals, and the Relative Efficiency (RE) are shown. RE is calculated as the ratio of the mean squared error loss (MSE) of estimates from the fitted models to the mean squared error loss (MSE) of estimates for the same parameters from the true model. All estimates were calculated based on 500 replicates.

The results in Table 7 ($\rho_{12} = \rho_{13} = \rho_{23} = 0.9$ and $n = 50$), indicate that when the true processes were correlated through correlated random effects both joint model and separate model fits provided unbiased estimates but the estimated posterior means were more biased for separate models with larger SD. These biases were larger for the ordinal outcome which may be due to the less informative nature of ordinal data as compared to continuous data. The gain in efficiency using joint model relative to the separate model was as high as 11% for the fixed effects parameters and 52% for the variance parameters. The gain in efficiency was more pronounced in the ordinal and survival outcomes. In addition, the results in Table 7 indicate that nominal coverage of 95% HPD intervals was maintained for the continuous outcome in both joint and separate models but not for the ordinal and survival outcomes. Increasing the sample size to $n = 100$ (Table 8) resulted in improved gain in efficiency, less bias, and better coverage of 95% HPD intervals for the three outcomes in both joint and separate models.

The results for moderate correlation of $\rho_{12} = \rho_{13} = \rho_{23} = 0.6$ and sample size $n = 50$ are shown in Table 9. The results showed a similar trend as in Table 7. In particular, when we fitted the correct model (JC) the biases were smaller than those when fitted the separate (SP) models. The standard errors (SD) were smaller for the JC (true) model compared to SP model. Apart from the survival outcome, the gain in efficiency reduced with reduced correlation among the outcomes. In both models, coverage of 95% HPD intervals were adequate for all estimates. Meanwhile, the results in Table 10 showed reduced gain in efficiency when the sample was increased to $n = 100$ with same correlation of $\rho_{12} = \rho_{13} = \rho_{23} = 0.6$. The coverage probabilities were robust to the sample size and bias reduced with sample size.

The results in Tables 11 & 12 when the true processes were uncorrelated, that is, ($\rho_{12} = \rho_{13} = \rho_{23} = 0$) indicated the estimates from the separate and joint correlated random effects models were quite similar, though there was some gain in efficiency for ordinal and survival outcomes' estimates.

The results for unstructured correlation structure ($\rho_{12} = 0.9, \rho_{13} = 0.6, \rho_{23} = 0.3$) in Tables 13 & 14 also indicated similar results as for strong exchangeable correlation structure in Tables 7 & 8 with the estimates for the ordinal and survival outcomes having more gain in efficiency compared to those of the continuous outcome.

In all scenarios above, the MCSD and SD were quite similar implying the Monte Carlo simulations performed as well as the MCMC sampling.

Similarly, when we fitted separate models to data that were correlated through shared random effects (see Appendix A: Tables 23-28), the estimates had larger standard errors and slightly more biased on average. There was gain in efficiency when a joint model (JS) was fitted and was more pronounced in the ordinal outcome. The MCSD and SD were similar in all scenarios and the coverage of 95% HPD intervals were adequate and robust to the sample size.

Table 7: Results when data were simulated under a correlated random effects model with strong correlation ($\rho_{12} = \rho_{13} = \rho_{23} = 0.9$) and $n = 50$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Correlated (JC)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=50)	***Continuous Process***										
	β_{10}: intercept	13.50	0.010	0.219	0.412	0.95	0.014	0.219	0.417	0.95	1.00
	β_{11}: time	-0.56	0.001	0.027	0.029	0.96	0.000	0.028	0.030	0.96	1.03
	β_{12}: treatment	-3.35	0.022	0.422	0.414	0.95	0.019	0.420	0.417	0.94	0.99
	β_{13}: time×treatment	-0.57	0.000	0.030	0.029	0.95	0.000	0.030	0.030	0.94	1.02
	Ordinal Process										
	α_2: threshold	0.45	-0.025	0.076	0.078	0.93	-0.026	0.076	0.079	0.93	1.01
	β_{20}: intercept	1.70	-0.044	0.289	0.418	0.95	-0.064	0.300	0.425	0.96	1.10
	β_{21}: time	-0.27	0.007	0.039	0.035	0.94	0.008	0.041	0.036	0.95	1.07
	β_{22}: treatment	0.65	0.005	0.422	0.409	0.93	0.016	0.433	0.419	0.95	1.06
	β_{23}: time×treatment	-0.10	0.004	0.036	0.034	0.96	0.003	0.038	0.035	0.96	1.09
	Survival Process										
	β_{30}: intercept	-3.31	0.104	0.290	0.381	0.93	0.078	0.303	0.391	0.95	1.04
	β_{31}: treatment	0.37	0.001	0.353	0.357	0.93	0.000	0.372	0.365	0.92	1.11
	Association Parameters & Variances										
	ρ_{12}	0.90	-0.010	0.034	0.043	0.97	-	-	-	-	-
	ρ_{13}	0.90	0.006	0.042	0.066	0.93	-	-	-	-	-
	ρ_{23}	0.90	0.006	0.043	0.066	0.95	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.013	0.655	1.375	0.94	-0.022	0.660	1.382	0.94	1.02
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	-0.407	1.209	1.554	0.92	-0.522	1.372	1.669	0.92	1.32
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	3.20	-0.439	1.282	1.488	0.92	-0.225	1.657	1.830	0.94	1.52
	$\sigma^2_{e_1}$: error	7.40	-0.077	0.442	0.476	0.95	-0.056	0.442	0.475	0.95	0.98

Table 8: Results when data were simulated under a correlated random effects model with strong correlation ($\rho_{12} = \rho_{13} = \rho_{23} = 0.9$) and $n = 100$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Correlated (JC)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=100)	***Continuous Process***										
	β_{10}: intercept	13.50	0.020	0.161	0.286	0.95	0.007	0.162	0.293	0.95	1.00
	β_{11}: time	-0.56	0.001	0.020	0.020	0.96	0.001	0.021	0.021	0.96	1.07
	β_{12}: treatment	-3.35	-0.025	0.307	0.290	0.96	-0.029	0.308	0.292	0.96	1.01
	β_{13}: time×treatment	-0.57	0.000	0.020	0.020	0.96	0.000	0.021	0.021	0.96	1.05
	Ordinal Process										
	α_2: threshold	0.45	-0.014	0.055	0.055	0.94	-0.016	0.056	0.055	0.94	1.02
	β_{20}: intercept	1.70	0.004	0.191	0.283	0.95	-0.021	0.194	0.292	0.95	1.04
	β_{21}: time	-0.27	0.003	0.024	0.024	0.95	0.004	0.024	0.025	0.94	1.07
	β_{22}: treatment	0.65	-0.025	0.285	0.280	0.95	-0.034	0.292	0.287	0.95	1.06
	β_{23}: time×treatment	-0.10	0.001	0.021	0.023	0.94	0.002	0.022	0.024	0.93	1.14
	Survival Process										
	β_{30}: intercept	-3.31	0.044	0.199	0.256	0.95	0.019	0.204	0.264	0.95	1.01
	β_{31}: treatment	0.37	-0.004	0.235	0.241	0.95	-0.001	0.246	0.249	0.95	1.10
	Association Parameters & Variances										
	ρ_{12}	0.90	-0.009	0.026	0.032	0.95	-	-	-	-	-
	ρ_{13}	0.90	-0.008	0.033	0.047	0.97	-	-	-	-	-
	ρ_{23}	0.90	-0.002	0.037	0.049	0.96	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.028	0.410	0.945	0.95	-0.004	0.414	0.953	0.95	1.02
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	-0.101	0.703	1.017	0.96	-0.183	0.771	1.078	0.94	1.24
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	3.20	-0.080	0.795	0.943	0.96	0.064	1.064	1.175	0.95	1.78
	$\sigma^2_{e_1}$: error	7.40	-0.029	0.326	0.333	0.95	-0.010	0.325	0.332	0.95	0.98

Table 9: Results when data were simulated under a correlated random effects model with moderate correlation ($\rho_{12} = \rho_{13} = \rho_{23} = 0.6$) and $n = 50$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Correlated (JC)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=50)	***Continuous Process***										
	β_{10}: intercept	13.50	0.030	0.231	0.412	0.95	0.027	0.231	0.416	0.96	0.99
	β_{11}: time	-0.56	-0.002	0.029	0.030	0.95	-0.002	0.029	0.030	0.95	1.03
	β_{12}: treatment	-3.35	-0.011	0.413	0.413	0.95	-0.014	0.411	0.416	0.95	0.99
	β_{13}: time×treatment	-0.57	-0.001	0.030	0.029	0.95	-0.002	0.030	0.030	0.95	1.00
	Ordinal Process										
	α_2: threshold	0.45	-0.023	0.077	0.078	0.94	-0.027	0.078	0.079	0.94	1.05
	β_{20}: intercept	1.70	-0.024	0.280	0.415	0.96	-0.061	0.281	0.427	0.94	1.05
	β_{21}: time	-0.27	0.006	0.037	0.036	0.94	0.010	0.037	0.036	0.94	1.05
	β_{22}: treatment	0.65	-0.007	0.410	0.410	0.94	-0.012	0.421	0.419	0.95	1.05
	β_{23}: time×treatment	-0.10	0.001	0.034	0.034	0.95	0.002	0.035	0.035	0.95	1.05
	Survival Process										
	β_{30}: intercept	-3.31	0.038	0.268	0.336	0.95	0.049	0.303	0.383	0.95	1.28
	β_{31}: treatment	0.37	0.019	0.364	0.321	0.95	-0.007	0.391	0.359	0.95	1.15
	Association Parameters & Variances										
	ρ_{12}	0.60	-0.016	0.074	0.110	0.95	-	-	-	-	-
	ρ_{13}	0.60	-0.099	0.125	0.148	0.89	-	-	-	-	-
	ρ_{23}	0.60	-0.094	0.132	0.152	0.91	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.083	0.682	1.363	0.95	-0.013	0.679	1.379	0.96	0.98
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	-0.233	1.164	1.560	0.95	-0.515	1.221	1.669	0.94	1.25
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	3.20	0.775	1.059	1.243	0.93	-0.064	1.521	1.754	0.95	1.34
	$\sigma^2_{e_1}$: error	7.40	-0.052	0.466	0.475	0.94	-0.030	0.462	0.472	0.94	0.97

Table 10: Results when data were simulated under a correlated random effects model with moderate correlation ($\rho_{12} = \rho_{13} = \rho_{23} = 0.6$) and $n = 100$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Correlated (JC)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=100)	***Continuous Process***										
	β_{10}: intercept	13.50	0.016	0.165	0.289	0.96	0.007	0.165	0.292	0.96	0.99
	β_{11}: time	-0.56	0.000	0.020	0.020	0.94	0.000	0.020	0.021	0.94	1.01
	β_{12}: treatment	-3.35	0.015	0.285	0.290	0.94	0.012	0.284	0.291	0.94	0.99
	β_{13}: time×treatment	-0.57	0.000	0.020	0.021	0.96	0.000	0.020	0.021	0.95	1.00
	Ordinal Process										
	α_2: threshold	0.45	-0.008	0.054	0.055	0.94	-0.010	0.054	0.055	0.94	1.02
	β_{20}: intercept	1.70	0.004	0.188	0.286	0.95	-0.018	0.189	0.293	0.94	1.02
	β_{21}: time	-0.27	0.003	0.025	0.025	0.96	0.005	0.025	0.025	0.95	1.05
	β_{22}: treatment	0.65	0.022	0.278	0.283	0.95	0.014	0.283	0.288	0.95	1.03
	β_{23}: time×treatment	-0.10	0.000	0.024	0.024	0.96	0.001	0.024	0.024	0.96	1.04
	Survival Process										
	β_{30}: intercept	-3.31	0.008	0.197	0.241	0.95	0.016	0.213	0.266	0.94	1.17
	β_{31}: treatment	0.37	0.041	0.246	0.230	0.95	0.027	0.255	0.250	0.95	1.06
	Association Parameters & Variances										
	ρ_{12}	0.60	-0.005	0.049	0.079	0.95	-	-	-	-	-
	ρ_{13}	0.60	-0.086	0.096	0.112	0.84	-	-	-	-	-
	ρ_{23}	0.60	-0.080	0.104	0.115	0.87	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.076	0.401	0.941	0.96	0.027	0.405	0.949	0.96	0.99
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	-0.100	0.700	1.050	0.96	-0.240	0.726	1.089	0.94	1.17
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	3.20	0.622	0.961	0.982	0.94	-0.001	1.152	1.197	0.96	1.01
	$\sigma^2_{e_1}$: error	7.40	-0.022	0.326	0.333	0.95	-0.010	0.325	0.332	0.95	0.99

Table 11: Results when data were simulated under a correlated random effects model but with $\rho_{12} = \rho_{13} = \rho_{23} = 0.0$ and $n = 50$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Correlated (JC)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=50)	***Continuous Process***										
	β_{10}: intercept	13.50	0.026	0.234	0.412	0.95	0.028	0.233	0.416	0.95	0.99
	β_{11}: time	-0.56	-0.002	0.029	0.030	0.94	-0.002	0.029	0.030	0.94	1.00
	β_{12}: treatment	-3.35	-0.004	0.408	0.414	0.95	-0.009	0.406	0.416	0.95	0.99
	β_{13}: time×treatment	-0.57	-0.002	0.030	0.030	0.95	-0.002	0.030	0.030	0.95	0.99
	Ordinal Process										
	α_2: threshold	0.45	-0.027	0.075	0.078	0.94	-0.031	0.076	0.079	0.93	1.05
	β_{20}: intercept	1.70	-0.040	0.273	0.411	0.94	-0.067	0.277	0.422	0.93	1.07
	β_{21}: time	-0.27	0.005	0.037	0.036	0.94	0.009	0.038	0.036	0.94	1.06
	β_{22}: treatment	0.65	-0.025	0.380	0.407	0.94	-0.031	0.385	0.415	0.95	1.03
	β_{23}: time×treatment	-0.10	0.001	0.033	0.035	0.95	0.002	0.034	0.035	0.95	1.03
	Survival Process										
	β_{30}: intercept	-3.31	0.002	0.253	0.312	0.95	0.062	0.301	0.387	0.95	1.18
	β_{31}: treatment	0.37	0.032	0.338	0.304	0.95	-0.014	0.376	0.362	0.96	1.08
	Association Parameters & Variances										
	ρ_{12}	0.00	-0.003	0.087	0.161	0.95	-	-	-	-	-
	ρ_{13}	0.00	0.015	0.230	0.277	0.95	-	-	-	-	
	ρ_{23}	0.00	0.000	0.254	0.282	0.96	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.072	0.686	1.363	0.96	-0.019	0.683	1.379	0.96	0.98
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	-0.140	1.170	1.545	0.95	-0.427	1.222	1.638	0.93	1.21
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	3.20	1.370	1.381	1.303	0.96	-0.138	1.642	1.794	0.95	0.72
	$\sigma^2_{e_1}$: error	7.40	-0.039	0.462	0.474	0.94	-0.030	0.462	0.473	0.94	0.99

Table 12: Results when data were simulated under a correlated random effects model but with $\rho_{12} = \rho_{13} = \rho_{23} = 0.0$ and $n = 100$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Correlated (JC)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=100)	***Continuous Process***										
	β_{10}: intercept	13.50	0.010	0.165	0.291	0.95	0.008	0.165	0.292	0.95	1.00
	β_{11}: time	-0.56	0.000	0.020	0.021	0.94	0.000	0.020	0.021	0.94	1.00
	β_{12}: treatment	-3.35	0.021	0.287	0.291	0.95	0.018	0.286	0.292	0.95	1.00
	β_{13}: time×treatment	-0.57	0.000	0.020	0.021	0.95	0.000	0.020	0.021	0.95	1.00
	Ordinal Process										
	α_2: threshold	0.45	-0.015	0.054	0.055	0.94	-0.017	0.055	0.055	0.94	1.03
	β_{20}: intercept	1.70	-0.011	0.193	0.285	0.95	-0.026	0.198	0.293	0.94	1.07
	β_{21}: time	-0.27	0.002	0.026	0.025	0.94	0.004	0.026	0.025	0.93	1.05
	β_{22}: treatment	0.65	-0.001	0.284	0.283	0.96	-0.012	0.285	0.287	0.96	1.01
	β_{23}: time×treatment	-0.10	0.001	0.023	0.024	0.95	0.003	0.023	0.024	0.96	1.00
	Survival Process										
	β_{30}: intercept	-3.31	-0.026	0.190	0.240	0.95	0.014	0.205	0.265	0.95	1.14
	β_{31}: treatment	0.37	0.041	0.241	0.230	0.95	0.023	0.251	0.250	0.95	1.06
	Association Parameters & Variances										
	ρ_{12}	0.00	0.003	0.055	0.115	0.95	-	-	-	-	-
	ρ_{13}	0.00	-0.011	0.136	0.172	0.94	-	-	-	-	-
	ρ_{23}	0.00	0.000	0.143	0.178	0.95	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.069	0.403	0.943	0.95	0.021	0.403	0.950	0.95	0.97
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	-0.064	0.798	1.053	0.96	-0.216	0.821	1.084	0.95	1.13
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	3.20	0.733	1.145	1.095	0.91	0.022	1.076	1.183	0.96	0.63
	$\sigma^2_{e_1}$: error	7.40	-0.015	0.325	0.333	0.95	-0.010	0.324	0.332	0.95	1.00

Table 13: Results when data were simulated under a correlated random effects model with unstructured correlation ($\rho_{12} = 0.9, \rho_{13} = 0.6, \rho_{23} = 0.3$) and $n = 50$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Correlated (JC)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=50)	***Continuous Process***										
	β_{10}: intercept	13.50	-0.019	0.228	0.413	0.96	-0.027	0.233	0.416	0.95	1.05
	β_{11}: time	-0.56	0.002	0.028	0.029	0.95	0.002	0.029	0.030	0.94	1.09
	β_{12}: treatment	-3.35	0.008	0.408	0.416	0.94	0.011	0.409	0.416	0.95	1.01
	β_{13}: time×treatment	-0.57	0.001	0.029	0.029	0.95	0.002	0.030	0.030	0.95	1.05
	Ordinal Process										
	α_2: threshold	0.45	0.027	0.078	0.079	0.94	0.028	0.078	0.079	0.94	1.01
	β_{20}: intercept	1.70	0.058	0.278	0.418	0.94	0.076	0.287	0.427	0.94	1.09
	β_{21}: time	-0.27	-0.010	0.037	0.036	0.94	-0.011	0.038	0.036	0.93	1.07
	β_{22}: treatment	0.65	0.020	0.402	0.412	0.95	0.021	0.415	0.418	0.95	1.07
	β_{23}: time×treatment	-0.10	-0.002	0.034	0.034	0.95	-0.002	0.036	0.035	0.95	1.12
	Survival Process										
	β_{30}: intercept	-3.31	0.006	0.269	0.345	0.95	-0.022	0.289	0.378	0.95	1.16
	β_{31}: treatment	0.37	0.000	0.335	0.330	0.96	0.014	0.350	0.355	0.94	1.10
	Association Parameters & Variances										
	ρ_{12}	0.90	-0.007	0.036	0.047	0.95	-	-	-	-	-
	ρ_{13}	0.60	0.034	0.127	0.161	0.95	-	-	-	-	-
	ρ_{23}	0.30	0.104	0.178	0.208	0.91	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.023	0.675	1.377	0.96	0.018	0.683	1.380	0.96	1.02
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	0.380	1.141	1.550	0.94	0.533	1.247	1.673	0.92	1.27
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	3.20	-0.586	1.284	1.392	0.97	-0.028	1.543	1.705	0.96	1.19
	$\sigma^2_{e_1}$: error	7.40	0.023	0.459	0.470	0.94	0.030	0.461	0.472	0.94	1.01

Table 14: Results when data were simulated under a correlated random effects model with unstructured correlation ($\rho_{12} = 0.9, \rho_{13} = 0.6, \rho_{23} = 0.3$) and $n = 100$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Correlated (JC)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JC}}$
Correlated (n=100)	***Continuous Process***										
	β_{10}: intercept	13.50	-0.018	0.165	0.290	0.96	-0.008	0.166	0.292	0.95	1.00
	β_{11}: time	-0.56	0.000	0.020	0.020	0.95	0.000	0.020	0.021	0.95	1.05
	β_{12}: treatment	-3.35	-0.025	0.284	0.289	0.95	-0.020	0.286	0.292	0.95	1.01
	β_{13}: time×treatment	-0.57	0.001	0.019	0.020	0.95	0.000	0.020	0.021	0.95	1.05
	Ordinal Process										
	α_2: threshold	0.45	0.012	0.052	0.055	0.95	0.013	0.052	0.055	0.94	1.00
	β_{20}: intercept	1.70	0.009	0.184	0.289	0.95	0.020	0.187	0.293	0.95	1.05
	β_{21}: time	-0.27	-0.005	0.024	0.024	0.96	-0.005	0.025	0.025	0.95	1.05
	β_{22}: treatment	0.65	-0.019	0.270	0.282	0.96	-0.015	0.276	0.288	0.96	1.04
	β_{23}: time×treatment	-0.10	-0.001	0.023	0.023	0.94	-0.002	0.024	0.024	0.95	1.09
	Survival Process										
	β_{30}: intercept	-3.31	-0.017	0.199	0.255	0.95	-0.029	0.210	0.268	0.95	1.13
	β_{31}: treatment	0.37	-0.016	0.255	0.242	0.95	-0.013	0.262	0.252	0.95	1.05
	Association Parameters & Variances										
	ρ_{12}	0.90	-0.006	0.025	0.034	0.94	-	-	-	-	-
	ρ_{13}	0.60	0.025	0.088	0.112	0.94	-	-	-	-	-
	ρ_{23}	0.30	0.047	0.120	0.144	0.93	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.020	0.394	0.947	0.96	-0.020	0.405	0.949	0.96	1.06
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.89	0.188	0.701	1.038	0.94	0.244	0.753	1.089	0.94	1.19
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	3.20	-0.215	0.971	1.081	0.97	0.058	1.126	1.216	0.96	1.28
	$\sigma^2_{e_1}$: error	7.40	0.008	0.323	0.331	0.94	0.011	0.324	0.332	0.95	1.01

4.5 ANALYSIS OF UGANDAN DIABETES DATA

In this section, we apply our proposed joint model to the diabetes data introduced in Section 4.1. These data were collected retrospectively from three hospitals (Mulago, Nsambya, and Rubaga all in Kampala, Uganda) to determine the factors associated with time to blood glucose normalization (Buhule et al., 2007) [8]. Because the blood glucose test is more expensive than the urine glucose test, quite a number of individuals had no blood glucose measurements taken on most of the hospital visits. About 825 of the 1010 patients had no blood glucose measurements taken on most of the occasions but instead urine glucose was taken. In the original 2007 study, only a survival (unshared frailty) model was fitted to these data from which inferences were made. For the current study we considered data for type 2 diabetic patients who had blood glucose measurements taken and were not in the normal range on the first hospital visit. We defined time-to-event as time to when blood glucose level reached normal range of $70 - 180$ mg/dl to include those who may have not been fasting (scenario 1) and $70 - 130$ mg/dl the fasting glucose normal range (scenario 2). For scenario 1, we had a total of 500 patients and out of these, 314 experienced the event of interest (37% censoring). While for scenario 2, we had 543 patients and 248 experienced the event of interest, which was a 54% censoring rate.

Because high blood pressure and BMI are believed to be associated with type 2 diabetes we used them as repeatedly measured biomarkers in this analysis. Our goal was to understand how the biomarkers are related to time to normal blood glucose level for diabetic patients given different treatments and other covariates. The covariates of interest included treatment (Biguanides, Sulphonyureas, and Insulin (baseline)), baseline age in years, gender (male=1, female=0), time of hospital visits in months, and time and treatment interaction. The summary statistics of these covariates under the two scenarios are given in Table 15. We observed similar distributions for all variables across the two scenarios. The average baseline age was about 54 years with a standard deviation of 11 years and the majority of the patients were female (about 70%). The median failure time was about 12 months while the median censoring time was about 24 months. Most of the patients were treated with Biguanides at baseline.

Table 15: Descriptive statistics for baseline characteristics and survival times

	Blood Glucose Normal Range (mg/dl)	
	$70 - 180$	$70 - 130$
Variable	$n = 500$	$n = 543$
Age (years), $\bar{x}$ (s)	53.8 (11.2)	53.9 (11.1)
Gender		
Male, n (%)	151 (30.2)	159 (29.3)
Female, n (%)	349 (69.8)	384 (70.7)
Treatment		
Biguanides, n (%)	198 (39.6)	212 (39.0)
Sulphonyureas, n (%)	149 (29.8)	166 (30.6)
Insulin, n (%)	153 (30.6)	165 (30.4)
Survival Times		
Failure, $\tilde{x}$ (s)	11.9 (22.5)	12.3 (22.6)
Censoring, $\tilde{x}$ (s)	23.6 (29.9)	24.4 (28.5)

We jointly modeled the two biomarkers and time to normal blood glucose level (scenarios 1 & 2 separately) through correlated and shared random effects and compared the parameter estimates to separate models.

For the continuous biomarker, we considered systolic blood pressure in millimeters of mercury (mm Hg) which was square root transformed to improve normality. The conditional transformed measurements were then modeled as Normal $(y_{ij,1}|\mathbf{b}_{i1} \sim N(\mu_{ij,1}, \sigma^2_{e_1})$ and thus linked to the linear predictor with an identity link. That is,

$$y_{ij,1} = \mu_{ij,1} = \beta_{10} + \beta_{11} \times time_{ij} + \beta_{12} \times Biguan_i + \beta_{13} \times Sulphon_i + \beta_{14} \times age_i + \beta_{15} \times gender_i + \beta_{16} \times Biguan_i \times time_{ij} + \beta_{17} \times Sulphon_i \times time_{ij} + \mathbf{b}_{i1} + \epsilon_{ij,1},$$

where, $y_{ij,1}$ is the square root of systolic blood pressure (mm Hg) for the i^{th} subject measured at the j^{th} hospital visit, $\mathbf{b}_{i1}$ is the random intercept and $\epsilon_{ij,1} \sim N(0, \sigma^2_{e_1})$ is the measurement error independent of $\mathbf{b}_{i1}$.

Lastly, we defined the ordinal biomarker by grouping BMI into $K = 4$ categorizes as below:

$$y_{ijk,2} = \begin{cases} 1 \quad \text{for} \quad \text{BMI} < 18.5 & \Rightarrow \text{Underweight} \\ 2 \quad \text{for} \quad 18.5 \leq \text{BMI} \leq 24.9 & \Rightarrow \text{Normal} \\ 3 \quad \text{for} \quad 25 \leq \text{BMI} \leq 29.9 & \Rightarrow \text{Overweight} \\ 4 \quad \text{for} \quad \text{BMI} \geq 30 & \Rightarrow \text{Obesity} \end{cases}$$

The conditional $K-1$ measurements were modeled through a multinomial distribution $((y_{ij1,2}, \ldots, y_{ijK-1,2})|\mathbf{b}_{i2} \sim \text{multinomial}(\pi_{ij1}, \ldots, \pi_{ijK-1}))$. The marginal probabilities were linked to the covariates through a cumulative logit link as follows:

$$\begin{aligned} \log\left(\frac{Pr(y_{ij,2} \leq k)}{1 - Pr(y_{ij,2} \leq k)}\right) = \alpha_k - [&\beta_{21} \times time_{ij} + \beta_{22} \times Biguan_i + \beta_{23} \times Sulphon_i \\ &+ \beta_{24} \times age_i + \beta_{25} \times gender_i + \beta_{26} \times Biguan_i \times time_{ij} \\ &+ \beta_{27} \times Sulphon_i \times time_{ij} + \mathbf{b}_{i2}], \end{aligned}$$

where $k = 1, \ldots, 4$ and $\mathbf{b}_{i2}$ is the random intercept.

Lastly, time to normal blood glucose level T was modeled through a Weibull distribution with frailty truncated to the left by the censoring times. Thus, $T_i|\mathbf{b}_{i3} \sim Weibull(\lambda, \mu_{i3})$, where $\log(\mu_{i3}) = \beta_{30} + \beta_{31} \times Biguan_i + \beta_{32} \times Sulphon_i + \beta_{33} \times age_i + \beta_{34} \times gender_i + \mathbf{b}_{i3}$ and $\mathbf{b}_{i3}$ is the frailty term. For simplicity, we set $\lambda = 1$ which gave an Exponential distribution.

For the correlated random effects joint model, the random effects were modeled as

$$\begin{pmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Gamma = \begin{pmatrix} \sigma^2_{b_1} & \rho_{12}\sigma_{b_1}\sigma_{b_2} & \rho_{13}\sigma_{b_1}\sigma_{b_3} \\ \rho_{12}\sigma_{b_2}\sigma_{b_1} & \sigma^2_{b_2} & \rho_{23}\sigma_{b_2}\sigma_{b_3} \\ \rho_{13}\sigma_{b_3}\sigma_{b_1} & \rho_{23}\sigma_{b_3}\sigma_{b_2} & \sigma^2_{b_3} \end{pmatrix} \right).$$

While for the shared random effects joint model, we assumed $b_{i2} = \gamma_1 b_{i1}$ and $b_{i3} = \gamma_2 b_{i1}$ with $b_{i1} \sim N(0, \sigma^2_{b_1})$.

For all parameters, vague or non-informative priors were employed to allow the data to dominate the inferences made. Specifically, multivariate normal priors with mean zero and

precision matrices $0.01I_8$, $0.01I_7$, and $0.01I_5$ were employed for β_1, β_2, and β_3, respectively, where I_q indicates an $q \times q$ identity matrix. For the thresholds or cut points α_k, a truncated normal $N(0, 10^6)I(\alpha_{k-1}, \alpha_{k+1})$; $k = 1, \ldots, K-1$, where $I(\cdot, \cdot)$ denotes truncation to specified interval was employed. In addition, an inverse gamma prior was assumed for error variance $\sigma^2_{e_1}$, that is, $\sigma^2_{e_1} \sim$ Inverse Gamma$(1, 1)$ or $1/\sigma^2_{e_1} \sim$ Gamma$(1, 1)$ and an Inverse Wishart for the variance convariance matrix of the random effects, which are both conjugate priors. For the separate models, we assumed half-Cauchy(s=25) priors for $\sigma_{b_1}, \sigma_{b_2}$, and σ_{b_3}, and for the shape parameter λ of the Weibull distribution a conjugate Gamma prior ($\lambda \sim$ Gamma$(0.1, 0.1)$) was assumed when we didn't set it to 1. The association parameters for shared random effects model had normal priors, that is, $\gamma_1 \sim N(0, 100)$ and $\gamma_2 \sim N(0, 100)$.

To accelerate computation, time was standardized while age was centered, and the MCMC was run for $30,000$ iterations with the first $5,000$ discarded as burn-in. The models were fitted in OpenBUGS (version 3.2.2) and its R interface BRugs Version 0.8.3 and based on standard MCMC diagnostic plots (see Appendix D Figures 11-18) and diagnostic tests (Table 16) the estimated parameters converged though more iterations would be required for better convergence of the ordinal outcome regression estimates.

Tables 17 & 18 give a summary of the posterior estimates of the regression coefficients and their 95% credible intervals (CIs) for the joint versus separate analyses under scenario 1 and scenario 2, respectively.

The results in Table 17 for normal blood glucose range of $70 - 180$ mg/dl showed similar point estimates across the three models but subtle differences were observed in their credible intervals (CIs). The CIs for the parameter estimates for the ordinal and survival outcomes shrunk in both the joint correlated and joint shared analyses but all shrunk in the joint shared analysis indicating improved efficiency of parameter estimates. Age was found to be significantly associated with systolic blood pressure in all three models. Diabetes patients had increasing systolic blood pressure levels with age, indicating increased risk of systolic hypertension among the elderly. These findings are consistent with studies that have indicated that hypertension which is associated with type 2 diabetes is more common in women than in men and that the age-related increase in systolic blood pressure is steeper in women (Williams, 2003)[98]. The results in Table 17 also showed that age, gender, and the Sulphonyureas

treatment were significantly associated with BMI levels in all three models and Sulphonyureas and time interaction was significant in the joint correlated and separate models. Male patients tended to have lower BMI levels as compared to their female counterparts, and the elderly patients had lower BMI levels. Patients who were treated with Sulphonyureas had higher BMI levels at baseline as compared to those treated with Insulin but over time their BMI levels decreased. This is supported by the fact that Sulphonyureas causes weight gain in the first years which levels off with time. For the survival outcome, Sulphonyureas treatment was significantly associated with time to normal blood glucose level in all three models while gender was significant in the shared random effects model. The patients who were treated with Sulphonyureas had their blood glucose levels reach normal range faster than those treated with Insulin. These finding are different from the original study [8] where Biguanides was found to work better than Insulin and Sulphonyureas. However, Sulphonyureas is only given to Type 2 diabetics, thus by including the Type 1 diabetics in the original analysis could have masked the effect of Sulphonyureas. The male patients had their blood glucose levels reach normal range later than the female patients, which is consistent with the original study [8].

Furthermore, the posterior estimates of the association parameters ρ_{12} and ρ_{23} in the joint correlated random effects model and γ_1 and γ_2 in the joint shared random effects model were positive and significantly different from zero. Implying, positive association between the systolic blood pressure and BMI sub-models (ρ_{12} and γ_1), BMI and survival sub-models (ρ_{23}), and systolic blood pressure and survival sub-models (γ_2). Thus, the initial level of systolic blood pressure was positively associated with the BMI levels and also with time to normal blood glucose levels. In addition, the initial levels of BMI were positively associated with time to normal blood glucose levels. The goodness of fit measure (DIC), indicated that the joint correlated random effects model fit our data better than the shared random effects model.

Table 16: Convergence Diagnostic tests results for analysis of Type 2 Diabetes Data: Normal blood glucose 70 – 180 mg/dl.

	Joint-Correlated (JC)				Joint-Shared (JS)				Separate (SP)			
	Gelman	Geweke	Heidelberg		Gelman	Geweke	Heidelberg		Gelman	Geweke	Heidelberg	
Parameter	Estimate	Z-score	Stationarity	Halfwidth	Estimate	Z-score	Stationarity	Halfwidth	Estimate	Z-score	Stationarity	Halfwidth
Continuous Process: systolic blood pressure												
Intercept	1.00	-1.015	passed	passed	1.00	1.242	passed	passed	1.00	-0.429	passed	passed
Time	1.00	0.472	passed	failed	1.00	-1.834	passed	passed	1.00	0.673	passed	passed
Treatment												
Biguanides	1.00	-0.437	passed	passed	1.00	-0.621	passed	failed	1.00	0.076	passed	passed
Sulphonyureas	1.00	-1.063	passed	passed	1.00	-0.471	passed	passed	1.00	-1.324	passed	passed
Age in years	1.00	0.233	passed	passed	1.00	2.267	failed	$< NA >$	1.00	0.672	passed	passed
Male	1.00	1.771	passed	passed	1.00	-1.197	passed	passed	1.00	1.666	passed	passed
Treatment×Time												
Biguanides×Time	1.00	1.374	passed	passed	1.00	1.963	passed	passed	1.00	-0.699	passed	passed
Sulphonyureas×Time	1.00	1.473	passed	passed	1.00	-0.043	passed	passed	1.00	-0.371	passed	passed
Ordinal Process: BMI categories												
Threshold-1	1.02	0.137	passed	passed	1.03	-0.647	passed	passed	1.02	0.845	passed	passed
Threshold-2	1.02	1.447	passed	passed	1.03	-0.620	passed	passed	1.06	1.848	passed	passed
Threshold-3	1.01	1.279	passed	passed	1.01	-0.541	passed	passed	1.08	2.697	failed	$< NA >$
Time	1.00	-1.561	passed	passed	1.00	-1.457	passed	passed	1.00	-0.992	passed	failed
Treatment												
Biguanides	1.03	1.960	passed	failed	1.03	-0.267	passed	failed	1.07	0.971	passed	failed
Sulphonyureas	1.04	0.829	passed	passed	1.02	-0.178	passed	passed	1.08	-0.693	passed	failed
Age in years	1.01	1.234	passed	passed	1.01	1.621	failed	$< NA >$	1.00	-0.183	passed	passed
Male	1.06	2.267	passed	passed	1.03	-0.521	passed	passed	1.06	-0.979	passed	passed
Treatment×Time												
Biguanides×Time	1.00	1.236	passed	passed	1.00	1.719	passed	passed	1.00	-0.775	passed	passed
Sulphonyureas×Time	1.00	0.490	passed	passed	1.01	0.049	passed	passed	1.00	0.336	passed	passed
Survival Process: time to a normal blood glucose level												
Intercept	1.00	-1.870	passed	passed	1.00	0.833	passed	passed	1.00	0.959	passed	passed
Treatment												
Biguanides	1.00	-1.971	passed	passed	1.00	-0.370	failed	$< NA >$	1.00	0.231	passed	passed
Sulphonyureas	1.00	1.073	passed	passed	1.00	0.240	passed	passed	1.00	-1.215	passed	passed
Age in years	1.00	1.314	passed	passed	1.00	2.363	failed	$< NA >$	1.00	1.447	passed	passed
Male	1.00	1.620	passed	passed	1.00	-0.157	passed	passed	1.00	0.668	passed	passed
Association Parameters & Variances												
ρ_{12}	1.00	-1.797	passed	passed	-	-	-	-	-	-	-	-
ρ_{13}	1.00	2.049	passed	failed	-	-	-	-	-	-	-	-
ρ_{23}	1.02	-1.054	passed	passed	-	-	-	-	-	-	-	-
γ_1	-	-	-	-	1.05	-1.355	passed	passed	-	-	-	-
γ_2	-	-	-	-	1.00	-1.598	passed	passed	-	-	-	-
$\sigma^2_{b_1}$	1.00	1.099	passed	passed	1.01	1.655	passed	passed	1.00	1.306	passed	passed
$\sigma^2_{b_2}$	1.00	-0.732	passed	passed	-	-	-	-	1.00	1.226	passed	passed
$\sigma^2_{b_3}$	1.01	1.483	passed	passed	-	-	-	-	1.00	-0.703	passed	passed
$\sigma^2_{e_1}$	1.00	-0.877	passed	passed	1.00	0.521	passed	passed	1.00	0.876	passed	passed
Multivariate Test	1.10	-	-	-	1.07	-	-	-	1.06	-	-	-

The results for normal blood glucose range of $70 - 130$ mg/dl are shown in Table 18. The convergence diagnostic results (see Appendix D: Figures 15-18) indicated that all parameters converged except for regression estimates of the ordinal outcome. The point estimates were also quite similar across the three models but the CIs differed. There were slightly more significant variables than in scenario 1 (Table 17) but the trends were similar. Sulphonyureas treatment and age were found to be significantly associated with systolic blood pressure. The patients who were treated with Sulphonyureas at baseline had systolic blood pressure that were lower as compared to those treated with Insulin. Again, the elderly were associated with higher levels of systolic blood pressure. For the ordinal outcome, the significant variables remained the same as in Table 17 but Sulphonyureas treatment and time interaction was significantly associated with BMI levels in all the three models. Similar results as in Table 17 were obtained for the survival outcome, with Sulphonyureas significant in all models and gender in shared random effects model. Here, the only association parameters significantly different from zero were ρ_{12} and γ_1; indicating positive correlation between systolic blood pressure and BMI but no association with survival outcome. Also, the DIC indicated joint correlated random effects model fit the data better than the shared random effects model.

Table 17: Analysis of Ugandan Type 2 Diabetes Data: Normal blood glucose $70-180$ mg/dl

Parameter	Joint-Correlated Random Effects		Joint-Shared Random Effects		Separate	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
Continuous Process: systolic blood pressure						
Intercept	11.88	(11.720, 12.050)	11.93	(11.800, 12.060)	11.88	(11.720, 12.050)
Time	0.015	(-0.095, 0.125)	0.028	(-0.078, 0.134)	0.033	(-0.072, 0.137)
Treatment						
Biguanides	-0.053	(-0.254, 0.148)	-0.017	(-0.187, 0.153)	-0.051	(-0.254, 0.152)
Sulphonyureas	-0.127	(-0.346, 0.092)	-0.131	(-0.304, 0.043)	-0.119	(-0.336, 0.095)
Age in years	0.018	(0.011, 0.026)	0.022	(0.016, 0.029)	0.018	(0.010, 0.026)
Male	-0.069	(-0.260, 0.120)	-0.124	(-0.287, 0.040)	-0.074	(-0.261, 0.112)
Treatment×Time						
Biguanides×Time	-0.056	(-0.215, 0.104)	-0.016	(-0.177, 0.147)	-0.079	(-0.236, 0.081)
Sulphonyureas×Time	0.112	(-0.066, 0.289)	0.157	(-0.001, 0.314)	0.115	(-0.061, 0.288)
Ordinal Process: BMI categories						
Threshold-1	-14.4	(-17.55, -11.780)	-12.1	(-14.25, -10.110)	-14.71	(-17.73, -12.090)
Threshold-2	-4.294	(-6.108, -2.602)	-3.822	(-5.155, -2.491)	-4.506	(-6.134, -2.938)
Threshold-3	2.788	(1.233, 4.497)	2.105	(0.895, 3.407)	2.694	(1.209, 4.156)
Time	0.483	(-0.088, 1.057)	0.464	(-0.021, 0.943)	0.244	(-0.353, 0.810)
Treatment						
Biguanides	1.499	(-0.305, 3.396)	0.99	(-0.531, 2.475)	1.343	(-0.500, 3.056)
Sulphonyureas	3.034	(1.219, 5.200)	2.283	(0.905, 3.767)	2.814	(0.834, 4.711)
Age in years	-0.084	(-0.155, -0.014)	-0.071	(-0.127, -0.014)	-0.082	(-0.157, -0.011)
Male	-3.712	(-5.409, -1.938)	-3.226	(-4.593, -1.835)	-3.697	(-5.598, -1.948)
Treatment×Time						
Biguanides×Time	-0.5	(-1.356, 0.356)	-0.441	(-1.181, 0.305)	-0.531	(-1.418, 0.353)
Sulphonyureas×Time	-0.898	(-1.789, -0.068)	-0.626	(-1.378, 0.132)	-0.891	(-1.794, -0.034)
Survival Process: time to a normal blood glucose level						
Intercept	-3.922	(-4.239, -3.620)	-3.882	(-4.128, -3.643)	-3.893	(-4.211, -3.590)
Treatment						
Biguanides	0.286	(-0.078, 0.654)	0.285	(-0.009, 0.580)	0.242	(-0.127, 0.610)
Sulphonyureas	0.925	(0.549, 1.305)	0.766	(0.475, 1.059)	0.894	(0.519, 1.283)
Age in years	0.003	(-0.011, 0.016)	0.002	(-0.009, 0.013)	0.002	(-0.011, 0.016)
Male	-0.242	(-0.561, 0.079)	-0.357	(-0.610, -0.110)	-0.23	(-0.555, 0.107)
Association Parameters & Variances						
ρ_{12}	0.195	(0.071, 0.315)	-	-	-	-
ρ_{13}	-0.032	(-0.289, 0.215)	-	-	-	-
ρ_{23}	0.357	(0.160, 0.555)	-	-	-	-
γ_1	-	-	21.04	(17.590, 24.490)	-	-
γ_2	-	-	0.987	(0.488, 1.531)	-	-
$\sigma^2_{b_1}$	0.513	(0.415, 0.622)	0.092	(0.064, 0.127)	0.512	(0.415, 0.622)
$\sigma^2_{b_2}$	60.87	(42.970, 84.500)	-	-	63.25	(45.010, 89.450)
$\sigma^2_{b_3}$	0.764	(0.373, 1.233)	-	-	0.836	(0.003, 1.395)
$\sigma^2_{e_1}$	0.695	(0.631, 0.766)	1.297	(1.196, 1.406)	0.696	(0.631, 0.767)
Goodness of fit (DIC)						
Continuous	3316	-	3815	-	3321	-
Ordinal	1091	-	1196	-	1083	-
Survival	2815	-	2880	-	2825	-
Total	7222	-	7890	-	-	-

Table 18: Analysis of Ugandan Type 2 Diabetes Data: Normal blood glucose $70-130$ mg/dl

	Joint-Correlated Random Effects		Joint-Shared Random Effects		Separate	
Parameter	Mean	95% CI	Mean	95% CI	Mean	95% CI
Continuous Process: systolic blood pressure						
Intercept	11.93	(11.780, 12.080)	11.94	(11.830, 12.050)	11.93	(11.770, 12.080)
Time	0.005	(-0.087, 0.097)	0.032	(-0.055, 0.119)	0.027	(-0.062, 0.116)
Treatment						
Biguanides	-0.091	(-0.279, 0.096)	-0.061	(-0.203, 0.080)	-0.083	(-0.274, 0.107)
Sulphonyureas	-0.242	(-0.442, -0.041)	-0.27	(-0.418, -0.125)	-0.227	(-0.424, -0.027)
Age in years	0.017	(0.010, 0.024)	0.016	(0.011, 0.021)	0.017	(0.010, 0.024)
Male	-0.116	(-0.292, 0.060)	-0.125	(-0.260, 0.010)	-0.124	(-0.300, 0.053)
Treatment×Time						
Biguanides×Time	0.048	(-0.086, 0.182)	0.041	(-0.092, 0.174)	0.026	(-0.109, 0.158)
Sulphonyureas×Time	0.059	(-0.069, 0.188)	0.091	(-0.028, 0.211)	0.062	(-0.068, 0.189)
Ordinal Process: BMI categories						
Threshold-1	-15.65	(-18.240, -13.300)	-13.34	(-15.44, -11.380)	-15.87	(-18.750, -13.400)
Threshold-2	-4.791	(-6.324, -3.281)	-4.342	(-5.729, -2.896)	-4.825	(-6.464, -3.348)
Threshold-3	2.73	(1.235, 4.210)	2.221	(0.924, 3.700)	2.789	(1.348, 4.187)
Time	0.261	(-0.227, 0.771)	0.236	(-0.181, 0.639)	0.183	(-0.340, 0.710)
Treatment						
Biguanides	1.813	(-0.050, 3.537)	1.211	(-0.433, 2.889)	1.798	(0.185, 3.544)
Sulphonyureas	2.651	(0.832, 4.290)	2.161	(0.136, 3.960)	2.69	(0.932, 4.551)
Age in years	-0.105	(-0.174, -0.031)	-0.093	(-0.152, -0.038)	-0.098	(-0.170, -0.022)
Male	-4.177	(-5.891, -2.389)	-3.564	(-5.225, -2.144)	-4.05	(-5.919, -2.081)
Treatment×Time						
Biguanides×Time	0.008	(-0.749, 0.764)	-0.035	(-0.668, 0.620)	0.038	(-0.720, 0.830)
Sulphonyureas×Time	-0.874	(-1.588, -0.174)	-0.741	(-1.361, -0.102)	-0.853	(-1.568, -0.114)
Survival Process: time to a normal blood glucose level						
Intercept	-4.268	(-4.631, -3.927)	-4.15	(-4.412, -3.898)	-4.283	(-4.658, -3.935)
Treatment						
Biguanides	-0.057	(-0.470, 0.363)	0.012	(-0.301, 0.331)	-0.077	(-0.513, 0.361)
Sulphonyureas	0.554	(0.135, 0.987)	0.426	(0.112, 0.736)	0.562	(0.120, 1.002)
Age in years	0.008	(-0.007, 0.024)	0.005	(-0.007, 0.016)	0.009	(-0.007, 0.025)
Male	-0.298	(-0.682, 0.083)	-0.335	(-0.615, -0.062)	-0.299	(-0.685, 0.086)
Association Parameters & Variances						
ρ_{12}	0.183	(0.071, 0.292)	-	-	-	-
ρ_{13}	-0.107	(-0.314, 0.096)	-	-	-	-
ρ_{23}	0.159	(-0.013, 0.332)	-	-	-	-
γ_1	-	-	24.42	(21.390, 28.080)	-	-
γ_2	-	-	0.402	(-0.103, 0.921)	-	-
$\sigma^2_{b_1}$	0.533	(0.445, 0.633)	0.086	(0.063, 0.115)	0.531	(0.444, 0.630)
$\sigma^2_{b_2}$	70.7	(53.830, 92.350)	-	-	73.29	(54.400, 95.570)
$\sigma^2_{b_3}$	1.249	(0.627, 1.978)	-	-	1.484	(0.832, 2.210)
$\sigma^2_{e_1}$	0.683	(0.635, 0.735)	1.256	(1.177, 1.340)	0.683	(0.635, 0.735)
Goodness of fit (DIC)						
Continuous	5042	-	5886	-	5047	-
Ordinal	1426	-	1524	-	1421	-
Survival	2427	-	2529	-	2416	-
Total	8895	-	9938	-	-	-

4.6 DISCUSSION

In this chapter we have developed a full Bayesian hierarchical multivariate generalized linear mixed effects model for repeatedly measured continuous and ordinal measures of disease severity and time-to-event outcome. This model extended the work in Chapter 3 to include time-to-event data. Although this model is more complex, use of Bayesian MCMC methods for parameter estimation makes it more user-friendly, given that in many clinical studies there is more than one biomarker that is associated with the event of interest. Moreover, we have used a parametric approach to model the baseline hazard in proportional hazards model for survival outcome which is more flexible than the semi-parametric approach especially when the Cox proportional hazards assumption is violated. In addition, we have illustrated how the three outcomes can be modeled jointly through correlated and shared random effects with a real data example after examining their performance through simulations. In addition, our proposed joint model can easily be fit in OpenBUGS with code we provide (see Appendix F).

The results from the simulation study illustrated that joint modeling leads to efficient estimates and adequate 95% coverage probabilities for the population parameters. For the correlated random effects joint model, the efficiency gain was larger for the ordinal and survival outcomes estimates than for the continuous outcome. While, for the shared random effects joint model, the gain in efficiency was larger for the ordinal outcomes compared to that of the continuous and survival outcomes. Overall, the gain in efficiency increased with the increase in correlation among the three outcomes and decrease in the sample size. Furthermore, the diabetes data analysis results showed improved efficiency or more precise estimates when systolic blood pressure, body mass index, and time to normalization of blood glucose were modeled jointly. However, the large treatment effects could be due to the fact that these treatments are given at different stages of the disease.

While in the final stages of this dissertation, two papers by Luo (2014)[66] and Baghfalaki, Ganjali, and Berridge (2014)[2] came out that have dealt with joint modeling of multivariate longitudinal outcomes and time-to-event data. Luo [66] worked on joint modeling of binary, ordinal, and continuous longitudinal outcomes, and time-to-event data and used Bayesian approach for parameter estimation. In modeling the longitudinal outcomes, multilevel item

response theory (MLIRT) model was employed where each of the outcomes was modeled first as a function of latent measure of disease severity θ_{ij} and then at the second level, θ_{ij} was regressed on the covariates of interest. For the time-to-event data, the accelerated failure time (AFT) model was employed, and the longitudinal outcomes were linked to time-to-event data through shared random effects. Baghfalaki et al. [2] on the other hand have worked on continuous and ordinal longitudinal outcomes and time-to-event data, and have also employed Bayesian approach for parameter estimation like we did. However, for the ordinal outcome, they considered a continuous latent variable model (logistic) and for the time-to-event data they employed AFT model, and linked the longitudinal and time-to-event processes through shared random effects. Our work differs from the above two in several ways: first we employed generalized linear mixed effects models and modeled the outcomes as multivariate generalized linear mixed effects model linked through both correlated (general case) and shared random effects. Secondly, time-to-event data was modeled through a parametric Weibull distribution with unshared frailty to account for unobserved heterogeneity within individuals as well as correlation with the longitudinal biomarkers.

5.0 CONCLUSION

5.1 SUMMARY

In this dissertation, we have developed a multivariate joint model for repeatedly measured mixed (continuous and ordinal) biomarkers of disease severity and time-to-event for highly unbalanced data. This work was motivated by a diabetes observational study with highly unbalanced data, because patients reported for check-ups at different time points and the number of hospital visits varied from patient to patient. The main markers of diabetes disease severity in this study were blood glucose in mg/dl (continuous) and urine glucose levels (ordinal). Other markers taken that are associated with type 2 diabetes in particular; included blood pressure in mmHG and body mass index.

This dissertation work was done in two parts, where in Part 1 (Chapter 3), we developed a multivariate joint model for highly unbalanced repeatedly measured continuous and ordinal markers of disease severity. Each of the outcomes was assumed to be from a distribution that is in the exponential family, where the conditional mean function was linked to the linear predictor through some known monotonic function. Thus, for the continuous outcomes, we assumed an identity link function while for the ordinal outcome a cumulative logit link function was assumed. Given the random effects, the two outcomes were assumed to be independent of each other and the repeated measures within an individual were independent observations from a known distribution in the exponential family. The two outcomes were then modeled as multivariate generalized linear mixed models linked through correlated and /or shared random effects. We employed the Bayesian MCMC methods for parameter estimation because they have the capacity to handle these complex models with ease. Simulation studies were conducted to assess the performance of our proposed joint model, and the results indicated

gain in efficiency, unbiased estimates, and adequate 95% coverage probabilities when the two correlated outcomes were modeled jointly. In addition, we fitted our proposed models to the diabetes data, and the results showed improved efficiency when blood glucose and urine glucose were modeled jointly. Lastly, a sensitivity analysis was conducted to asses the effect of the priors for the variance parameters on the regression parameters. Although differences were observed in the estimates of variance parameters especially when the hyper-parameters for Gamma priors were varied, the regression parameters of interest were not affected.

In Part 2 of this dissertation, we extended the work in Part 1 to include time-to-event data (Chapter 4). Following the same steps, the three outcomes were modeled as multivariate generalized linear mixed models linked through correlated and/or shared random effects. For time-to-event, we followed Aitkin and Clayton (1980) [1] and modeled the indicator variable δ_i (i.e. whether event occurred or not) as a Poisson variate with mean μ_i, where the mean was linked to the linear predictor through a log link function. The simulation study results also indicated gain in efficiency for estimates from the proposed joint model compared to separate models. The estimates from the proposed joint model were less biased and had adequate 95% coverage probabilities. Finally, the results from fitting the proposed joint model to diabetes data indicated more efficient estimates when systolic blood pressure, BMI, and time to when blood glucose reached the normal range were modeled jointly than separately.

5.2 EXTENSIONS AND FUTURE WORK

5.2.1 Predictions

The work in this thesis has mainly focused on the joint evaluation of the repeatedly measured biomarkers of disease severity and time-to-event data. However, one of the key aims of modeling markers and time-to-event simultaneously, is to predict the event of interest after adjusting for longitudinal or repeatedly measured markers. Thus, the immediate focus will be on predictions of either an event of interest given both the longitudinal measurements and survival data or vise verse. Specifically, we can predict the event of interest or longitudinal

value for a new subject or individual from the posterior predictive distribution. Supposing there are m individuals with full data on longitudinal markers (continuous and ordinal) and time-to-event summarized as $D_m = (\mathbf{y}_{11}, \ldots, \mathbf{y}_{m1},\ \mathbf{y}_{12}, \ldots, \mathbf{y}_{m2},\ \mathbf{y}_{13}, \ldots, \mathbf{y}_{m3})$, where $\mathbf{y}_{i1}$, $\mathbf{y}_{i2}$, and $\mathbf{y}_{i3} = (T_i, \delta_i)$ are the continuous, ordinal, and time-to-event data for the i^{th} subject or individual, respectively. As defined earlier, $T_i = \min(T_i^*, C_i)$ is the failure time and $\delta_i = I(T_i^* \leq C_i)$ is an event indicator which indicates whether the observed failure time is a true failure time, T_i^*, or a censoring time C_i for the i^{th} subject. Now suppose we have a new $(m+1^{th})$ subject who has survived up to time s and provided continuous and ordinal measurements up to this time all summarized as $\mathbf{y} = \left(\mathbf{y}_{(m+1)1},\ \mathbf{y}_{(m+1)2},\ \mathbf{y}_{(m+1)3}\right)$, where, $\mathbf{y}_{(m+1)3} = (T_{m+1} = s, \delta_{m+1} = 0)$. Then, given this data, the predictive distribution for a new observation $\tilde{y}$ (continuous or ordinal) from this distribution with random effects $\tilde{b}$ and hyper-parameters $\Theta = (\beta_1, \beta_2, \beta_3, \alpha, \sigma_\epsilon^2, \lambda, \Gamma)$ is

$$p(\tilde{y}|D_m, y) = \int\int p(\tilde{y}|\mathbf{y}, \tilde{b}, \Theta) p(\tilde{b}|\mathbf{y}, \Theta) p(\Theta|D_m) d\Theta d\tilde{b}.$$

In the same way, the predicted survival probability for the time to event $\tilde{T}$ for the new subject, at time t given survival up to time s is

$$p(\tilde{T} \geq t|D_m, \tilde{T} > s, \mathbf{y}_{(m+1)1}, \mathbf{y}_{(m+1)2}) =$$

$$\int\int p(\tilde{T} \geq t|\tilde{T} > s, \mathbf{y}_{(m+1)1}, \mathbf{y}_{(m+1)2}, \tilde{b}) p(\tilde{b}|\tilde{T} > s, \mathbf{y}_{(m+1)1}, \mathbf{y}_{(m+1)2}, \Theta) p(\Theta|D_m) d\Theta d\tilde{b},$$

where $p(\tilde{b}|\tilde{T} > s, \mathbf{y}_{(m+1)1}, \mathbf{y}_{(m+1)2}, \Theta)$ is the posterior distribution of the random effects for the new subject conditional on their data and the hyper-parameters Θ (Sweeting and Thompson, 2011) [89]. We will compare predictions of survival probabilities given different scenarios (i.e., only survival data, survival and ordinal, survival and continuous, and (survival, ordinal, and continuous)). In addition, several methods that include but limited to calibration measures (Schemper and Henderson, 2000; Henderson et al., 2002) [82, 43] and discrimination measures (Heagerty et al., 2000; Heagerty and Zheng, 2005) [37, 38] will be employed to assess the accuracy of these predictions.

5.2.2 Recurrent and multiple events

Based on our motivating diabetes data, diabetics always have their blood glucose levels fluctuating given other conditions surrounding them. Thus, the blood glucose levels can be normal, too low (hypoglycemia), or to high (hyperglycemia) and these conditions are recurrent. One of our desirable future work will therefore include extending the survival submodel to account for recurrent and/or multiple events.

5.2.2.1 Recurrent events Considering only the recurrent events (e.g. normal blood glucose), one way would be to use a shared frailty model as defined by Equation 2.2.41 or 2.2.42. Because these events are assumed to be correlated, the shared frailty will account for this correlation within individuals as well as among the outcomes when we jointly model the biomarkers and time-to-event.

5.2.2.2 Multiple recurrent events For the multiple recurrent events, a possible way will be to employ a Multistate Markov model (Huzurbazar, 2005) [50], where each of the conditions, hypoglycemia, normal, and hyperglycemia is taken as a state. That is, we depict the transitions among three possible states (1=hypoglycemia, 2=normal, and 3 =hyperglycemia) as a Markov chain. Let $\lambda_{jk}(s)$ denote the hazard of progression from state j to k at time t. Then normal blood glucose is reached from hypoglycemia according to hazard $\lambda_{12}(s)$, from hyperglycemia according to hazard $\lambda_{32}(s)$. Alternatively, someone can move from having normal blood glucose to hypoglycemia according to hazard $\lambda_{21}(s)$ or normal to hyperglycemia with hazard $\lambda_{23}(s)$. To include covariate information and link to the longitudinal submodels, we will construct a proportional hazards model with frailty for each transition. This allows great flexibility for different covariates and different linking functions in each hazard submodel.

5.3 PUBLIC HEALTH SIGNIFICANCE

Diabetes is one of the most challenging public health problems globally. Its associated complications such as heart attack and stroke are the leading causes of death especially in the developed world. Because diabetes is a chronic illness and the severity of its complications can be so awful, it is a very costly disease to the individuals, families, and to the health-personnel as well as institutions. Proper treatment, however, can control and prevent the development of these complications and hence improve the quality of life of millions of people, and reduce the associated costs. The work in this dissertation proposes more effective statistical methods that can be employed to estimate the treatment effects efficiently. This will help researchers as well as clinicians identify effective treatments that can slow down the disease progression.

APPENDIX A

CHAPTER 3 SIMULATION RESULTS WHEN THE TRUE PROCESSES ARE CORRELATED THROUGH SHARED RANDOM EFFECTS

Table 19: Results when data were simulated under a shared random effects model ($\gamma = 0.9$): SD and CP, stand for posterior standard deviation and coverage probabilities of the 95% HPD intervals, respectively.

True Model (n)	Parameter	Truth	Fitted Model: Joint-Correlated (JC) Bias	JC SD	JC CP	JC $RE_1 = \frac{MSE_{JC}}{MSE_{JS}}$	Joint-Shared (JS) Bias	JS SD	JS CP	Separate (SP) Bias	SP SD	SP CP	SP $RE_2 = \frac{MSE_{SP}}{MSE_{JS}}$
Shared (n=50)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.067	0.584	0.95	0.96	-0.129	0.606	0.94	-0.075	0.600	0.95	0.97
	β_{11}: time	-0.56	0.001	0.069	0.95	1.00	0.002	0.068	0.95	0.002	0.072	0.95	1.10
	β_{12}: treatment	-0.50	0.028	0.849	0.96	0.97	0.136	0.864	0.95	0.052	0.848	0.96	0.97
	β_{13}: time×treatment	0.30	-0.001	0.097	0.94	1.01	-0.001	0.097	0.94	-0.002	0.101	0.95	1.09
	Ordinal Process												
	α_2: threshold	1.25	0.073	0.167	0.92	1.05	0.059	0.169	0.94	0.056	0.166	0.94	1.00
	β_{20}: intercept	1.80	0.071	0.565	0.95	1.01	-0.031	0.582	0.95	0.090	0.615	0.94	1.10
	β_{21}: time	-0.35	-0.026	0.074	0.94	1.06	-0.021	0.071	0.95	-0.020	0.076	0.95	1.10
	β_{22}: treatment	-0.50	-0.054	0.805	0.94	1.01	0.080	0.819	0.95	-0.109	0.838	0.95	1.04
	β_{23}: time×treatment	0.10	0.012	0.100	0.96	1.03	0.010	0.094	0.96	0.011	0.103	0.95	1.10
	Association Parameters & Variances												
	γ	0.9	-	-	-	-	0.076	0.137	0.92	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.142	1.503	0.96	0.87	0.276	1.603	0.95	0.276	1.596	0.95	1.00
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.75	0.847	1.776	0.92	-	-	-	-	1.125	2.093	0.91	-
	$\sigma^2_{e_1}$: error	7.40	-0.047	0.623	0.94	0.98	0.041	0.631	0.94	0.044	0.644	0.94	1.06
Shared (n=100)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.020	0.420	0.94	1.02	-0.025	0.415	0.94	-0.060	0.426	0.94	1.03
	β_{11}: time	-0.56	0.006	0.048	0.96	1.00	0.006	0.048	0.96	0.006	0.050	0.95	1.08
	β_{12}: treatment	-0.50	0.036	0.604	0.94	1.00	0.064	0.592	0.94	0.067	0.606	0.94	0.99
	β_{13}: time×treatment	0.30	-0.002	0.068	0.95	1.00	-0.002	0.068	0.95	-0.003	0.071	0.95	1.07
	Ordinal Process												
	α_2: threshold	1.25	0.030	0.116	0.95	1.04	0.021	0.116	0.95	0.020	0.118	0.95	0.99
	β_{20}: intercept	1.80	0.054	0.399	0.96	1.04	0.018	0.389	0.95	-0.023	0.402	0.96	1.05
	β_{21}: time	-0.35	-0.011	0.050	0.95	1.02	-0.006	0.050	0.96	-0.004	0.053	0.94	1.07
	β_{22}: treatment	-0.50	0.007	0.559	0.94	0.99	0.061	0.536	0.95	0.092	0.549	0.94	1.06
	β_{23}: time×treatment	0.10	0.003	0.066	0.94	1.00	-0.001	0.066	0.95	-0.004	0.070	0.96	1.07
	Association Parameters & Variances												
	γ	0.9	-	-	-	-	0.035	0.090	0.93	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.102	1.046	0.94	0.96	0.108	1.060	0.94	0.137	1.083	0.94	1.06
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.75	0.472	1.151	0.93	-	-	-	-	0.472	1.260	0.94	-
	$\sigma^2_{e_1}$: error	7.40	-0.010	0.436	0.95	0.95	0.069	0.447	0.95	0.064	0.449	0.95	1.02

Table 20: Results when data were simulated under a shared random effects model ($\gamma = 0.6$): SD and CP, stand for posterior standard deviation and coverage probabilities of the 95% HPD intervals, respectively.

True Model (n)	Parameter	Truth	Fitted Model: Joint-Correlated (JC) Bias	JC SD	JC CP	JC $RE_1 = \frac{MSE_{JC}}{MSE_{JS}}$	Joint-Shared (JS) Bias	JS SD	JS CP	Separate (SP) Bias	SP SD	SP CP	SP $RE_2 = \frac{MSE_{SP}}{MSE_{JS}}$
Shared (n=50)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.031	0.592	0.95	0.98	-0.079	0.609	0.95	-0.047	0.607	0.95	1.00
	β_{11}: time	-0.56	0.002	0.071	0.94	1.00	0.003	0.070	0.94	0.002	0.073	0.94	1.06
	β_{12}: treatment	-0.50	0.030	0.849	0.94	0.98	0.114	0.869	0.94	0.060	0.852	0.94	1.00
	β_{13}: time×treatment	0.30	0.000	0.099	0.94	1.00	-0.001	0.099	0.94	0.000	0.102	0.94	1.07
	Ordinal Process												
	α_2: threshold	1.25	0.075	0.151	0.91	1.16	0.048	0.152	0.94	0.045	0.150	0.94	1.01
	β_{20}: intercept	1.80	0.119	0.446	0.94	1.10	0.030	0.438	0.96	0.076	0.457	0.94	1.09
	β_{21}: time	-0.35	-0.027	0.068	0.93	1.12	-0.016	0.064	0.93	-0.015	0.069	0.94	1.10
	β_{22}: treatment	-0.50	-0.040	0.613	0.96	1.05	0.045	0.609	0.95	-0.032	0.613	0.96	1.04
	β_{23}: time×treatment	0.10	0.013	0.091	0.95	1.05	0.009	0.084	0.94	0.008	0.092	0.94	1.11
	Association Parameters & Variances												
	γ	0.6	-	-	-	-	0.044	0.097	0.92	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.095	1.501	0.96	0.87	0.291	1.611	0.96	0.341	1.611	0.96	1.01
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	2.11	0.517	0.848	0.92	-	-	-	-	0.432	0.927	0.93	-
	$\sigma^2_{e_1}$: error	7.40	0.032	0.637	0.95	0.96	0.100	0.645	0.95	0.086	0.649	0.95	1.04
Shared (n=100)	***Continuous Process***												
	β_{10}: intercept	15.34	-0.035	0.419	0.96	1.00	-0.042	0.418	0.96	-0.066	0.425	0.95	1.01
	β_{11}: time	-0.56	0.002	0.049	0.95	1.00	0.001	0.049	0.95	0.002	0.050	0.95	1.07
	β_{12}: treatment	-0.50	0.038	0.598	0.94	0.99	0.063	0.592	0.95	0.063	0.604	0.95	1.00
	β_{13}: time×treatment	0.30	-0.001	0.069	0.95	1.00	-0.002	0.069	0.95	-0.003	0.071	0.95	1.08
	Ordinal Process												
	α_2: threshold	1.25	0.039	0.105	0.93	1.11	0.021	0.105	0.95	0.021	0.108	0.95	1.03
	β_{20}: intercept	1.80	0.061	0.303	0.96	1.10	0.010	0.298	0.96	-0.008	0.305	0.96	1.07
	β_{21}: time	-0.35	-0.016	0.045	0.93	1.12	-0.007	0.045	0.95	-0.007	0.048	0.95	1.12
	β_{22}: treatment	-0.50	-0.013	0.416	0.94	1.06	0.036	0.401	0.94	0.055	0.408	0.94	1.04
	β_{23}: time×treatment	0.10	0.008	0.060	0.96	1.05	0.002	0.060	0.96	0.000	0.063	0.96	1.07
	Association Parameters & Variances												
	γ	0.6	-	-	-	-	0.020	0.063	0.95	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.003	1.033	0.95	0.95	0.046	1.052	0.95	0.100	1.075	0.95	1.06
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	2.11	0.289	0.542	0.92	-	-	-	-	0.165	0.570	0.94	-
	$\sigma^2_{e_1}$: error	7.40	-0.029	0.438	0.95	0.97	0.030	0.450	0.94	0.015	0.446	0.95	1.01

APPENDIX B

RESULTS FROM FITTING THE JOINT AND SEPARATE MODELS IN CHAPTER 3 TO UGANDAN TYPE 2 DIABETES DATA

Table 21: Convergence Diagnostic tests results for analysis of Type 2 Diabetes data.

	Joint-Correlated (JC)				Joint-Shared (JS)				Separate (SP)			
	Gelman	Geweke	Heidelberg		Gelman	Geweke	Heidelberg		Gelman	Geweke	Heidelberg	
Parameter	Estimate	Z-score	Stationarity	Halfwidth	Estimate	Z-score	Stationarity	Halfwidth	Estimate	Z-score	Stationarity	Halfwidth
Continuous Process: blood glucose levels												
Intercept	1.00	-0.956	passed	passed	1.00	0.053	passed	passed	1.00	0.536	passed	passed
Time	1.00	0.512	passed	passed	1.00	-1.169	passed	failed	1.00	0.908	passed	passed
Treatment												
Biguanides	1.00	1.806	passed	passed	1.00	-0.063	passed	passed	1.00	0.945	passed	passed
Sulphonyureas	1.00	1.592	passed	passed	1.00	-0.466	passed	passed	1.00	1.575	passed	passed
BMI	1.00	1.055	passed	passed	1.00	-0.015	passed	passed	1.00	0.561	passed	passed
Male	1.00	-0.452	passed	passed	1.00	-1.583	passed	passed	1.00	-0.621	passed	passed
Treatment×Time												
Biguanides×Time	1.00	-0.821	passed	passed	1.00	1.410	passed	passed	1.00	0.191	passed	passed
Sulphonyureas×Time	1.00	0.574	passed	passed	1.00	1.528	passed	passed	1.00	0.521	passed	passed
Ordinal Process: urine glucose categories												
Threshold-1	1.00	-1.760	passed	passed	1.01	-0.160	passed	passed	1.00	0.778	passed	passed
Threshold-2	1.00	-1.512	passed	passed	1.01	-0.186	passed	passed	1.00	0.407	passed	passed
Threshold-3	1.00	-1.654	passed	passed	1.00	-0.266	passed	passed	1.00	0.571	passed	passed
Threshold-4	1.00	-1.401	passed	passed	1.00	-0.353	passed	passed	1.00	-0.631	passed	passed
Time	1.00	0.032	passed	passed	1.00	-0.990	passed	passed	1.00	0.141	passed	passed
Treatment												
Biguanides	1.00	-1.719	passed	passed	1.01	-0.191	passed	passed	1.00	0.228	passed	passed
Sulphonyureas	1.00	0.432	passed	passed	1.00	-0.156	passed	passed	1.00	-1.384	passed	passed
BMI	1.00	1.065	passed	passed	1.01	0.157	passed	passed	1.00	1.281	passed	passed
Male	1.00	-0.413	passed	passed	1.00	-1.282	passed	passed	1.00	-1.263	passed	passed
Treatment×Time												
Biguanides×Time	1.00	0.772	passed	passed	1.00	1.517	passed	passed	1.00	-0.494	passed	passed
Sulphonyureas×Time	1.00	0.272	passed	passed	1.01	1.203	passed	passed	1.00	0.385	passed	passed
Association Parameters & Variances												
ρ	1.00	-0.838	passed	passed	-	-	-	-	-	-	-	-
γ	-	-	-	-	1.00	-0.552	passed	passed	-	-	-	-
$\sigma^2_{b_1}$	1.00	1.630	passed	passed	1.00	1.124	passed	passed	1.00	-1.077	passed	passed
$\sigma^2_{b_2}$	1.00	1.412	passed	passed	-	-	-	-	1.27	0.995	passed	failed
$\sigma^2_{e_1}$	1.00	0.156	passed	passed	1.00	-0.795	passed	passed	1.00	0.607	passed	passed
Multivariate Test	1.01	-	-	-	1.03	-	-	-	1.00	-	-	-

Table 22: Analysis of Ugandan Type 2 Diabetes Data

	Joint-Correlated (JC)		Joint-Shared (JS)		Separate (SP)	
Parameter	Mean	95% CI	Mean	95% CI	Mean	95% CI
Continuous Process: blood glucose levels						
Intercept	15.82	(15.210, 16.440)	16.14	(15.570, 16.690)	15.74	(15.090, 16.380)
Time	-0.39	(-0.793, 0.020)	-0.08	(-0.474, 0.314)	-0.42	(-0.825, -0.006)
Treatment						
Biguanides	-0.31	(-0.968, 0.344)	-0.58	(-1.197, 0.041)	-0.26	(-0.940, 0.430)
Sulphonyureas	-0.39	(-1.118, 0.341)	-0.79	(-1.461, -0.103)	-0.20	(-0.942, 0.547)
BMI	-0.03	(-0.096, 0.030)	-0.04	(-0.097, 0.009)	-0.03	(-0.093, 0.034)
Male	-0.54	(-1.492, 0.451)	-1.06	(-1.890, -0.232)	-0.46	(-1.420, 0.523)
Treatment×Time						
Biguanides×Time	0.39	(-0.117, 0.909)	0.36	(-0.145, 0.860)	0.41	(-0.108, 0.925)
Sulphonyureas×Time	0.54	(-0.082, 1.156)	0.46	(-0.156, 1.073)	0.48	(-0.150, 1.112)
Ordinal Process: urine glucose categories						
Threshold-1	-0.91	(-1.391, -0.433)	-0.73	(-1.137, -0.312)	-1.08	(-1.519, -0.612)
Threshold-2	-0.45	(-0.926, 0.024)	-0.30	(-0.701, 0.107)	-0.62	(-1.059, -0.159)
Threshold-3	0.46	(-0.004, 0.935)	0.54	(0.149, 0.948)	0.29	(-0.145, 0.748)
Threshold-4	2.60	(2.102, 3.111)	2.51	(2.087, 2.963)	2.44	(1.959, 2.933)
Time	-0.42	(-0.699, -0.145)	-0.45	(-0.708, -0.199)	-0.28	(-0.569, 0.010)
Treatment						
Biguanides	-0.58	(-1.061, -0.080)	-0.44	(-0.864, -0.003)	-0.76	(-1.253, -0.239)
Sulphonyureas	-0.33	(-0.849, 0.188)	-0.20	(-0.664, 0.270)	-0.45	(-0.981, 0.101)
BMI	-0.04	(-0.082, 0.009)	-0.03	(-0.067, 0.015)	-0.04	(-0.084, 0.009)
Male	0.72	(0.020, 1.410)	0.85	(0.231, 1.442)	0.52	(-0.148, 1.231)
Treatment×Time						
Biguanides×Time	0.31	(-0.055, 0.668)	0.26	(-0.075, 0.599)	0.26	(-0.106, 0.628)
Sulphonyureas×Time	0.45	(-0.001, 0.895)	0.38	(-0.009, 0.782)	0.38	(-0.070, 0.830)
Association Parameters & Variances						
ρ	0.60	(0.465, 0.716)	-	-	-	-
γ	-	-	0.83	(0.603, 1.085)	-	-
$\sigma^2_{b_1}$	6.97	(5.310, 8.889)	4.07	(2.697, 5.862)	7.02	(5.389, 8.972)
$\sigma^2_{b_2}$	3.46	(2.444, 4.719)	-	-	3.33	(2.364, 4.512)
$\sigma^2_{e_1}$	7.42	(6.713, 8.211)	9.53	(8.383, 10.720)	7.40	(6.684, 8.177)
Goodness of Fit (DIC)						
Continuous		5160		5323		5170
Ordinal		2573		2639		2591
Total		7733		7962		-

APPENDIX C

CHAPTER 4 SIMULATION RESULTS WHEN THE TRUE PROCESSES ARE CORRELATED THROUGH SHARED RANDOM EFFECTS

Table 23: Results when data were simulated under a shared random effects model ($\gamma_1 = \gamma_2 = 0.9$) and $n = 50$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Shared (JS)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JS}}$
Shared (n=50)	***Continuous Process***										
	β_{10}: intercept	13.50	-0.013	0.410	0.412	0.94	-0.020	0.411	0.414	0.95	1.01
	β_{11}: time	-0.56	0.001	0.027	0.028	0.95	0.001	0.028	0.030	0.95	1.09
	β_{12}: treatment	-3.35	-0.005	0.426	0.409	0.96	0.001	0.429	0.415	0.95	1.02
	β_{13}: time×treatment	-0.57	0.000	0.030	0.028	0.96	0.000	0.031	0.030	0.95	1.10
	Ordinal Process										
	α_2: threshold	0.45	0.032	0.074	0.079	0.92	0.029	0.074	0.079	0.92	0.97
	β_{20}: intercept	1.70	0.062	0.403	0.419	0.94	0.051	0.409	0.422	0.94	1.02
	β_{21}: time	-0.27	-0.011	0.036	0.035	0.95	-0.008	0.038	0.036	0.96	1.05
	β_{22}: treatment	0.65	-0.014	0.418	0.405	0.96	-0.019	0.421	0.414	0.95	1.02
	β_{23}: time×treatment	-0.10	0.000	0.033	0.032	0.95	0.000	0.035	0.035	0.95	1.17
	Survival Process										
	β_{30}: intercept	-3.31	-0.107	0.435	0.440	0.95	-0.046	0.433	0.450	0.95	0.95
	β_{31}: treatment	0.37	0.021	0.401	0.403	0.94	0.014	0.413	0.416	0.95	1.06
	Association Parameters & Variances										
	γ_1: $\mathbf{b}_{i2} \propto \mathbf{b}_{i1}$	0.90	0.062	0.109	0.106	0.92	-	-	-	-	-
	γ_2: $\mathbf{b}_{i3} \propto \mathbf{b}_{i1}$	0.90	0.064	0.151	0.150	0.93	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.036	1.348	1.364	0.96	0.026	1.349	1.382	0.96	1.00
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.75	-	-	-	-	0.571	1.660	1.650	0.92	-
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	4.75	-	-	-	-	0.264	2.303	2.417	0.95	-
	$\sigma^2_{e_1}$: error	7.40	0.043	0.445	0.466	0.95	0.038	0.450	0.472	0.94	1.02

Table 24: Results when data were simulated under a shared random effects model ($\gamma_1 = \gamma_2 = 0.9$) and $n = 100$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Shared (JS)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JS}}$
Shared (n=100)	***Continuous Process***										
	β_{10}: intercept	13.50	-0.024	0.293	0.291	0.96	-0.028	0.293	0.293	0.96	1.00
	β_{11}: time	-0.56	0.002	0.020	0.020	0.95	0.002	0.021	0.021	0.94	1.06
	β_{12}: treatment	-3.35	-0.003	0.285	0.290	0.94	-0.005	0.285	0.292	0.94	1.00
	β_{13}: time×treatment	-0.57	0.000	0.021	0.020	0.95	0.000	0.022	0.021	0.96	1.12
	Ordinal Process										
	α_2: threshold	0.45	0.014	0.054	0.055	0.94	0.012	0.054	0.055	0.94	0.98
	β_{20}: intercept	1.70	0.025	0.275	0.287	0.95	0.016	0.283	0.290	0.95	1.05
	β_{21}: time	-0.27	-0.006	0.024	0.024	0.93	-0.004	0.026	0.025	0.94	1.15
	β_{22}: treatment	0.65	0.017	0.268	0.280	0.95	0.011	0.278	0.286	0.95	1.07
	β_{23}: time×treatment	-0.10	-0.003	0.023	0.022	0.94	-0.002	0.025	0.024	0.95	1.17
	Survival Process										
	β_{30}: intercept	-3.31	-0.062	0.318	0.299	0.95	-0.048	0.332	0.315	0.94	1.07
	β_{31}: treatment	0.37	0.006	0.270	0.276	0.94	-0.008	0.281	0.291	0.94	1.08
	Association Parameters & Variances										
	γ_1: $\mathbf{b}_{i2} \propto \mathbf{b}_{i1}$	0.90	0.030	0.074	0.071	0.93	-	-	-	-	-
	γ_2: $\mathbf{b}_{i3} \propto \mathbf{b}_{i1}$	0.90	0.029	0.099	0.099	0.93	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.021	0.957	0.951	0.96	0.041	0.958	0.958	0.96	1.00
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	4.75	-	-	-	-	0.322	1.127	1.080	0.94	-
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	4.75	-	-	-	-	0.230	1.670	1.631	0.95	-
	$\sigma^2_{e_1}$: error	7.40	0.030	0.327	0.329	0.94	0.029	0.331	0.333	0.95	1.03

Table 25: Results when data were simulated under a shared random effects model ($\gamma_1 = \gamma_2 = 0.6$) and $n = 50$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Shared (JS)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JS}}$
Shared (n=50)	***Continuous Process***										
	β_{10}: intercept	13.50	-0.014	0.408	0.412	0.95	-0.020	0.412	0.415	0.95	1.02
	β_{11}: time	-0.56	0.001	0.028	0.029	0.95	0.001	0.028	0.030	0.95	1.05
	β_{12}: treatment	-3.35	-0.004	0.426	0.410	0.95	0.002	0.429	0.415	0.95	1.01
	β_{13}: time×treatment	-0.57	0.000	0.030	0.029	0.96	0.000	0.031	0.030	0.95	1.06
	Ordinal Process										
	α_2: threshold	0.45	0.025	0.070	0.072	0.94	0.023	0.071	0.072	0.94	1.00
	β_{20}: intercept	1.70	0.052	0.304	0.316	0.94	0.047	0.310	0.319	0.94	1.03
	β_{21}: time	-0.27	-0.009	0.031	0.031	0.94	-0.007	0.033	0.033	0.95	1.10
	β_{22}: treatment	0.65	-0.006	0.313	0.305	0.95	-0.006	0.325	0.312	0.94	1.08
	β_{23}: time×treatment	-0.10	-0.001	0.031	0.030	0.94	-0.001	0.034	0.032	0.95	1.19
	Survival Process										
	β_{30}: intercept	-3.31	-0.088	0.344	0.342	0.94	-0.049	0.334	0.345	0.95	0.90
	β_{31}: treatment	0.37	0.023	0.320	0.315	0.94	0.017	0.329	0.327	0.95	1.05
	Association Parameters & Variances										
	γ_1: $\mathbf{b}_{i2} \propto \mathbf{b}_{i1}$	0.60	0.037	0.072	0.074	0.92	-	-	-	-	-
	γ_2: $\mathbf{b}_{i3} \propto \mathbf{b}_{i1}$	0.60	0.050	0.128	0.127	0.93	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.044	1.348	1.364	0.95	0.025	1.349	1.381	0.96	1.00
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	2.11	-	-	-	-	0.249	0.706	0.733	0.93	-
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	2.11	-	-	-	-	0.268	1.280	1.366	0.93	-
	$\sigma^2_{e_1}$: error	7.40	0.049	0.447	0.470	0.95	0.038	0.450	0.472	0.95	1.01

Table 26: Results when data were simulated under a shared random effects model ($\gamma_1 = \gamma_2 = 0.6$) and $n = 100$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Shared (JS)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JS}}$
Shared (n=100)	***Continuous Process***										
	β_{10}: intercept	13.50	-0.027	0.292	0.291	0.96	-0.028	0.293	0.293	0.96	1.01
	β_{11}: time	-0.56	0.002	0.020	0.020	0.95	0.002	0.021	0.021	0.94	1.03
	β_{12}: treatment	-3.35	-0.003	0.287	0.291	0.95	-0.005	0.286	0.292	0.94	0.99
	β_{13}: time×treatment	-0.57	0.000	0.021	0.020	0.95	0.000	0.022	0.021	0.96	1.09
	Ordinal Process										
	α_2: threshold	0.45	0.013	0.048	0.050	0.95	0.012	0.048	0.050	0.95	0.97
	β_{20}: intercept	1.70	0.031	0.210	0.218	0.94	0.025	0.215	0.222	0.94	1.04
	β_{21}: time	-0.27	-0.006	0.022	0.022	0.94	-0.005	0.024	0.023	0.94	1.10
	β_{22}: treatment	0.65	0.014	0.212	0.213	0.95	0.011	0.212	0.216	0.96	1.01
	β_{23}: time×treatment	-0.10	-0.002	0.022	0.021	0.95	-0.002	0.023	0.022	0.96	1.08
	Survival Process										
	β_{30}: intercept	-3.31	-0.055	0.242	0.231	0.93	-0.043	0.247	0.238	0.94	1.02
	β_{31}: treatment	0.37	0.008	0.216	0.215	0.95	-0.003	0.224	0.226	0.95	1.07
	Association Parameters & Variances										
	γ_1: $\mathbf{b}_{i2} \propto \mathbf{b}_{i1}$	0.60	0.018	0.051	0.050	0.93	-	-	-	-	-
	γ_2: $\mathbf{b}_{i3} \propto \mathbf{b}_{i1}$	0.60	0.025	0.090	0.084	0.95	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.019	0.960	0.953	0.96	0.041	0.959	0.958	0.96	1.00
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	2.11	-	-	-	-	0.133	0.471	0.485	0.94	-
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	2.11	-	-	-	-	0.137	0.923	0.943	0.95	-
	$\sigma^2_{e_1}$: error	7.40	0.032	0.329	0.331	0.94	0.029	0.331	0.333	0.95	1.01

Table 27: Results when data were simulated under a shared random effects model ($\gamma_1 = \gamma_2 = -0.5$) and $n = 50$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

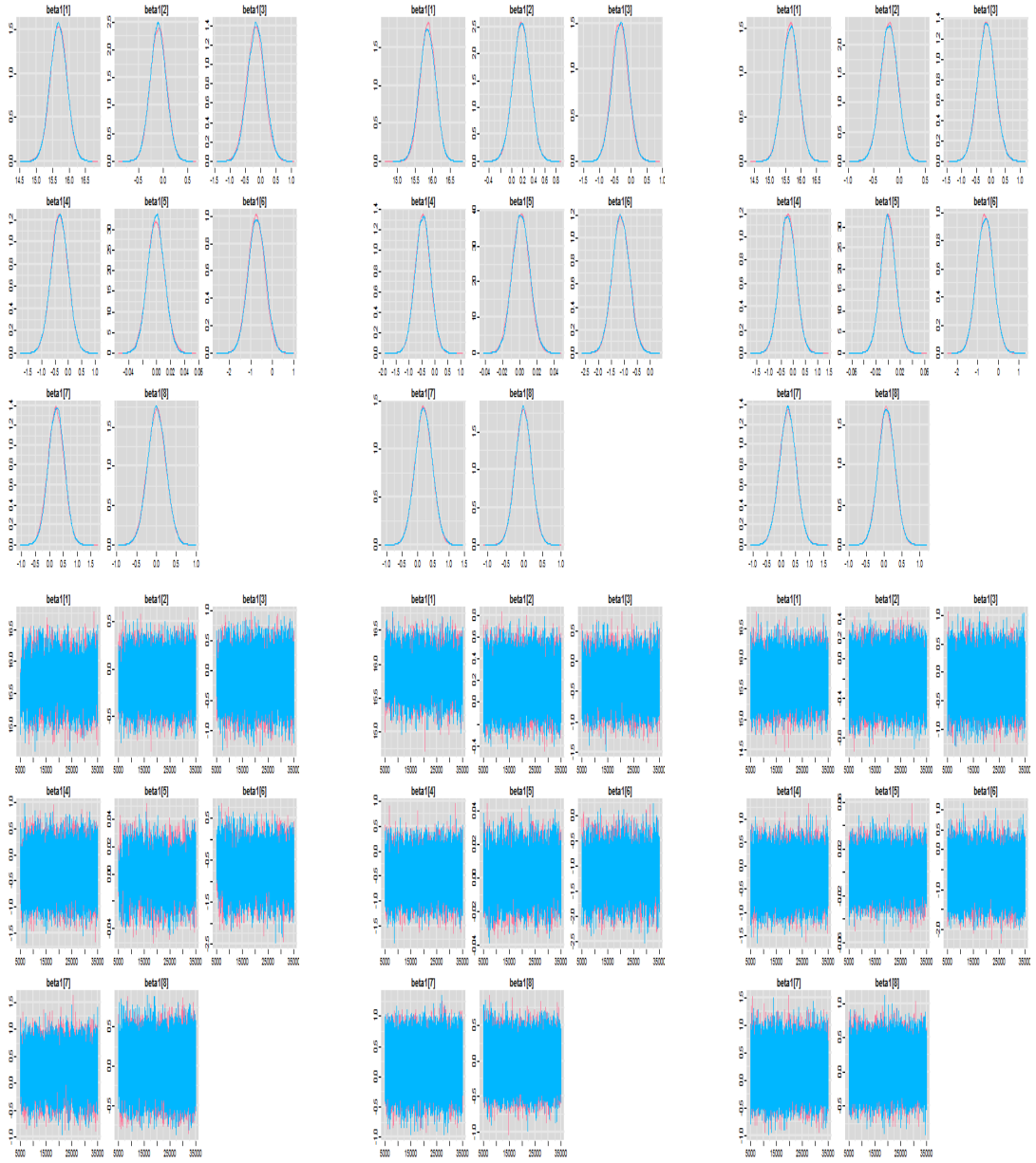
			Fitted Model								
			Joint-Shared (JS)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JS}}$
Shared (n=50)	***Continuous Process***										
	β_{10}: intercept	13.50	-0.017	0.414	0.414	0.95	-0.020	0.412	0.415	0.95	0.99
	β_{11}: time	-0.56	0.001	0.027	0.029	0.96	0.001	0.028	0.030	0.95	1.07
	β_{12}: treatment	-3.35	0.001	0.428	0.411	0.95	0.001	0.428	0.415	0.95	1.00
	β_{13}: time×treatment	-0.57	0.000	0.031	0.029	0.95	0.000	0.031	0.030	0.95	1.03
	Ordinal Process										
	α_2: threshold	0.45	0.019	0.068	0.069	0.94	0.017	0.068	0.069	0.95	0.98
	β_{20}: intercept	1.70	0.057	0.286	0.285	0.95	0.039	0.291	0.289	0.95	1.01
	β_{21}: time	-0.27	-0.009	0.032	0.030	0.95	-0.008	0.033	0.032	0.94	1.05
	β_{22}: treatment	0.65	-0.008	0.262	0.274	0.95	-0.007	0.267	0.280	0.96	1.04
	β_{23}: time×treatment	-0.10	0.001	0.029	0.029	0.95	0.001	0.032	0.031	0.95	1.16
	Survival Process										
	β_{30}: intercept	-3.31	-0.071	0.312	0.307	0.94	-0.058	0.316	0.317	0.94	1.01
	β_{31}: treatment	0.37	0.040	0.301	0.287	0.95	0.041	0.320	0.303	0.95	1.13
	Association Parameters & Variances										
	γ_1: $\mathbf{b}_{i2} \propto \mathbf{b}_{i1}$	-0.50	-0.025	0.065	0.066	0.94	-	-	-	-	-
	γ_2: $\mathbf{b}_{i3} \propto \mathbf{b}_{i1}$	-0.50	-0.036	0.124	0.119	0.94	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	-0.041	1.340	1.366	0.96	0.026	1.349	1.382	0.96	1.01
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	1.47	-	-	-	-	0.143	0.508	0.517	0.94	-
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	1.47	-	-	-	-	0.293	0.850	1.081	0.93	-
	$\sigma^2_{e_1}$: error	7.40	0.047	0.443	0.471	0.95	0.038	0.450	0.472	0.95	1.03

Table 28: Results when data were simulated under a shared random effects model ($\gamma_1 = \gamma_2 = -0.5$) and $n = 100$: MCSD, SD, and CP stand for Monte Carlo Standard Deviation, Posterior standard deviation, and Coverage probability of 95% HPD intervals, respectively.

			Fitted Model								
			Joint-Shared (JS)				Separate (SP)				
True Model (n)	Parameter	Truth	Bias	MCSD	SD	CP	Bias	MCSD	SD	CP	$RE = \frac{MSE_{SP}}{MSE_{JS}}$
Shared (n=100)	***Continuous Process***										
	β_{10}: intercept	13.50	-0.028	0.293	0.292	0.95	-0.028	0.292	0.293	0.96	0.99
	β_{11}: time	-0.56	0.002	0.020	0.020	0.94	0.002	0.021	0.021	0.94	1.03
	β_{12}: treatment	-3.35	-0.011	0.285	0.291	0.95	-0.005	0.286	0.292	0.94	1.00
	β_{13}: time×treatment	-0.57	0.001	0.021	0.020	0.94	0.000	0.022	0.021	0.96	1.07
	Ordinal Process										
	α_2: threshold	0.45	0.012	0.049	0.049	0.95	0.011	0.049	0.049	0.95	0.99
	β_{20}: intercept	1.70	0.039	0.207	0.200	0.94	0.028	0.206	0.201	0.94	0.98
	β_{21}: time	-0.27	-0.004	0.022	0.021	0.94	-0.003	0.023	0.022	0.95	1.01
	β_{22}: treatment	0.65	0.014	0.188	0.193	0.95	0.012	0.189	0.195	0.95	1.02
	β_{23}: time×treatment	-0.10	-0.001	0.020	0.020	0.94	-0.001	0.021	0.021	0.94	1.08
	Survival Process										
	β_{30}: intercept	-3.31	-0.036	0.214	0.209	0.95	-0.026	0.218	0.215	0.95	1.02
	β_{31}: treatment	0.37	0.023	0.190	0.195	0.94	0.013	0.202	0.206	0.94	1.11
	Association Parameters & Variances										
	γ_1: $\mathbf{b}_{i2} \propto \mathbf{b}_{i1}$	-0.50	-0.014	0.044	0.045	0.94	-	-	-	-	-
	γ_2: $\mathbf{b}_{i3} \propto \mathbf{b}_{i1}$	-0.50	-0.018	0.082	0.080	0.95	-	-	-	-	-
	$\sigma^2_{b_1}$: $\mathbf{b}_{i1}$	5.87	0.010	0.959	0.951	0.96	0.041	0.959	0.959	0.96	1.00
	$\sigma^2_{b_2}$: $\mathbf{b}_{i2}$	1.47	-	-	-	-	0.086	0.331	0.349	0.95	-
	$\sigma^2_{b_3}$: $\mathbf{b}_{i3}$	1.47	-	-	-	-	0.128	0.694	0.741	0.94	-
	$\sigma^2_{e_1}$: error	7.40	0.036	0.332	0.332	0.95	0.029	0.331	0.333	0.95	0.99

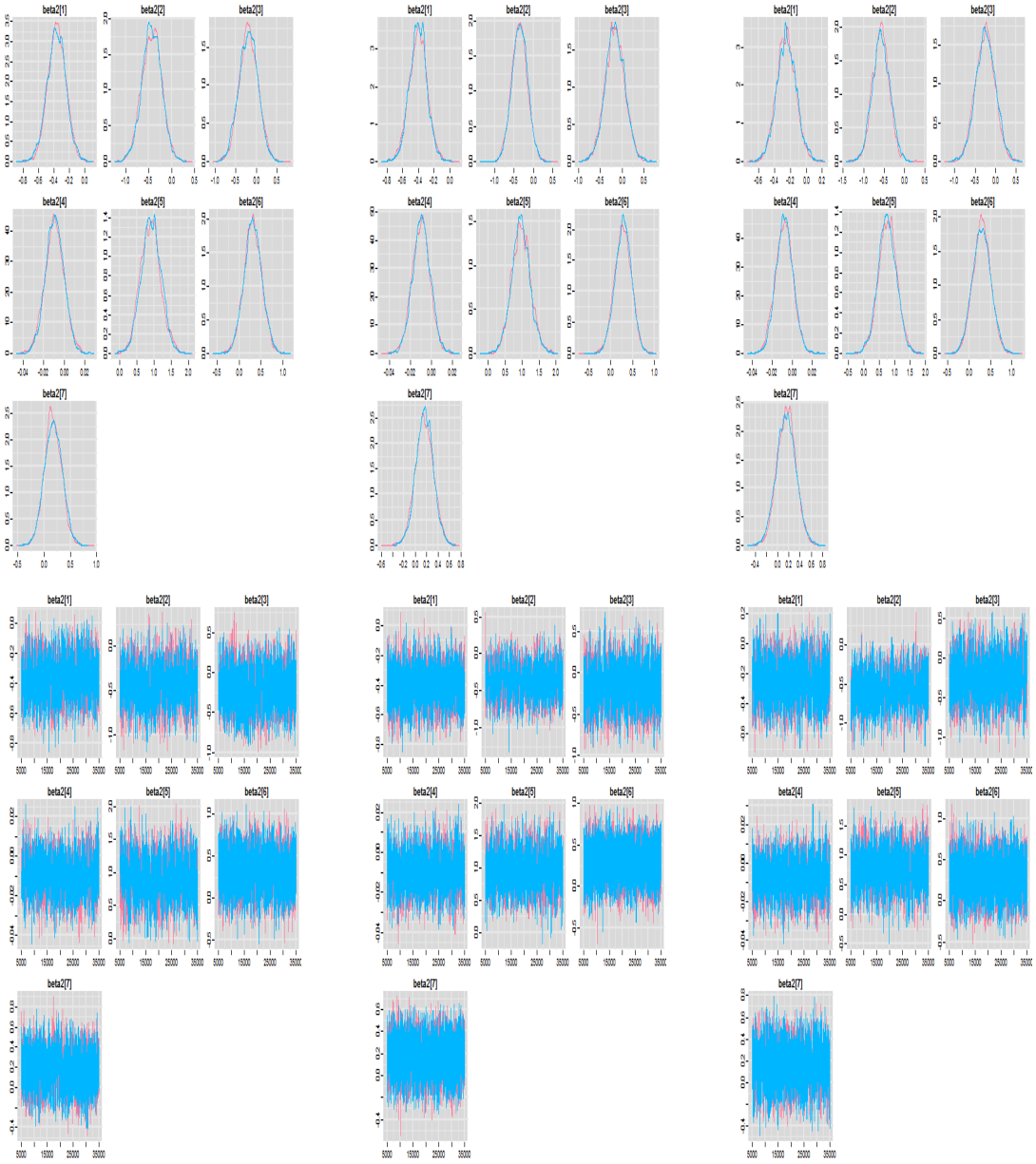
APPENDIX D

CONVERGENCE DIAGNOSTIC PLOTS



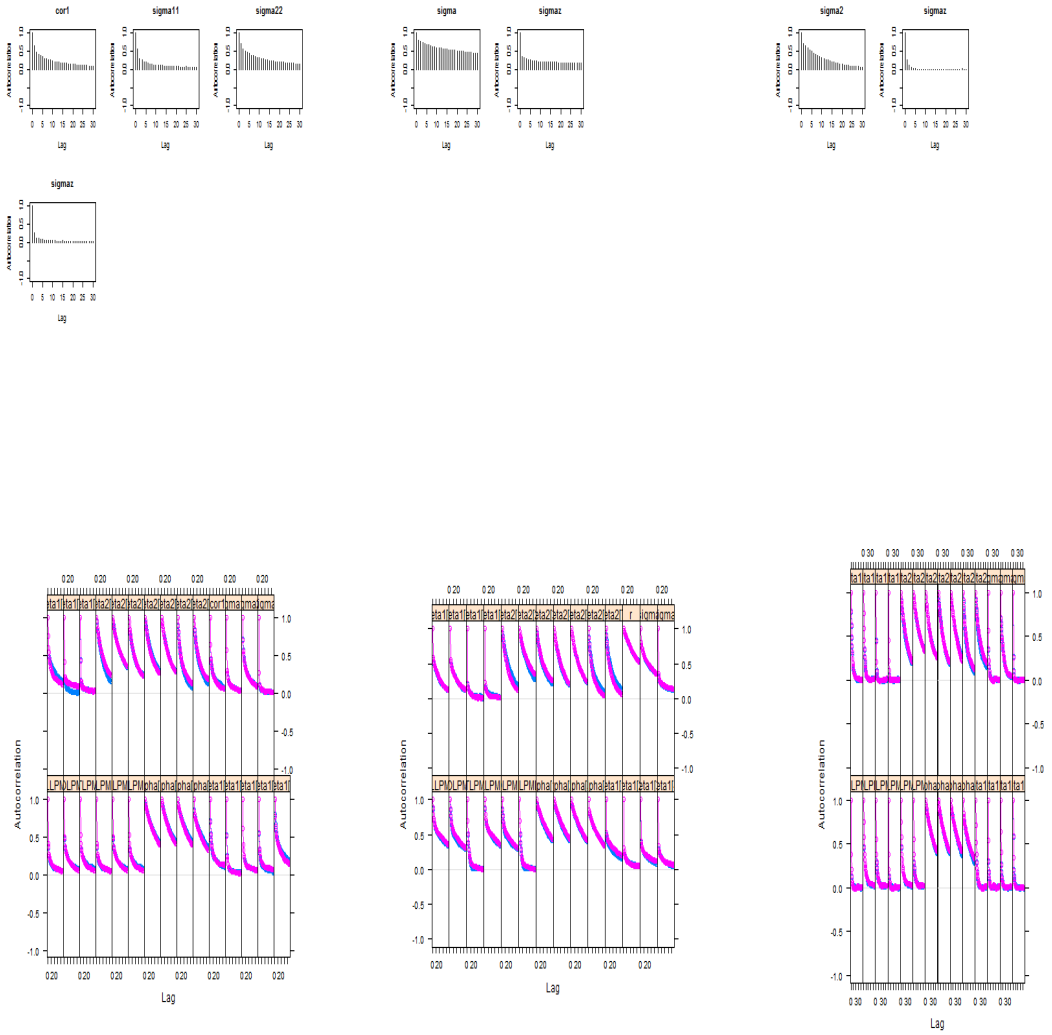
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_1: plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models to Ugandan diabetes data, respectively (results in Table 6).

Figure 5: Marginal posterior densities and trace plots of β_1 using joint and separate models



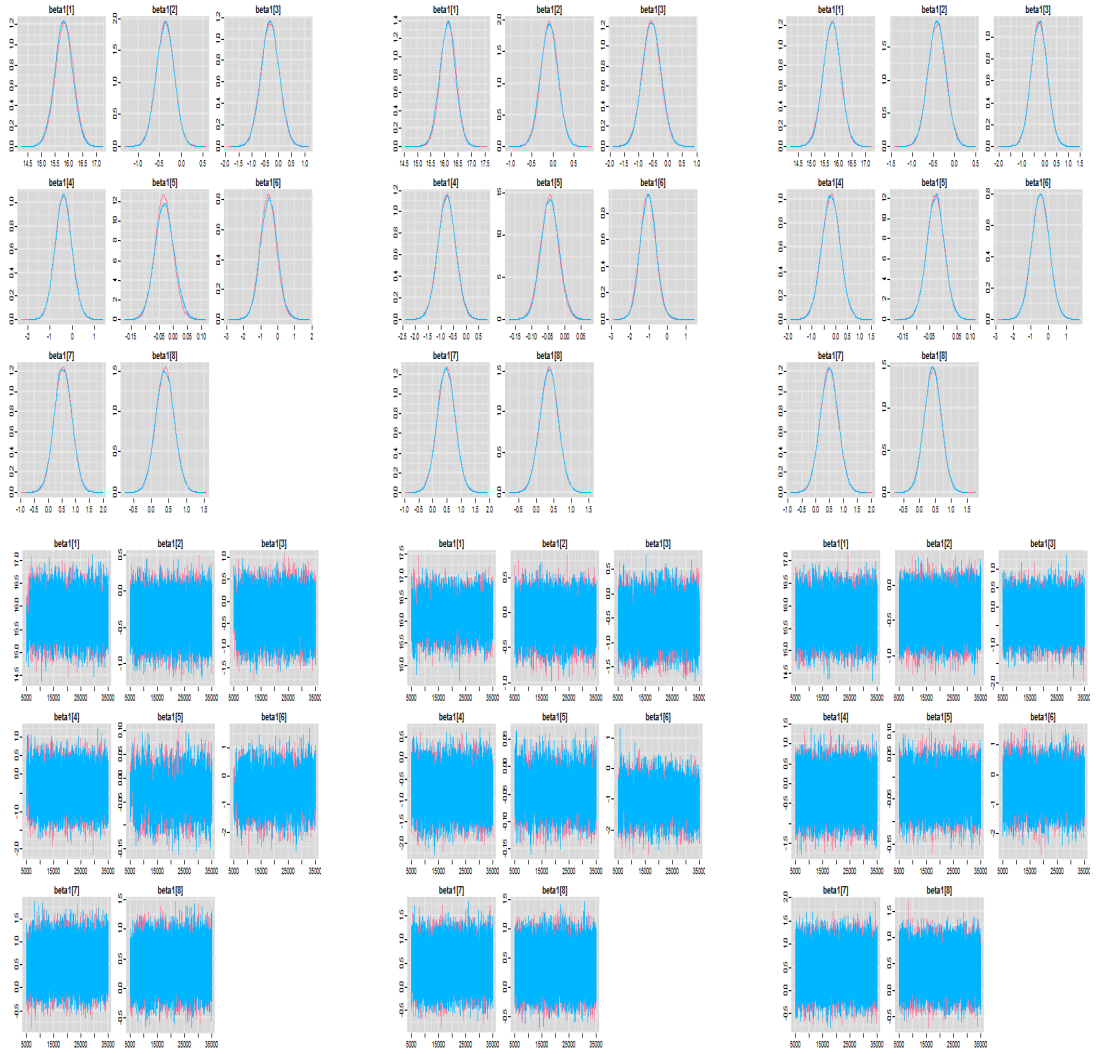
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_2: plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models to Ugandan diabetes data, respectively (results in Table 6).

Figure 6: Marginal posterior densities and trace plots of β_2 using joint and separate models



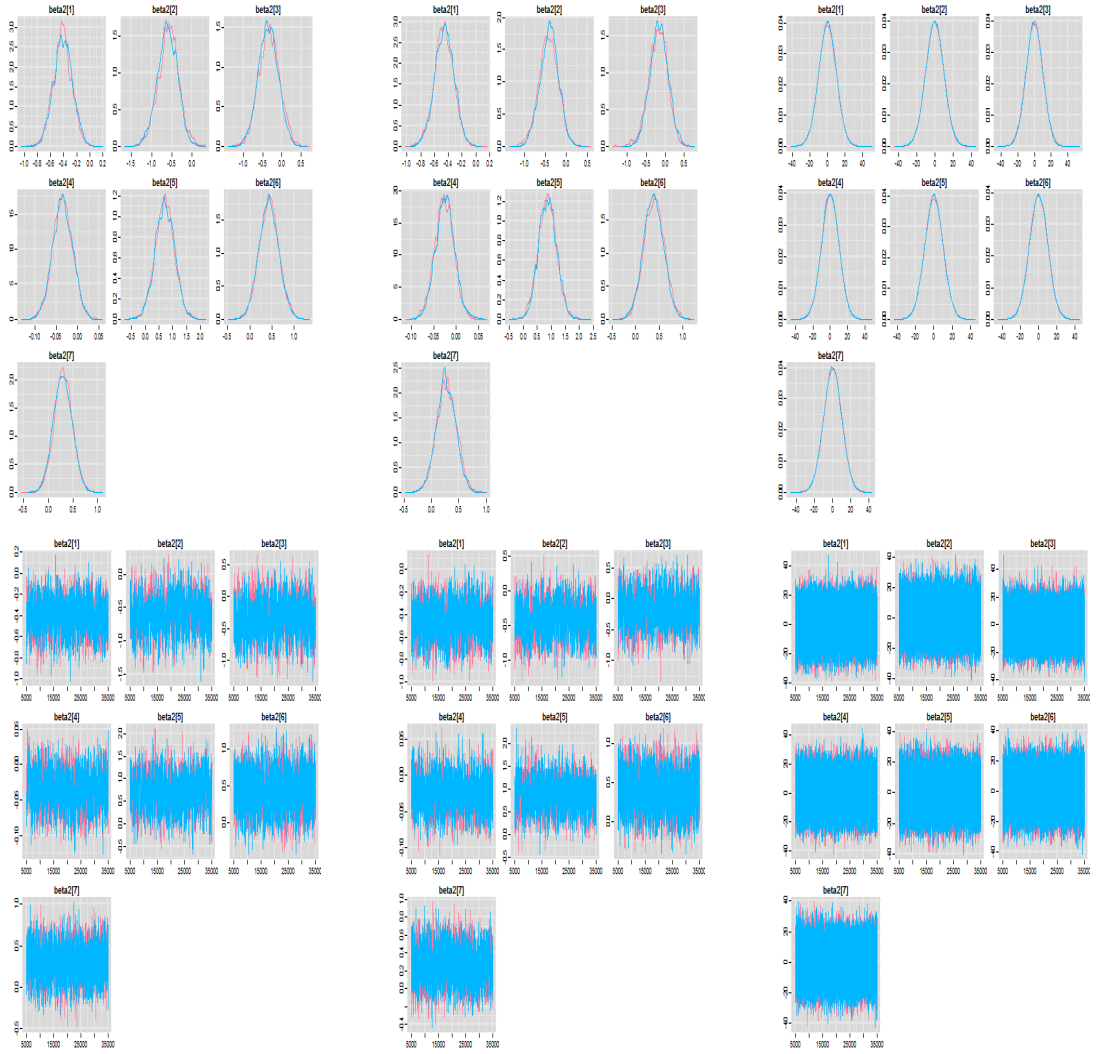
Autcorrelation function plots of parameter estimates using joint and separate models for Ugandan diabetes data (results in Table 6): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 7: Autcorrelation function plots of selected parameter estimates using joint and separate models



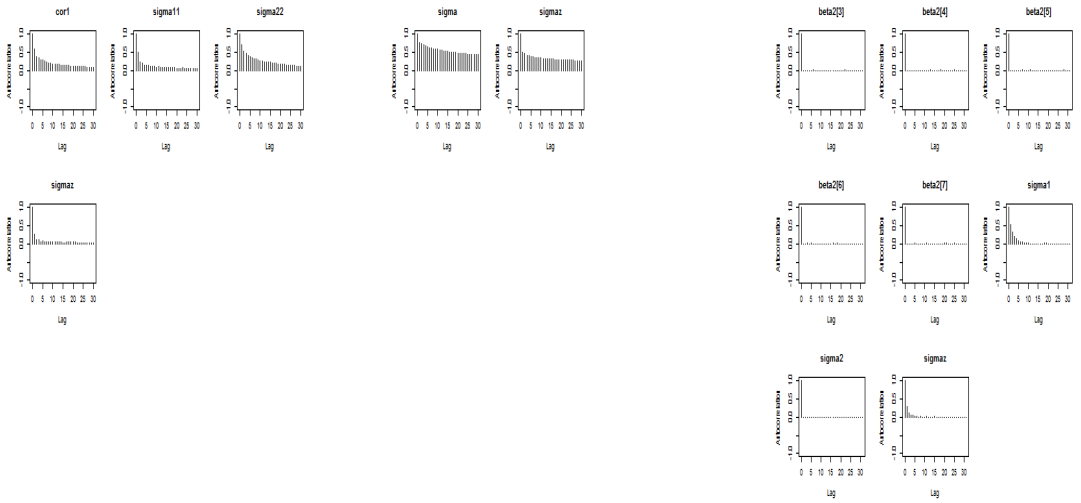
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_1: plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models to Ugandan Type 2 diabetes data, respectively (results in Table 22).

Figure 8: Marginal posterior densities and trace plots of β_1 using joint and separate models for Ugandan Type 2 diabetes data



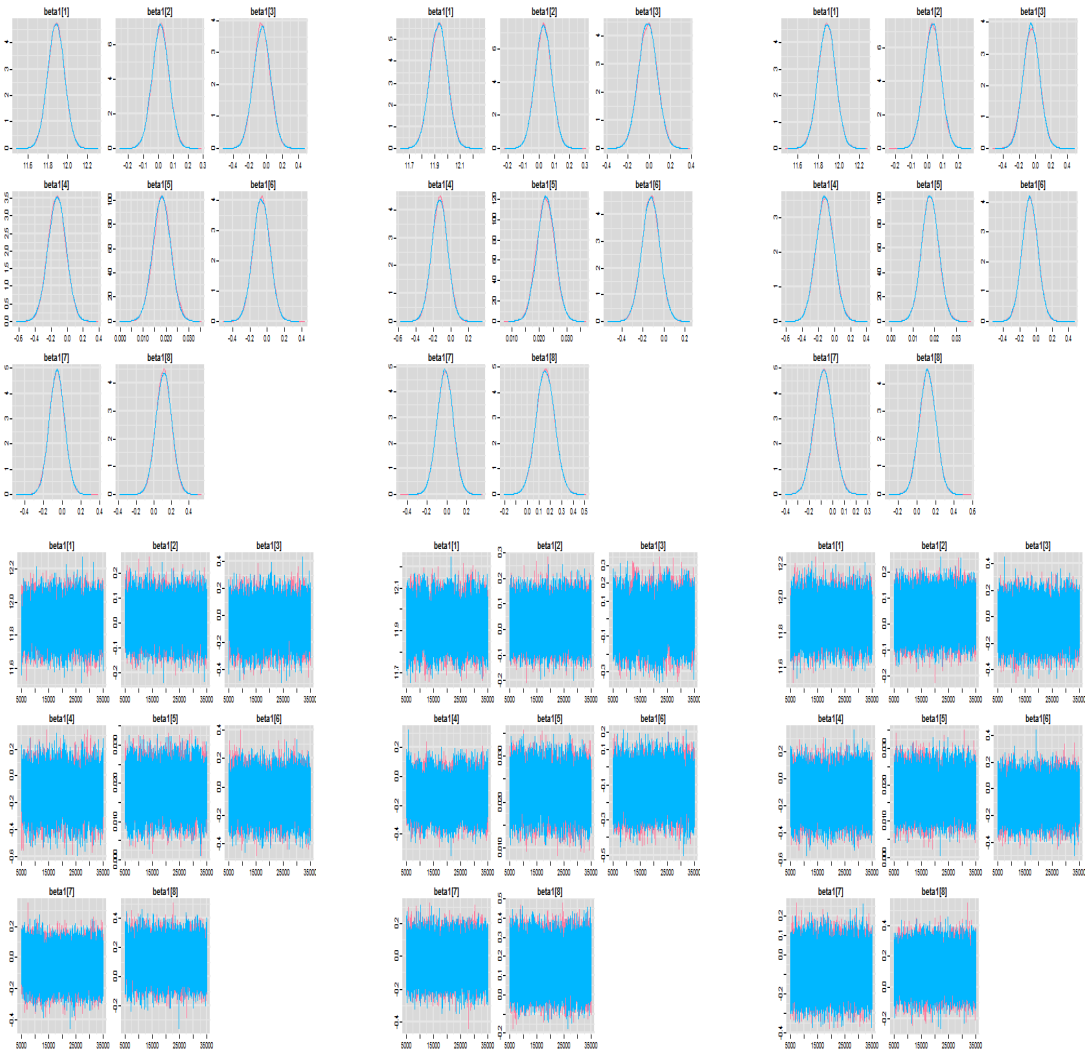
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_2: plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models to Ugandan Type 2 diabetes data, respectively (results in Table 22).

Figure 9: Marginal posterior densities and trace plots of β_2 using joint and separate models for Ugandan Type 2 diabetes data



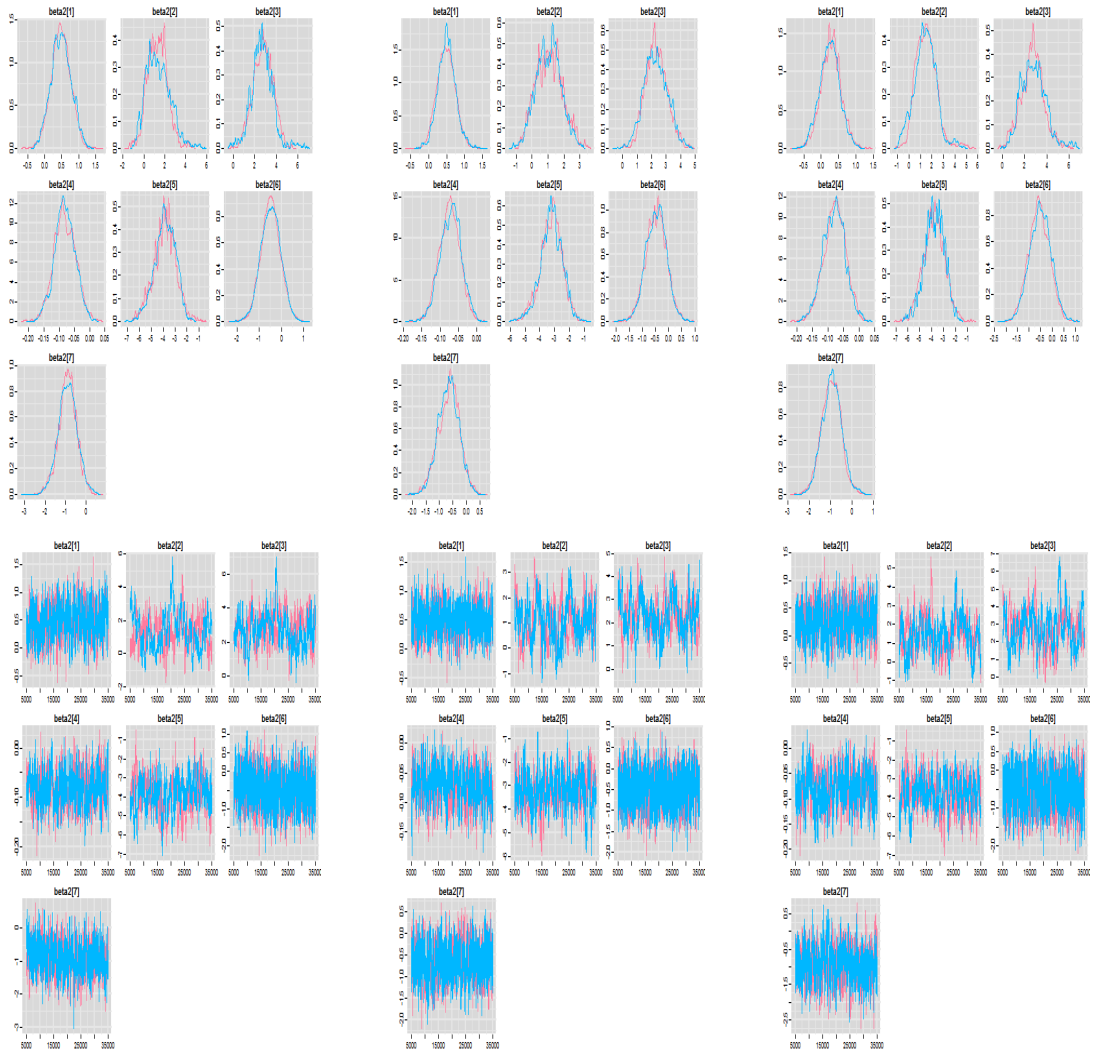
Autcorrelation function plots of parameter estimates using joint and separate models for Ugandan Type 2 diabetes data (results in Table 22): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 10: Autcorrelation function plots of selected parameter estimates using joint and separate models for Ugandan Type 2 diabetes data



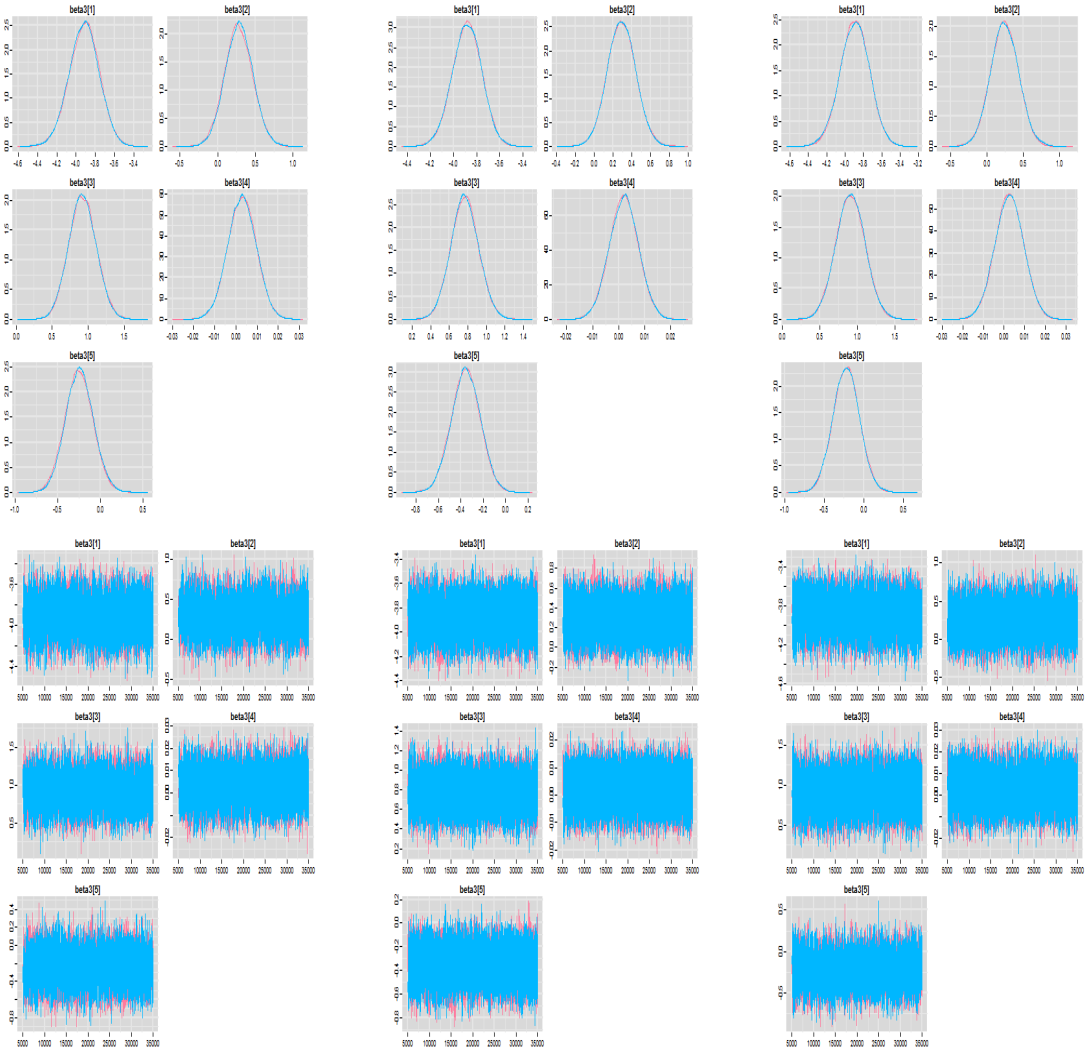
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_1 using joint and separate models for Ugandan Type 2 diabetes data (results in Table 17): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 11: Marginal posterior densities and trace plots of β_1 using joint and separate models (Normal blood glucose $70 - 180$ mg/dl)



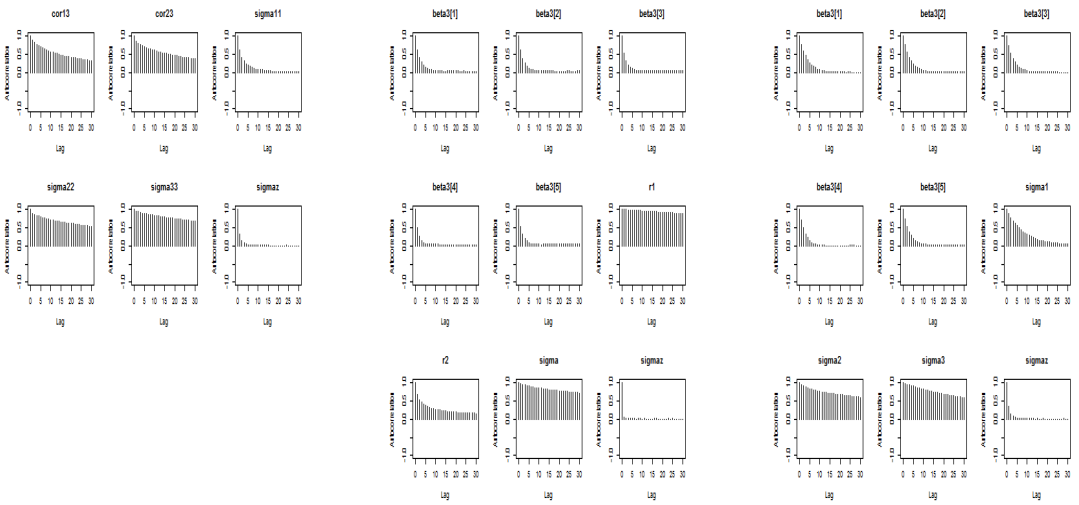
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_2 using joint and separate models for Ugandan Type 2 diabetes data (results in Table 17): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 12: Marginal posterior densities and trace plots of β_2 using joint and separate models (Normal blood glucose $70 - 180$ mg/dl)



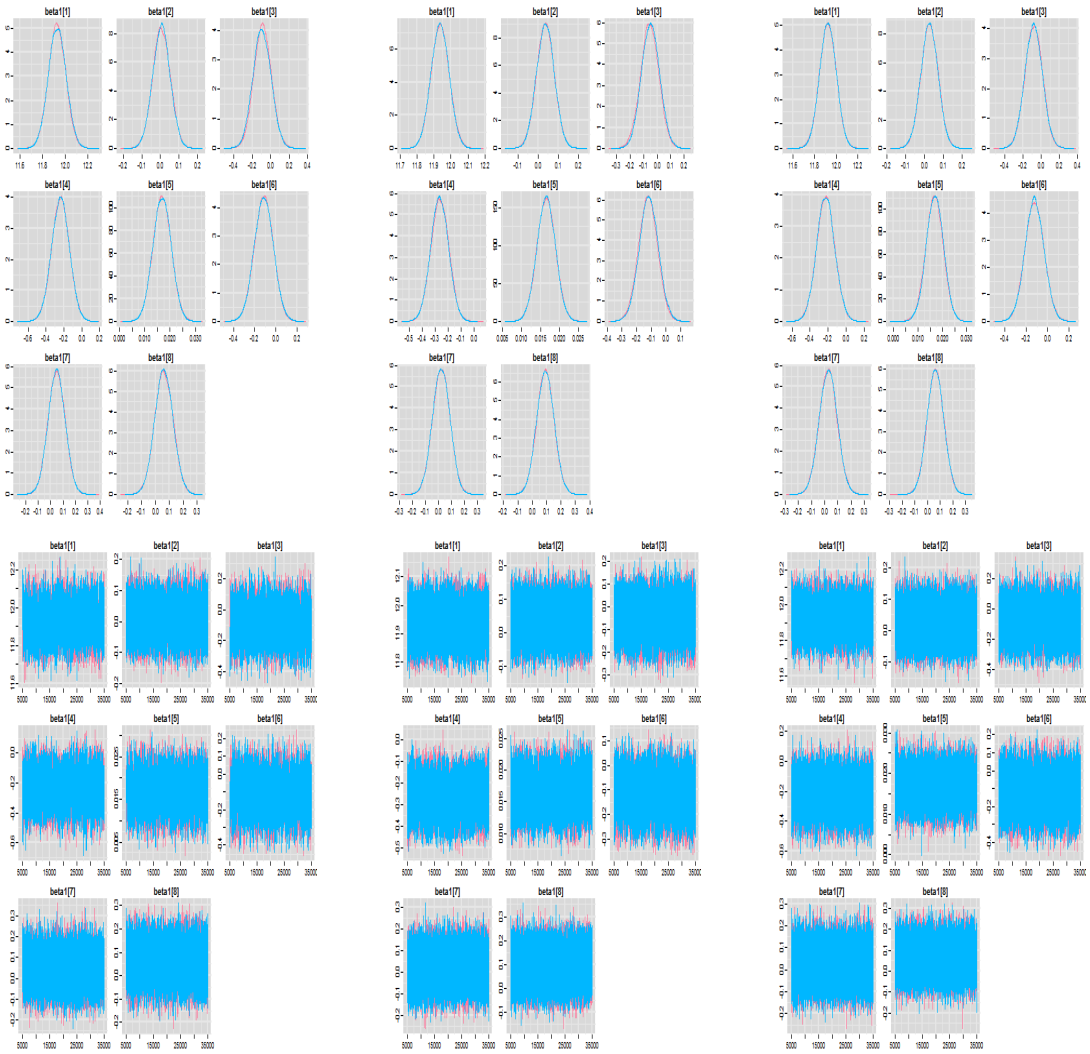
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_3 using joint and separate models for Ugandan Type 2 diabetes data (results in Table 17): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 13: Marginal posterior densities and trace plots of β_3 using joint and separate models (Normal blood glucose $70 - 180$ mg/dl)



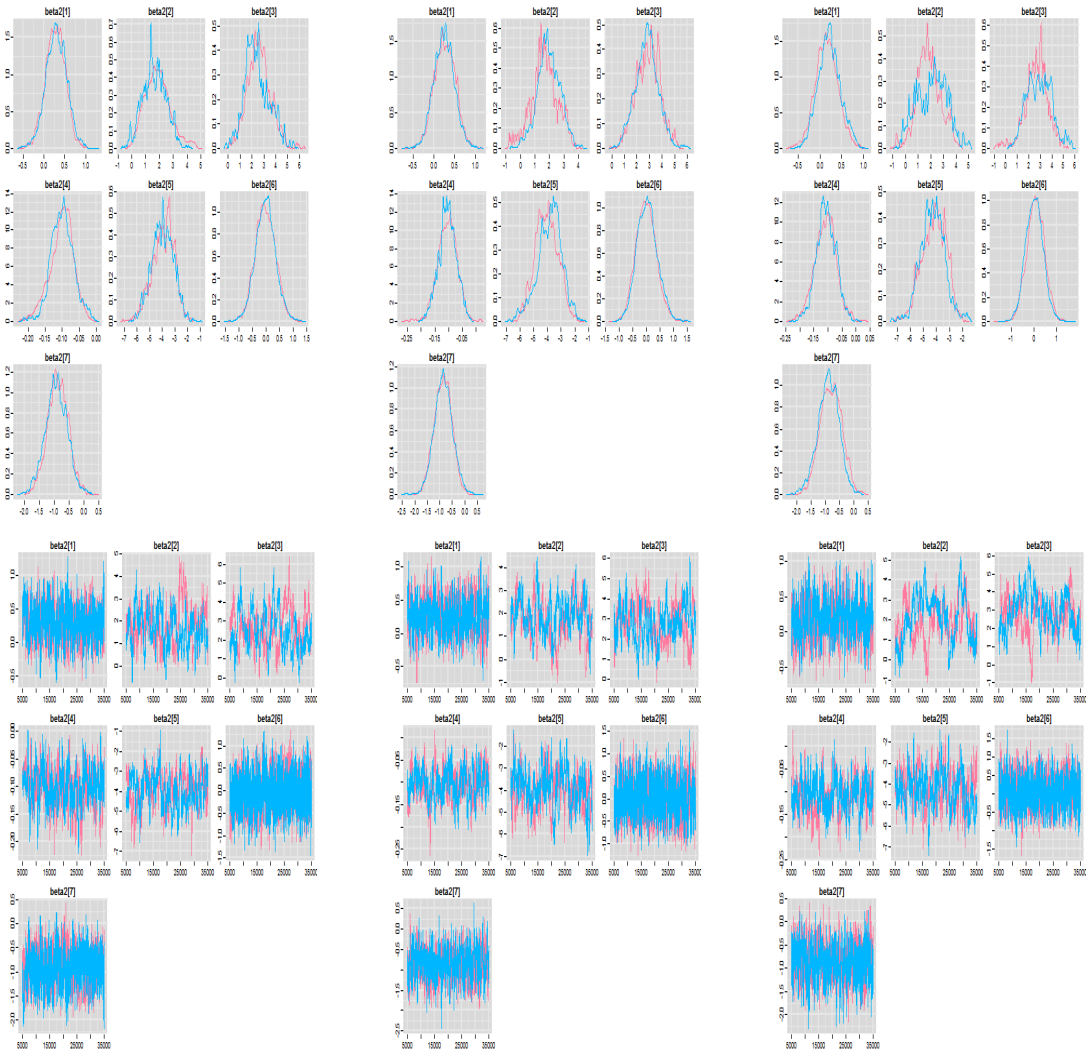
Autcorrelation function plots of parameter estimates using joint and separate models for Ugandan Type 2 diabetes data (results in Table 17): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 14: Autcorrelation function plots of selected parameter estimates using joint and separate models (Normal blood glucose $70 - 180$ mg/dl)



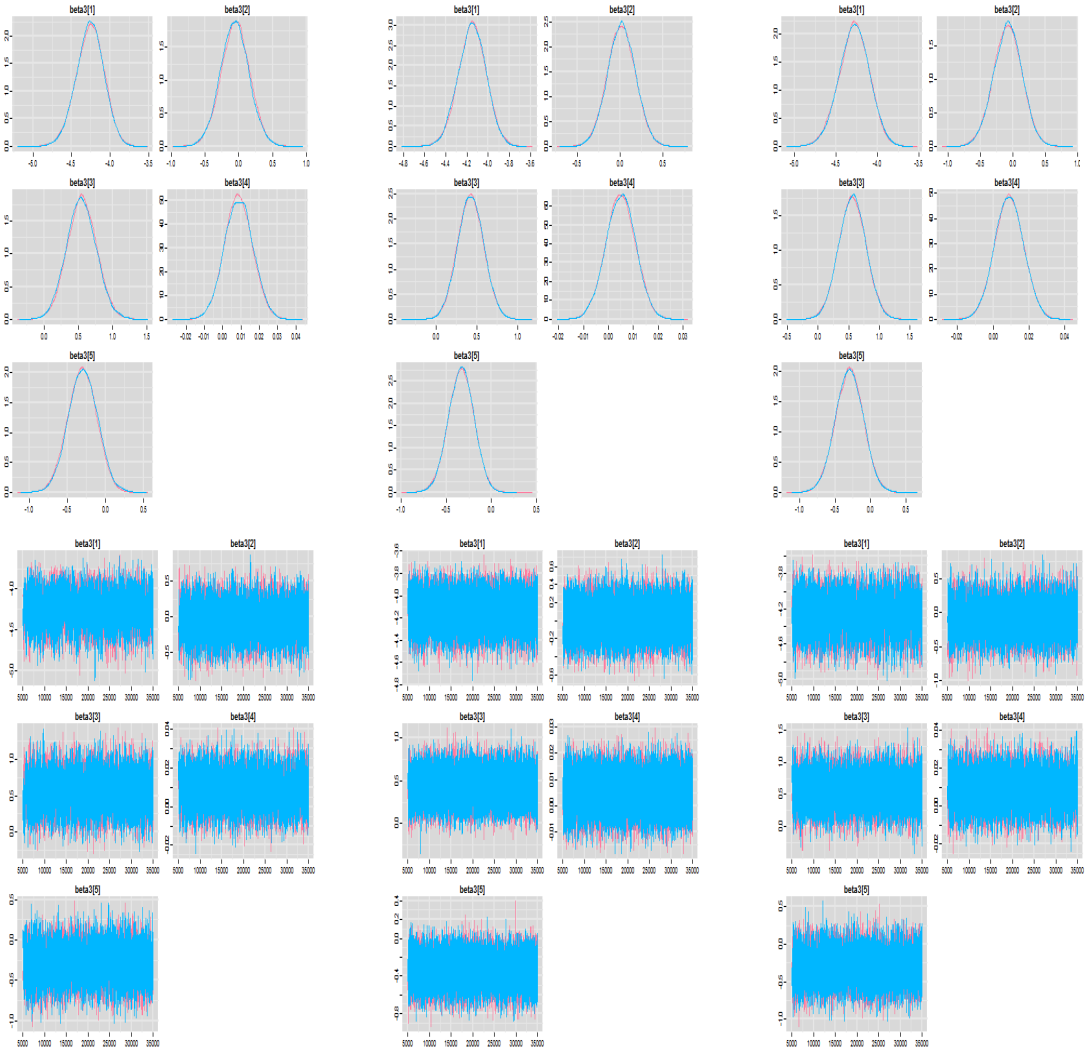
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_1 using joint and separate models for Ugandan Type 2 diabetes data (results in Table 18): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 15: Marginal posterior densities and trace plots of β_1 using joint and separate models (Normal blood glucose $70 - 130$ mg/dl)



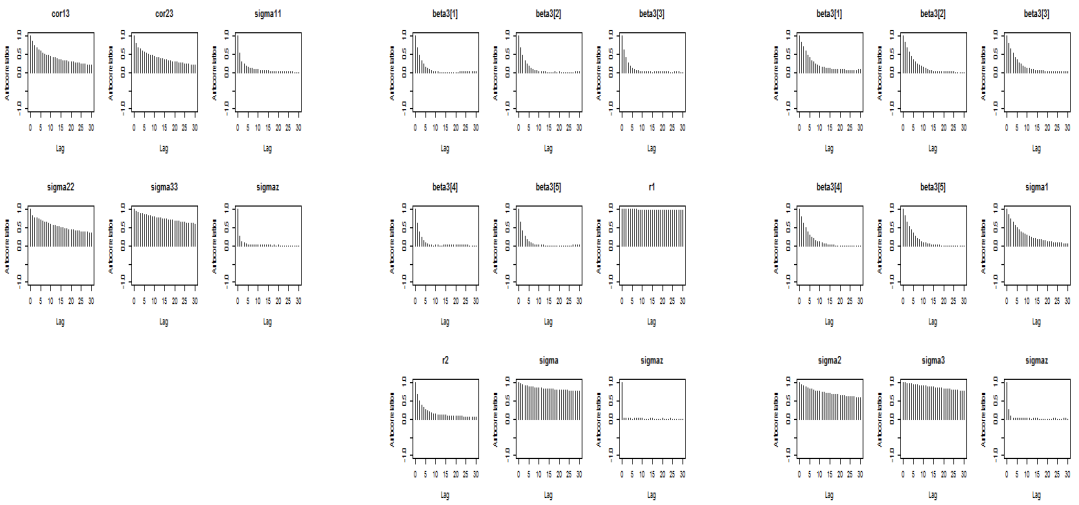
Marginal posterior densities (upper panel) and trace plots (lower panel) of β_2 using joint and separate models for Ugandan Type 2 diabetes data (results in Table 18): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 16: Marginal posterior densities and trace plots of β_2 using joint and separate models (Normal blood glucose $70 - 130$ mg/dl)



Marginal posterior densities (upper panel) and trace plots (lower panel) of β_3 using joint and separate models for Ugandan Type 2 diabetes data (results in Table 18): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 17: Marginal posterior densities and trace plots of β_3 using joint and separate models (Normal blood glucose $70 - 130$ mg/dl)



Autcorrelation function plots of parameter estimates using joint and separate models for Ugandan Type 2 diabetes data (results in Table 18): plots in the left, middle, and right panels are from fitting, joint correlated random effects, joint shared random effects, and separate models, respectively.

Figure 18: Autcorrelation function plots of selected parameter estimates using joint and separate models (Normal blood glucose $70 - 130$ mg/dl)

APPENDIX E

CODES USED TO FIT THE JOINT AND SEPARATE MODELS TO DIABETES DATA IN CHAPTER 3

The following codes were used to fit the joint and separate models to diabetes data (Results in Table 6)

```
#########################################################
##### Model 3.1: Joint correlated random effects #####
#########################################################
sink("diab-model_4-6-13.txt") ## save model to file
cat("
###################################
# OpenBUGS program              ###
###################################
model {# begin model
                for (j in 1:Npat){ # loop over subjects (random effects)
                U[j,1:2]~dmnorm(U0[],tau[,])
                U1[j]<-U[j,1]
                U2[j]<-U[j,2]
                }

                ## Means for the continuous  and ordinal variables
                for (i in 1:Nobs){ # begin loop over observations
                Z[i]<-sqrt(bglucse[i])
                Z[i]~dnorm(mu1[i],tauz)
                time[i]<-(month[i]-mean(month[]))/(sd(month[]))

                mu1[i]<-beta1[1]+beta1[2]*time[i]+beta1[3]*biguan[i]+beta1[4]*sulphony[i]
                +beta1[5]*(age[i]-mean(age[]))+beta1[6]*male[i]+beta1[7]*sulphony[i]*time[i]
                +beta1[8]*biguan[i]*time[i]+U1[patient[i]]

                mu2[i]<-beta2[1]*time[i]+beta2[2]*biguan[i]+beta2[3]*sulphony[i]
                +beta2[4]*(age[i]-mean(age[]))+beta2[5]*male[i]+beta2[6]*sulphony[i]*time[i]
                +beta2[7]*biguan[i]*time[i]+U2[patient[i]]

                ## cumulative logistic probabilities for the ordinal variable

                        logit(Q[i, 1])<-alpha[1]-mu2[i]
                        p[i,1]<-Q[i,1]

                        for (k in 2:4) {
                        logit(Q[i,k])<-alpha[k]-mu2[i]
                        p[i,k]<-Q[i,k] - Q[i,k-1]
```

```
                        }
                        p[i, 5]<-1- Q[i, 4]
                        ursugar[i]~dcat(p[i, 1:5])

                }# end loop over observations

                #priors for the threshold parameters for the ordinal outcome
                alpha[1]~dnorm(0, 1.0E-06) T(, alpha[2])
                alpha[2]~dnorm(0, 1.0E-06) T(alpha[1], alpha[3])
                alpha[3]~dnorm(0, 1.0E-06) T(alpha[2], alpha[4])
                alpha[4]~dnorm(0, 1.0E-06) T(alpha[3], )

                #construct variance for the error
                sigmaz<-1/tauz

                # construct variance-convariance matrix for the random effects
                sigma1[1:2,1:2]<-inverse(tau[,])
                sigma11<-sigma1[1,1]
                sigma22<-sigma1[2,2]
                sigma12<-sigma1[1,2]
                cor1<-sigma12/(sqrt(sigma11*sigma22))

                #prior for error precision
                tauz~dgamma(0.001,0.001)

                #prior for the precision matrix for the random effects
                tau[1:2,1:2]~dwish(R1[,],3)

                #prior for the regression coefficients
                beta1[1:8]~dmnorm(betamu1 [],Sigma1[,])# continuous outcome
                beta2[1:7]~dmnorm(betamu2 [],Sigma2[,])# ordinal outcome
}# end model
 ",fill=TRUE)
  sink()
####################################
# R program                        ###
####################################
#Bundle data
betamu1<-c(0,0,0,0,0,0,0,0)
betamu2<-c(0,0,0,0,0,0,0)
Sigma1<-diag(0.01,nrow=8,ncol=8)
Sigma2<-diag(0.01,nrow=7,ncol=7)
U0<-c(0,0)
R1<-diag(1,nrow=2, ncol=2)

# Place data in a list to be read by OpenBUGS
diab.data<-list(Nobs=1363,
                Npat=321,
                U0=U0,
                R1=R1,
                betamu1=betamu1,
                betamu2=betamu2,
                Sigma1=Sigma1,
                Sigma2=Sigma2,
                bglucse=data12[,1],
                ursugar=data12[,2],
                month=data12[,3],
                sulphony=data12[,4],
                biguan=data12[,5],
                age=data12[,6],
                male=data12[,8],
                patient=data12[,9])

# Initial values for MCMC sampling
inits<-function(){list(
                beta1=c(15.905,-0.024,-0.2362,-0.3229,-0.00015,-0.6978,0.0025,0.01017),
                tauz=1,
                alpha=c(-1.1512,-0.8419,-0.2657,1.260),
```

```
                beta2=c(-0.009,-0.6985,0.0145,-0.0052,0.7088,-0.0088,0.0128))
        }

# Parameters to estimate
params <-c("alpha","beta1","beta2","cor1","sigma11","sigma22","sigmaz")

# Start MCMC Sampling
outdiab08v1 <- BRugsFit(data=diab.data,
                        inits=inits,
                        parameters=params,
                        modelFile="diab-model_4-6-13.txt",
                        numChains=2,
                        nBurnin=5000,
                        nIter=30000,
                        nThin=1,
                        coda=FALSE,
                        DIC=TRUE,
                        digits=5,
                        BRugsVerbose=getOption("BRugsVerbose"))

#########################################################
######Model 3.2: Joint shared random effects    ######
#########################################################
sink("diab-model_4-7-13v2.txt")
cat("
###################################
# OpenBUGS program               ###
###################################
model {# begin model
                         # random effects
                for (j in 1:Npat) {U[j]~dnorm(0.0,tau)}

                ##Means for the continuous  and ordinal variables
                for (i in 1:Nobs){# begin loop over observations
                Z[i]<-sqrt(bglucse[i])
                Z[i]~dnorm(mu1[i], tauz)
                time[i]<-(month[i]-mean(month[]))/(sd(month[]))

                mu1[i]<-beta1[1] beta1[2]*time[i]+beta1[3]*biguan[i]+beta1[4]*sulphony[i]
                +beta1[5]*(age[i]-mean(age[]))+beta1[6]*male[i]+beta1[7]*sulphony[i]*time[i]
                +beta1[8]*biguan[i]*time[i]+U[patient[i]]

                mu2[i]<-beta2[1]*time[i]+beta2[2]*biguan[i]+beta2[3]*sulphony[i]
                +beta2[4]*(age[i]-mean(age[]))+beta2[5]*male[i]+beta2[6]*sulphony[i]*time[i]
                +beta2[7]*biguan[i]*time[i]+r*(U[patient[i]])

                 ## cumulative logistic probabilities for the ordinal outcome
                        logit(Q[i, 1])<-alpha[1]-mu2[i]
                        p[i,1]<-Q[i,1]

                        for (k in 2:4) {
                        logit(Q[i,k])<-alpha[k]-mu2[i]
                        p[i,k]<-Q[i,k]-Q[i,k-1]
                        }

                        p[i, 5]<-1-Q[i, 4]
                        ursugar[i]~dcat(p[i, 1:5])

                }# end loop over observations

                # prior for the threshold parameters for the ordinal outcome
                alpha[1]~dnorm(0, 1.0E-06) T(, alpha[2])
                alpha[2]~dnorm(0, 1.0E-06) T(alpha[1], alpha[3])
                alpha[3]~dnorm(0, 1.0E-06) T(alpha[2], alpha[4])
                alpha[4]~dnorm(0, 1.0E-06) T(alpha[3], )

                # construct error variance and random effects variance
                sigmaz<-1/tauz
```

```
                sigma<-1/tau

                # priors for precision
                tau~dgamma(0.001,0.001)
                tauz~dgamma(0.001,0.001)

                # priors for regresion coefficients
                beta1[1:8]~dmnorm(betamu1[],Sigma1[,])
                beta2[1:7]~dmnorm(betamu2[],Sigma2[,])

                # prior for the association parameter under shared random effects
                r~dnorm(0,0.1)

 }# end model
 ",fill=TRUE)
 sink()
####################################
# R program                        ###
####################################
# Bundle data
betamu1<-c(0,0,0,0,0,0,0,0)
betamu2<-c(0,0,0,0,0,0,0)
Sigma1<-diag(0.01,nrow=8,ncol=8)
Sigma2<-diag(0.01,nrow=7,ncol=7)

#Place data in a list to be read by OpenBUGS
diab.data<-list(Nobs=1363,
                Npat=321,
                betamu1=betamu1,
                betamu2=betamu2,
                Sigma1=Sigma1,
                Sigma2=Sigma2,
                bglucse=data12[,1],
                ursugar=data12[,2],
                month=data12[,3],
                sulphony=data12[,4],
                biguan=data12[,5],
                age=data12[,6],
                male=data12[,8],
                patient=data12[,9])

# Initial values for MCMC sampling
inits<-function(){list(
                beta1=c(15.905,-0.024,-0.2362,-0.3229,-0.00015,-0.6978,0.0025,0.01017),
                tauz=1,
                tau=1,
                r=0.0,
                alpha=c(-1.1512,-0.8419,-0.2657,1.260),
                beta2=c(-0.009,-0.6985,0.0145,-0.0052,0.7088,-0.0088,0.0128))
        }

# Parameters to estimate
params <- c("alpha","beta1","beta2","r","sigma","sigmaz")

# Start MCMC Sampling
outdiab08v3 <- BRugsFit(data=diab.data,
                        inits=inits,
                        parameters=params,
                        modelFile="diab-model_4-6-13v2.txt",
                        numChains=2,
                        nBurnin=5000,
                        nIter=30000,
                        nThin=1,
                        coda=FALSE,
                        DIC=TRUE,
                        digits=5,
                        BRugsVerbose=getOption("BRugsVerbose"))
```

```
#######################################################
##### Model 3.3: Separate models                  ######
#######################################################
sink("diab-model_4-7-13v3.txt")
cat("
###################################
# OpenBUGS program              ###
###################################
model {# begin model

                for (j in 1:Npat)  {# loop over subjects (random effects)
                U1[j]~dnorm(0.0,tau1)
                U2[j]~dnorm(0.0,tau2)
                }

                ## Means for the continuous  and ordinal variables
                for (i in 1:Nobs){# begin loop over observations
                Z[i]<-sqrt(bglucse[i])
                Z[i]~dnorm(mu1[i], tauz)
                time[i]<-(month[i]-mean(month[]))/(sd(month[]))

                mu1[i]<-beta1[1]+beta1[2]*time[i]+beta1[3]*biguan[i]+beta1[4]*sulphony[i]
                +beta1[5]*(age[i]-mean(age[]))+beta1[6]*male[i]+beta1[7]*sulphony[i]*time[i]
                +beta1[8]*biguan[i]*time[i]+U1[patient[i]]

                mu2[i]<-beta2[1]*time[i]+beta2[2]*biguan[i]+beta2[3]*sulphony[i]
                +beta2[4]*(age[i]-mean(age[]))+beta2[5]*male[i]+beta2[6]*sulphony[i]*time[i]
                +beta2[7]*biguan[i]*time[i]+U2[patient[i]]

                ## cumulative logistic probabilities for the ordinal outcome
                        logit(Q[i, 1])<-alpha[1]-mu2[i]
                        p[i,1]<-Q[i,1]

                        for (k in 2:4) {
                        logit(Q[i,k])<-alpha[k]-mu2[i]
                        p[i,k]<-Q[i,k]-Q[i,k-1]
                        }
                        p[i, 5]<-1-Q[i, 4]
                        ursugar[i]~dcat(p[i, 1:5])

                }# end loop over observations

                # prior for the threshold parameters for the ordinal outcome
                alpha[1]~dnorm(0, 1.0E-06) T(, alpha[2])
                alpha[2]~dnorm(0, 1.0E-06) T(alpha[1], alpha[3])
                alpha[3]~dnorm(0, 1.0E-06) T(alpha[2], alpha[4])
                alpha[4]~dnorm(0, 1.0E-06) T(alpha[3], )

                # construct error variance and random effects variances
                sigmaz<-1/tauz
                sigma1<-1/tau1
                sigma2<-1/tau2

                # priors for precisions
                tauz~dgamma(0.001,0.001)
                tau1~dgamma(0.001,0.001)
                tau2~dgamma(0.001,0.001)

                # priors for regresion coefficients
                beta1[1:8]~dmnorm(betamu1[],Sigma1[,])
                beta2[1:7]~dmnorm(betamu2[],Sigma2[,])

} #end model
 ",fill=TRUE)
 sink()

###################################
# R program                     ###
```

```
###################################
# Bundle data
betamu1<-c(0,0,0,0,0,0,0,0)
betamu2<-c(0,0,0,0,0,0,0)
Sigma1<-diag(0.01,nrow=8,ncol=8)
Sigma2<-diag(0.01,nrow=7,ncol=7)

# Place data in a list to be read by OpenBUGS
diab.data<-list(Nobs=1363,
                Npat=321,
                betamu1=betamu1,
                betamu2=betamu2,
                Sigma1=Sigma1,
                Sigma2=Sigma2,
                bglucse=data12[,1],
                ursugar=data12[,2],
                month=data12[,3],
                sulphony=data12[,4],
                biguan=data12[,5],
                age=data12[,6],
                male=data12[,8],
                patient=data12[,9])

# Initial values for MCMC sampling
inits<-function(){list(
                beta1=c(15.905,-0.024,-0.2362,-0.3229,-0.00015,-0.6978,0.0025,0.01017),
                tauz=1,
                tau1=1,
                tau2=1,
                alpha=c(-1.1512,-0.8419,-0.2657,1.260),
                beta2=c(-0.009,-0.6985,0.0145,-0.0052,0.7088,-0.0088,0.0128))
        }

# Parameters to estimate
params <- c("alpha","beta1","beta2","sigma1","sigma2","sigmaz")

# Start MCMC Sampling
outdiab08v4 <- BRugsFit(data=diab.data,
                        inits=inits,
                        parameters=params,
                        modelFile="diab-model_4-6-13v3.txt",
                        numChains=2,
                        nBurnin=5000,
                        nIter=30000,
                        nThin=1,
                        coda=FALSE,
                        DIC=TRUE,
                        digits=5,
                        BRugsVerbose=getOption("BRugsVerbose"))
```

APPENDIX F

CODES USED TO FIT THE JOINT AND SEPARATE MODELS TO DIABETES DATA IN CHAPTER 4

The following codes were used to fit the joint and separate models to diabetes data (Results in Table 17)

```
long.data ## Longitudinal data
surv.data ## Survival data

################################################
## Generating initial values for MCMC sampling #
################################################
# Continuous outcome
fit1<-lme(sqrt(ubp)~month + biguan + sulphony + month*biguan + month*sulphony
        + mage + male,data=long.data,method="ML",random=~1|patient)
b1 <- unlist(fit1$coef[1])

# Ordinal outcome
fit2< vglm(bmi_grp2~month + biguan + sulphony + month*biguan + month*sulphony
        + mage + male,family=cumulative(parallel=TRUE),data=long.data)
b2 <- coef(fit2, matrix=F)

# Survival outcome
temp<-ifelse(is.na(surv.data$surt ),surv.data$ cens2,surv.data$surt)
these <- temp>0
test.fit3<-survreg(Surv(temp,surv.data$cens2==0)~surv.data$biguan + surv.data$sulphony
        + surv.data$mage + surv.data$male,subset=these,dist="weibull")
b3<- -as.vector(test.fit3$coefficients)/test.fit3$scale

#######################################################
######Model 4.1: Joint correlated random effects #####
#######################################################
sink("diab-model_4-6-14.txt") # Save the model to file in your working directory
cat("
####################################
# OpenBUGS program              ###
####################################
model { # begin model

                for (i in 1:Nobs){  #Loop over observations to handle unbalanced data
                Z[i]<-sqrt(ubp[i])## transform systolic blood pressure
                Z[i]~dnorm(mu1[i], tauz)
                time[i]<-(month[i]-mean(month[]))/(sd(month[])) ## standardize time

                mu1[i]<-beta1[1]+beta1[2]*time[i]+beta1[3]*Biguan[i]+beta1[4]*Sulphony[i]
                +beta1[5]*(age[i]-mean(age[]))+beta1[6]*male[i]+beta1[7]*Biguan[i]*time[i]
```

```
+beta1[8]*Sulphony[i]*time[i]+U1[patient[i]]

mu2[i]<-beta2[1]*time[i]+beta2[2]*Biguan[i]+beta2[3]*Sulphony[i]
+beta2[4]*(age[i]-mean(age[]))+beta2[5]*male[i]+beta2[6]*Biguan[i]*time[i]
+beta2[7]*Sulphony[i]*time[i]+U2[patient[i]]

## Cumulative logistic probabilities for the ordinal outcome
        logit(Q[i, 1])<-alpha[1]-mu2[i]
        p[i,1] <-Q[i,1]

        for (k in 2:3){
        logit(Q[i,k]) <- alpha[k]-mu2[i]
        p[i,k]<-Q[i,k]-Q[i,k-1]
        }
        p[i, 4]<-1- Q[i, 3]
        bmi_grp2[i]~dcat(p[i, 1:4])

} #end loop over observations

for (j in 1:Npat) { # begin loop over subjects/patients
# Survival Model
log(mut[j])<-beta3[1]+beta3[2]*biguan[j]+beta3[3]*sulphony[j]
        +beta3[4]*(mage[j]-mean(mage[]))+beta3[5]*malet[j]+U3[j]

surt[j]~dweib(lamda,mut[j])C(cens[j],)

## Random effects
U[j,1:3]~dmnorm(U0[],tau[,])
U1[j]<-U[j,1]
U2[j]<-U[j,2]
U3[j]<-U[j,3]

} # end loop over subjects/patients

## Prior for the threshold parameters for the ordinal outcome
alpha[1]~dnorm(0, 1.0E-06) T(, alpha[2])
alpha[2]~dnorm(0, 1.0E-06) T(alpha[1], alpha[3])
alpha[3]~dnorm(0, 1.0E-06) T(alpha[2], )

## Prior for the shape parameter for the Weibull
lamda<-1  ## exponetial
lamda~dgamma(.1,.1) ## Weibull

## construct error variance
sigmaz<-1/tauz

## prior for error precision
tauz~dgamma(1.0, 1.0)

## construct variance-covariance matrix for random effects
sigma1[1:3,1:3]<-inverse(tau[,])
sigma11<-sigma1[1,1]
sigma22<-sigma1[2,2]
sigma33<-sigma1[3,3]
sigma12<-sigma1[1,2]
sigma13<-sigma1[1,3]
sigma23<-sigma1[2,3]
cor12<-sigma12/(sqrt(sigma11*sigma22))
cor13<-sigma13/(sqrt(sigma11*sigma33))
cor23<-sigma23/(sqrt(sigma22*sigma33))

## prior for  precision of random effects
tau[1:3,1:3]~dwish(R1[,], 4)

## Priors for regression coefficients (Betas)
beta1[1:8]~dmnorm(betamu1[],Sigma1[,])# continuous outcome
beta2[1:7]~dmnorm(betamu2[],Sigma2[,])# ordinal outcome
beta3[1:5]~dmnorm(betamu3[],Sigma3[,])# survival outcome
```

```
} # end model
",fill=TRUE)
sink()

###################################
# R program                       ###
###################################
## Bundle data
betamu1<-c(0,0,0,0,0,0,0,0)
betamu2<-c(0,0,0,0,0,0,0)
betamu3<-c(0,0,0,0,0)
Sigma1<-diag(0.01,nrow=8,ncol=8)
Sigma2<-diag(0.01,nrow=7,ncol=7)
Sigma3<-diag(0.01,nrow=5,ncol=5)
U0<-c(0,0,0)
R1<-diag(1,nrow=3,ncol=3)

# Place data in a list to be read by OpenBUGS
diab.data<-list(Nobs=1225,
                Npat=500,
                U0=U0,
                R1=R1,
                betamu1=betamu1,
                betamu2=betamu2,
                betamu3=betamu3,
                Sigma1=Sigma1,
                Sigma2=Sigma2,
                Sigma3=Sigma3,
                ubp=long.data$ubp,
                bmi_grp2=long.data$bmi_grp2,
                month=long.data$month,
                Sulphony=long.data$sulphony,
                Biguan=long.data$biguan,
                age=long.data$mage,
                male=long.data$male,
                patient=long.data$patient,
                surt=surv.data$surt,
                cens=surv.data$cens2,
                biguan=surv.data$biguan,
                sulphony=surv.data$sulphony,
                malet=surv.data$male,
                mage=surv.data$mage)

# Initial values for MCMC sampling
inits<-function(){list(
                beta1=c(b1[1],b1[2],b1[3],b1[4],b1[5],b1[6],b1[7],b1[8]),
                tauz=1,
                tau=diag(3),
                U=matrix(rnorm(500*3,0,0.5),500,3),
                alpha=c(b2[1],b2[2],b2[3]),
                beta2=c(-b2[4],-b2[5],-b2[6],-b2[7],-b2[8],-b2[9],-b2[10]),
                beta3=c(b3[1],b3[2],b3[3],b3[4],b3[5]),
                surt=ifelse(is.na(surv.data$surt),runif(500,surv.data$cens2,surv.data$cens2
                    +10),NA))
        }
# Parameters to estimate
params<-c("alpha","beta1","beta2","beta3","cor12","cor13","cor23","sigma11",
        "sigma22","sigma33","sigmaz")

# Start MCMC Sampling
long.surv.out<-BRugsFit(data=diab.data,
                        inits=inits,
                        parameters=params,
                        modelFile="diab-model_4-6-14.txt",
                        numChains=2,
                        nBurnin=5000,
                        nIter=30000,
```

```
                        nThin=1,
                        coda=FALSE,
                        DIC=TRUE,
                        digits=5,
                        BRugsVerbose=getOption("BRugsVerbose"))
print(long.surv.out)

#######################################################
######Model 4.2: Joint shared random effects    ######
#######################################################
sink("diab-model_4-6-14v2.txt")
cat("
###################################
# OpenBUGS program              ###
###################################
model { # begin model

                for (i in 1:Nobs){ # begin loop over observations
                Z[i]<-sqrt(ubp[i])
                Z[i]~dnorm(mu1[i], tauz)
                time[i]<-(month[i]-mean(month[]))/(sd(month[]))

                mu1[i]<-beta1[1]+beta1[2]*time[i]+beta1[3]*Biguan[i]+beta1[4]*Sulphony[i]
                +beta1[5]*(age[i]-mean(age[]))+beta1[6]*male[i]+beta1[7]*Biguan[i]*time[i]
                +beta1[8]*Sulphony[i]*time[i]+U1[patient[i]]

                mu2[i]<-beta2[1]*time[i]+beta2[2]*Biguan[i]+beta2[3]*Sulphony[i]
                +beta2[4]*(age[i]-mean(age[]))+beta2[5]*male[i]+beta2[6]*Biguan[i]*time[i]
                +beta2[7]*Sulphony[i]*time[i]+r1*(U1[patient[i]])

                ## cumulative logistic probabilities for the ordinal outcome
                        logit(Q[i, 1])<-alpha[1]-mu2[i]
                        p[i,1] <-Q[i,1]

                        for (k in 2:3) {
                        logit(Q[i,k])<-alpha[k]-mu2[i]
                        p[i,k]<-Q[i,k]-Q[i,k-1]
                        }
                        p[i, 4]<-1-Q[i, 3]
                        bmi_grp2[i]~dcat(p[i, 1:4])

                } # end loop over observations

                for (j in 1:Npat) { # begin loop over subjects

                # Survival Model
                log(mut[j])<-beta3[1]+beta3[2]*biguan[j]+beta3[3]*sulphony[j]
                        +beta3[4]*(mage[j]-mean(mage[]))+beta3[5]*malet[j]+r2*U1[j]

                surt[j]~dweib(lamda,mut[j])C(cens[j],)

                ## Random effects
                U1[j]~ dnorm(0.0,tau)

                } # end loop over subjects

                ## Prior for the threshold parameters for the ordinal outcome
                alpha[1]~dnorm(0, 1.0E-06) T(, alpha[2])
                alpha[2]~dnorm(0, 1.0E-06) T(alpha[1], alpha[3])
                alpha[3]~dnorm(0, 1.0E-06) T(alpha[2], )

                ## Prior for the shape parameter for the Weibull
                lamda<-1
                #lamda~dgamma(.1,.1)

                ## construct error variance
                sigmaz<-1/tauz
```

```
                ## prior for error precision
                tauz~dgamma(1,1)

                ## construct  variance for random effect
                sigma<-1/tau

                ## prior (half-Cauchy(s=25)) for standard deviation of random effect
                numertau~dnorm(0,1)
                denomtau~dnorm(0,0.0016)
                tau<-pow(numertau/denomtau,2)

                ## Priors for regression coefficients (Betas)
                beta1[1:8]~dmnorm(betamu1[],Sigma1[,])
                beta2[1:7]~dmnorm(betamu2[],Sigma2[,])
                beta3[1:5]~dmnorm(betamu3[],Sigma3[,])

                # prior for the association parameters (gamma)
                r1~dnorm(0.0,0.01)
                r2~dnorm(0.0,0.01)

} # end model
  ",fill=TRUE)
   sink()

###################################
# R program                   ###
###################################

# Bundle data
betamu1<-c(0,0,0,0,0,0,0,0)
betamu2<-c(0,0,0,0,0,0,0)
betamu3<-c(0,0,0,0,0)
Sigma1<-diag(0.01,nrow=8,ncol=8)
Sigma2<-diag(0.01,nrow=7,ncol=7)
Sigma3<-diag(0.01,nrow=5,ncol=5)

# Place data in a list to be read by OpenBUGS
diab.data<-list(Nobs=1225,
                Npat=500,
                betamu1=betamu1,
                betamu2=betamu2,
                betamu3=betamu3,
                Sigma1=Sigma1,
                Sigma2=Sigma2,
                Sigma3=Sigma3,
                ubp=long.data$ubp,
                bmi_grp2=long.data$bmi_grp2,
                month=long.data$month,
                Sulphony=long.data$sulphony,
                Biguan=long.data$biguan,
                age=long.data$mage,
                male=long.data$male,
                patient=long.data$patient,
                surt=surv.data$surt,
                cens=surv.data$cens2,
                biguan=surv.data$biguan,
                sulphony=surv.data$sulphony,
                malet=surv.data$male,
                mage=surv.data$mage)

# Initial values for MCMC sampling
inits<-function(){list(
                beta1=c(b1[1],b1[2],b1[3],b1[4],b1[5],b1[6],b1[7],b1[8]),
                tauz=1,
                numertau=rnorm(1),
                denomtau=rnorm(1),
                r1=rnorm(1,0,.1),
                r2=rnorm(1,0,.1),
```

```
                U1=rnorm(500,0,0.5),
                alpha = c(b2[1], b2[2], b2[3]),
                beta2=c(-b2[4],-b2[5],-b2[6],-b2[7],-b2[8],-b2[9],-b2[10]),
                beta3=c(b3[1],b3[2],b3[3],b3[4],b3[5]),
                surt=ifelse(is.na(surv.data$surt),runif(500,surv.data$cens2,surv.data$cens2
                    +10),NA))
        }

# Parameters to estimate
params<-c("alpha","beta1","beta2","beta3","r1","r2","sigma","sigmaz")

# Start MCMC Sampling
long.surv.out2<-BRugsFit(data=diab.data,
                         inits=inits,
                         parameters=params,
                         modelFile="diab-model_4-6-14v2.txt",
                         numChains=2,
                         nBurnin=5000,
                         nIter=30000,
                         nThin=1,
                         coda=FALSE,
                         DIC=TRUE,
                         digits=5,
                         BRugsVerbose=getOption("BRugsVerbose"))
print(long.surv.out2)

#######################################################
##### Model 4.3: Separate models                ######
#######################################################
sink("diab-model_4-6-14v3.txt")
cat("
####################################
# OpenBUGS program              ###
####################################
model { # begin model

                for (i in 1:Nobs){ # begin loop over observations
                Z[i]<-sqrt(ubp[i])
                Z[i]~dnorm(mu1[i],tauz)
                time[i]<-(month[i]-mean(month[]))/(sd(month[]))

                mu1[i]<-beta1[1]+beta1[2]*time[i]+beta1[3]*Biguan[i]+beta1[4]*Sulphony[i]
                +beta1[5]*(age[i]-mean(age[]))+beta1[6]*male[i]+beta1[7]*Biguan[i]*time[i]
                +beta1[8]*Sulphony[i]*time[i]+U1[patient[i]]

                mu2[i]<-beta2[1]*time[i]+beta2[2]*Biguan[i]+beta2[3]*Sulphony[i]
                +beta2[4]*(age[i]-mean(age[]))+beta2[5]*male[i]+beta2[6]*Biguan[i]*time[i]
                +beta2[7]*Sulphony[i]*time[i]+U2[patient[i]]

                ## Cumulative logistic probabilities for ordinal outcome
                        logit(Q[i, 1])<-alpha[1]-mu2[i]
                        p[i,1] <-Q[i,1]

                        for (k in 2:3) {
                        logit(Q[i,k])<-alpha[k]-mu2[i]
                        p[i,k]<-Q[i,k] - Q[i,k-1]
                        }
                        p[i, 4]<-1- Q[i, 3]
                        bmi_grp2[i]~dcat(p[i, 1:4])

                } # end loop over observations

                for (j in 1:Npat) { # begin loop over subjects
                # Survival Model
                log(mut[j])<- beta3[1]+beta3[2]*biguan[j]+beta3[3]*sulphony[j]
                        +beta3[4]*(mage[j]-mean(mage[]))+beta3[5]*malet[j]+U3[j]

                surt[j]~dweib(lamda,mut[j])C(cens[j],)
```

```
                ## Random effects
                U1[j]~dnorm(0.0,tau1)
                U2[j]~dnorm(0.0,tau2)
                U3[j]~dnorm(0.0,tau3)

                } # end loop over subjects

                ## Prior for the threshold parameters for the ordinal outcome
                alpha[1]~dnorm(0, 1.0E-06) T(, alpha[2])
                alpha[2]~dnorm(0, 1.0E-06) T(alpha[1], alpha[3])
                alpha[3]~dnorm(0, 1.0E-06) T(alpha[2], )

                ## Prior for the shape parameter for the Weibull
                lamda<-1        # exponetial
                #lamda~dgamma(.1,.1) # Weibull

                ## construct error variance
                sigmaz<-1/tauz

                ## prior for error precision
                tauz~dgamma(1.0, 1.0)

                ## construct  variances for random effects
                sigma1<-1/tau1
                sigma2<-1/tau2
                sigma3<-1/tau3

                ## prior (half-Cauchy(s=25)) for standard deviations of random effects
                numertau1~dnorm(0,1)
                denomtau1~dnorm(0,0.0016)
                tau1<-pow(numertau1/denomtau1,2)

                numertau2~dnorm(0,1)
                denomtau2~dnorm(0,0.0016)
                tau2<-pow(numertau2/denomtau2,2)

                numertau3~dnorm(0,1)
                denomtau3~dnorm(0,0.0016)
                tau3<-pow(numertau3/denomtau3,2)

                ## Priors for regression coefficients (Betas)
                beta1[1:8]~dmnorm(betamu1[],Sigma1[,])
                beta2[1:7]~dmnorm(betamu2[],Sigma2[,])
                beta3[1:5]~dmnorm(betamu3[],Sigma3[,])

} # end model
",fill=TRUE)
sink()

###################################
# R program                     ###
###################################
# Bundle data
betamu1<-c(0,0,0,0,0,0,0,0)
betamu2<-c(0,0,0,0,0,0,0)
betamu3<-c(0,0,0,0,0)
Sigma1<-diag(0.01,nrow=8,ncol=8)
Sigma2<-diag(0.01,nrow=7,ncol=7)
Sigma3<-diag(0.01,nrow=5,ncol=5)

# Place data in a list to be read by OpenBUGS
diab.data<-list(Nobs=1225,
                Npat=500,
                betamu1=betamu1,
                betamu2=betamu2,
                betamu3=betamu3,
                Sigma1=Sigma1,
```

```
                Sigma2=Sigma2,
                Sigma3=Sigma3,
                ubp=long.data$ubp,
                bmi_grp2=long.data$bmi_grp2,
                month=long.data$month,
                Sulphony=long.data$sulphony,
                Biguan=long.data$biguan,
                age=long.data$mage,
                male=long.data$male,
                patient=long.data$patient,
                surt=surv.data$surt,
                cens=surv.data$cens2,
                biguan=surv.data$biguan,
                sulphony=surv.data$sulphony,
                malet=surv.data$male,
                mage=surv.data$mage)

# Initial values for MCMC sampling
inits<-function(){list(
                beta1=c(b1[1],b1[2],b1[3],b1[4],b1[5],b1[6],b1[7],b1[8]),
                numertau1=rnorm(1),
                denomtau1=rnorm(1),
                numertau2=rnorm(1),
                denomtau2=rnorm(1),
                numertau3=rnorm(1),
                denomtau3=rnorm(1),
                alpha=c(b2[1],b2[2],b2[3]),
                beta2=c(-b2[4],-b2[5],-b2[6],-b2[7],-b2[8],-b2[9],-b2[10]),
                beta3=c(b3[1],b3[2],b3[3],b3[4],b3[5]),
                U1=rnorm(500,0,0.5),
                U2=rnorm(500,0,0.5),
                U3=rnorm(500,0,0.5),
                surt=ifelse(is.na(surv.data$surt),runif(500,surv.data$cens2,surv.data$cens2
                    +10),NA))

        }

# Parameters to estimate
params<-c("alpha","beta1","beta2","beta3","sigma1","sigma2","sigma3","sigmaz")

# Start MCMC Sampling
long.surv.out3<-BRugsFit(data=diab.data,
                        inits=inits,
                        parameters=params,
                        modelFile="diab-model_4-6-14v3.txt",
                        numChains=2,
                        nBurnin=5000,
                        nIter=30000,
                        nThin=1,
                        coda=FALSE,
                        DIC=TRUE,
                        digits=5,
                        BRugsVerbose=getOption("BRugsVerbose"))
print(long.surv.out3)
```

BIBLIOGRAPHY

[1] M. Aitkin and D. Clayton. The fitting of exponential, weibull and extreme value distributions to complex censored survival data using GLIM. *Applied Statistics*, 29:156–63, 1980.

[2] T. Baghfalaki, M. Ganjali, and D. Berridge. Joint modeling of multivariate longitudinal mixed measurements and time to event data using a Bayesian approach. *Journal of Applied Statistics*, 41:1934–1955, 2014.

[3] N. G. Best, D. J. Spiegelhalter, A. Thomas, and C. E. G. Brayne. Bayesian analysis of realistically complex models. *Statistical Society, Series A*, 159:323342, 1996.

[4] N. E. Breslow and D. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25, 1993.

[5] Stephen P. Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1997.

[6] E. R. Brown and J. G. Ibrahim. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59:221–228, 2003a.

[7] E. R. Brown and J. G. Ibrahim. Bayesian approaches to joint cure rate and logitudinal models with application to cancer vaccine trials. *Biometrics*, 59:686–693, 2003b.

[8] O. D. Buhule, J. J. Odwee, and L.K. Atuhaire. Determinants of recovery time of diabetic patients in Uganda. *Makerere University Research Journal*, 2:1–10, 2007.

[9] Bradley P. Carlin and Thomas A. Louis. *Bayesian Methods for Data Analysis*. Chapman and Hall/CRC, Baco Raton, 2009.

[10] J. Y. Carter and O. Lema. In *Practical Laboratory Manual for Health Centres in Eastern Africa*, page 218, Nairobi, 2003. African Medical and Research Foundation (AMREF).

[11] P. J. Catalano. Bivariate modelling of clustered continuous and ordered categorical outcomes. *Statistics in Medicine*, 16:883–900, 1997.

[12] P. J. Catalano and L. M. Ryan. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, 419:651–658, 1992.

[13] Mary Kathryn Cowles. Accelerating monte carlo markov chain convergence for cumultive-link generalized linear models. *Statistics and Computing*, 6:101–111, 1996.

[14] D. R. Cox. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–200, 1972.

[15] D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

[16] V. DeGruttola and X. M. Tu. Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics*, 50:1003–1014, 1994.

[17] Peter J. Diggle, Patrick J. Heagerty, Kung-Yee Liang, and Scott L. Zeger. *Analysis of Longitudinal Data*. Oxford University Press, UK, 2nd edition, 2002.

[18] C. J. Faucett and D. C. Thomas. Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*, 15:1663–1685, 1996.

[19] G. M. Fitzmaurice and N. M. Laird. Regression models for a bivariate discrete and continuous outcomes with clustering. *Journal of the American Statistical Association*, 90:845–852, 1995.

[20] D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7:57–68, 1997.

[21] S. Geisser and W. Eddy. A predictive approach to model selection. *Journal of the American Statistical Association*, 74:153–160, 1979.

[22] A. E. Gelfand and D. Dey. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B*, 56:501–514, 1994.

[23] A. E. Gelfand and A. F. M. Smith. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[24] A. Gelman. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.

[25] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, London, 1995.

[26] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457472, 1992.

[27] Andrew Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:515–533, 2006.

[28] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, 2009.

[29] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–41, 1984.

[30] J. Geweke. In AP Dawid JM Bernado, JO Berger and AFM Smith, editors, *Bayesian Statistics 4*, Clarendon Press, Oxford, UK, 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments.

[31] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1996.

[32] R. V. Gueorguieva and A. Agresti. A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association*, 96:1102–1112, 2001.

[33] R. V. Gueorguieva and G. Sanacora. Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. *Statistics in Medicine*, 25:1307–1322, 2006.

[34] X. Guo and B. P. Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *Am. Stat*, 58:16–24, 2004.

[35] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[36] L. A. Hatfield, M. E. Boye, and B. P. Carlin. Joint modeling of multiple longitudinal patient-reported outcomes and survival. *Journal of Biopharmaceutical Statistics*, 21:971–991, 2011.

[37] P. J. Heagerty, T. Lumley, and M. Pepe. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics*, 56:337–344, 2000.

[38] P. J. Heagerty and Y. Zheng. Survival model predictive accuracy and roc curves. *Biometrics*, 61:92–105, 2005.

[39] D. Hedeker and R. D. Gibbons. A random effects ordinal regerssion model for multilevel analysis. *Biometrics*, 50:933–944, 1994.

[40] D. Hedeker and R. J. Mermelstein. A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research*, 33:427–455, 1998.

[41] P. Heidelberger and P. D. Welch. Simulation run length control in the presence of an initial transient. *Operations Research*, 31:11091144, 1983.

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

[43] R. Henderson, P. Diggle, and A. Dobson. Identification and efficacy of longitudinal markers for survival. *Biostatistics*, 3:33–55, 2002.

[44] J. W. Hogan and N. M. Laird. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, 16:239–257, 1997a.

[45] J. W. Hogan and N. M. Laird. Mixture models for the joint distributions of repeated measures and event times. *Statistics in Medicine*, 16:239–257, 1997a.

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

[47] P. Hougaard. Frailty models for survival data. *Lifetime Data Analysis*, 1:255–273, 1995.

[48] C. Hu and M.E. Sale. A joint model for nonlinear longitudinal data with informative dropout. *Journal of Pharmacokinetics and Pharmacodynamics*, 30:83103, 2003.

[49] Y. Huang, X. J. Hu, and G. A. Dagne. Jointly modeling time-to-evnt and longitudinal data: a Bayesian approach. *Stat Methods Appl*, 23:95–121, 2014.

[50] A. V. Huzurbazar. *Flowgraph Models for Multistate Time-to-Event Data*. John Wiley & Sons, Hoboken, 2005.

[51] J. G. Ibrahim, M. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer, New York, 2001.

[52] J. G. Ibrahim, H. Chu, and L. M. Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28:2796–2801, 2010.

[53] SAS Institute. SAS/STAT user's guide. In *Version 9.2 ed*, Cary, NC, 2007. SAS Institute Inc.

[54] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley, New York, 2nd edition, 2002.

[55] N. Keiding, P. K. Anderson, and J. P. Klein. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16:215–224, 1997.

[56] John P. Klein and Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science + Business Media, LLC, New York, 2003.

[57] N. M. Laird and J. H. Ware. Random effects models for longitudinal data. *Biometrics*, 38:963–74, 1982.

[58] M. P. Lavalley and V. DeGruttola. Model for empirical bayes estimators of longitudinal CD4 counts. *Statistics in Medicine*, 15:2289–2305, 1996.

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

[60] K. Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.

[61] H. Lin, B. W. Turnbull, C. E. McCulloch, and E. H. Slate. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, 97:53–65, 2002.

[62] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. John Wiley and Sons, Inc, New Jersey, 2002.

[63] J. S. Liu, W. H. Wong, and A. Kong. Covariance structure of the gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81:27–40, 1994.

[64] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The bugs project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, 28:3049–3082, 2009.

[65] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS: a Bayesian modelling framework: concepts, structure, an extensibility. *Statistical Computing*, 10:325–337, 2000.

[66] S. Luo. A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in Medicine*, 33:580–594, 2014.

[67] P. McCullagh. Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B*, 42:109–142, 1980.

[68] P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, 1989.

[69] Charles McCulloch. Joint modelling of mixed outcome types using latent variables. *Statistical Methods in Medical Research*, 17:53–73, 2008.

[70] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–92, 1953.

[71] P. Muller. A generic approach to posterior integration and gibbs sampling. *Journal of the American Statistical Association*, 1994.

[72] J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society,Series A*, 135:370–84, 1972.

[73] Y. Pawitan and S. Self. Modeling disease marker process in AIDS. *Journal of the American Statistical Association*, 83:719–726, 1993.

[74] M. Plummer, N. G. Best, K. Cowles, and K. Vines. Coda: convergence diagnosis and output analysis for mcmc. *R News*, 6:711, 2006.

[75] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006.

[76] R. Prentice. Covariate measurement errors and parameter estimates in a failure time regression model. *Biometrika*, 69:331–342, 1982.

[77] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[78] D. Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819–829, 2011.

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

[80] D. Rizopoulos, G. Verbeke, and G. Molenberghs. Shared parameter models under random effects misspecification. *Biometrika*, 95:63–74, 2008.

[81] D. B. Rudin. Inference and missing data (with discussion). *Biometrika*, 63:581–592, 1976.

[82] M. Schemper and R. Henderson. Predictive accuracy and explained variation in cox regression. *Biometrics*, 56:249–255, 2000.

[83] S. Self and Y. Pawitan. Modeling a marker of disease progression and onset of disease. In N. P. Jewell, K. Dietz, and V. T. Farewell, editors, *AIDS Epidemiology: Methodoligical Issues*, Birkhauser, Boston, 1992.

[84] Z. Shun and P. McCullagh. Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society, Series B*, 57:749–760, 1995.

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

[86] X. Song, M. Davidian, and A. A. Tsiatis. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58:742–753, 2002b.

[87] D. J. Spiegelhalter, N. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit(with discussion). *Journal of the Royal Statistical Society, Series B*, 64:583–639, 2002.

[88] D. J. Spiegelhalter, A. Thomas, N. Best, and W. R. Gilks. Bugs examples 0.30. In *BUGS Examples 0.30*, Cambridge, 1993. MRC Biostatistics Unit.

[89] M. J. Sweeting and S. G. Thompson. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 5:750–763, 2011.

[90] L. Tierney. Markov chains for exploring posterior distributions(with discussion). *Annals of Statistics*, 22:1701–62, 1994.

[91] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81:82–86, 1986.

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

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

[94] A. A. Tsiatis, V. DeGruttola, and M. S. Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90:27–37, 1995.

[95] J. W. Vaupel, K. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439–54, 1979.

[96] P. Wang, W. Shen, and M. E. Boye. Joint modeling of longitudinal outcomes and survival using latent growth modeling approach in a mesothelioma trial. *Health Serv Outcomes Res Method*, 12:182–199, 2012.

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

[98] B. Williams. In Martin Dunitz, editor, *Epidemiology and pathogenesis of hypertension in diabetes*, page 323, London, 2003. Williams B, ed. Hypertension in diabetes.

[99] R. Wolfinger. Fitting nonlinear mixed models with the new NLMIXED procedure. In *24th SAS Users Group International Conference Proceedings*, Miami Beach, FL, Paper 287,1999.

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

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

[102] J. Xu and S. L. Zeger. The evaluation of multiple surrogate endpoints. *Biometrics*, 57:81–87, 2001b.

[103] S. L. Zeger and K. Y. Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42:121–130, 1986.