University of Fort Hare

Together in Excellence

Joint Modelling Inference for Longitudinal and Time To Event Data with Application to Biomarkers in Medical and Clinical Studies

By

Azeez Adeboye Nurudeen

(201400816)

A Thesis Submitted in Fulfilment of the Requirements for the Degree of

Doctor of Philosophy in Biostatistics

Department of Statistics

Faculty of Science and Agriculture

University of Fort Hare, Alice, South Africa

Supervisor: Prof YongSong Qin

Co-Supervisors: Dr James Ndege

Dr Ruffin Mutambayi

Declaration

I declare that this thesis titled **'A Joint Modelling Inference for Longitudinal and Time To Event Data with Application to Biomarkers in Medical and Clinical Studies'** is my own research work and all the sources cited or quoted have been referenced and acknowledged through of complete references. I further confirm that this research work has not been submitted before for any other degree at any other institution.

Student's Name: Azeez Adeboye 08/12/2020

Supervisor: Professor Yongsong Qin 08/12/2020

Co-Supervisor: Dr. James Ndege 08/12/2020

Co-Supervisor: Dr. Ruffin Mutambayi 08/12/2020

Acknowledgement

My first gratitude goes to Almighty God for this life, good health, wisdom, knowledge, and understanding bestowed on me to accomplish this Doctoral degree.

My sincere gratitude to my supervisors Professor Y. Qin, Dr James Ndege and Dr. Ruffin Mutambayi for all their available time and their constructive criticisms in guiding me through the process of my entire Ph.D. research. Your valuable comments have immensely contributed to enrich this thesis. Thank you for all the patience and support given to me to overcome all the mistakes that have arisen therein. I have learnt a lot in the process and all of our meetings were extremely enlightened. It was a great pleasure to do this research work under your supervision.

Special thanks to Govan Mbeki Research and Development Centre for providing me with Supervisor-Linked Bursary to support my research in terms of data collection, software purchase and attending conferences. Also to the Research Square Centre, USA, for the exposure in manuscript writing and reviewing. A special thanks to the Nigerian Bioinformatics organization for providing me travel grant to attend their first ever conference in Nigeria.

I want to, especially thank my parents and siblings for having faith in me and encouraging me in every decision I make in life, always believing in the choices that led me to this place. Thank you for all the love and never-ending support even when it was not so easy.

To all my well-wishers, thank you for being part of my life. All glory and adoration to God, the Lord of the world.

Dedication

I dedicate this thesis to the Glory of Almighty

Abstract

In the past couple of decades, longitudinal and survival data analysis have emerged as important and popular concepts of biostatistics and statistics for disease modelling. In recent years, these two statistics concepts have been combined to develop a joint model for longitudinal and survival data analysis. Joint model is a simultaneous modelling application of longitudinal and survival data while taking into account a possible association between them. In this thesis, three sub-topics (Conditional score approach, estimating equation approach, and modified Cholesky decomposition approach) are utilised to model the association if the independence assumption is violated.

Using the conditional score approach, the study investigated the association between longitudinal covariates and time-to-event process to examine the within-subject measurement error that could influence estimation when the assumption of normality and mutual independence is violated. Given the assumption violation, I proposed an estimating equation approach based on the conditional score to relax parametric distributional assumptions for repeated measures random effects. I jointly modelled the time-dependent biomarkers and event times using Cox model with intermittent time-dependent covariates measure, in which the longitudinal model was used to characterize the biomarker underlying (unobservable) trajectory and incorporated as a latent time-dependent covariate in the survival model to predict failure times. Estimates of the parameters were obtained by a restricted maximum likelihood estimate (REML). A modified Cholesky decomposition method was used to capture the within-subject covariance for a positive definite and symmetric matrix,

with the assumption that the observed data from different subjects are independent. I illustrated the proposed method by a real data set from a lung study and simulation. An extension to the joint model of longitudinal-survival data was also proposed, in which the longitudinal data has a cumulative and weighted effect on the hazard event function. Using a Bayesian parametric method, I proposed a skewed weighted probability density function to estimate the parameters. The weighted cumulative effect used enabled different longitudinal profile to be incorporated over time in calculating the hazard ratio between the subjects. The proposed functions provide greater flexibility for modelling the association structure of different longitudinal and survival sub-model. The focus was on the association between the biomarker (serum creatinine, sCr) and the development of an end-stage renal disease (ESRD). Since the effect of sCr biomarker is anticipated to be a cumulative effect, with the development of sCr biomarker over time leading to progressively higher damage of the kidney. The approach was applied as simulation for validation of the proposed method.

Keywords: Biomarker, Cumulative effect, Density function, Laplace Approximation, Weight function, Conditional score, Estimating equation, Bayesian method, Cholesky decomposition

Table of Contents

List of Tables

Table of Figure

CHAPTER ONE

1.0 Introduction to Joint Model of Longitudinal and Survival Data

1.1 Background

In the past couple of decades, longitudinal and survival data analysis have emerged as some of the fastest growing concepts in biostatistics and statistics to for disease modelling. Longitudinal-data/ panel-data analysis generally refers to statistical methods for repeated measurements of data analysis from a longitudinal study. In various medical researches, repeated measurements (biomarkers) data include multiple observations of an outcome variable such as body mass index (BMI), that are measured overtime on the same study unit during the course of follow up or the outcome of data on time to the happening of a particular event. For instance, time-to-event outcomes in Randomized controlled trials (Argyropoulos and Unruh 2015), in which patients follow-up are recorded over a period and biomarkers are repeatedly collected at multiple time intervals (repeated measurement). Such repeated biomarkers may be CD4+ count or viral load biomarkers for HIV/AIDS, Geriatric cognitive performance study, Systolic blood pressure and a coronary event in Cardiovascular study, Prostate-specific antigen biomarker and recurrence in Cancer study and so on. The fundamental concept of longitudinal data analysis is how to resolve correlations within-subjects and handling missing observations.

On the other hand, survival analysis deals with survival data or time-to-event data for which the outcome variable is time to the occurrence of an event. Data obtained in this manner is referred to as survival data of the time to the occurrence of a particular event (Austin, Lee, and Fine 2016). An event could be, for instance, death, the end of the period spent in remission from disease, relief from diseases,

symptoms, equipment failure, a disease recurrence, or discharge from a hospital. The survival data is the time to an event, i.e. the time at which a particular event of interest occurred. Such events may be *Adverse* like the relapse or recurrence of a disease like a tumour or malaria; *Positive* event like discharge from the hospital, *Neutral* event such as cessation of breastfeeding. Time-to-event data are usually incomplete, and thus cannot be handled by standard statistical tools for complete data. Many studies focus on the effect of patients' information on different survival predictions and modelling repeated data with event time outcomes in order to construct dynamic prediction models that modify the event over-time with accumulating evidence (Andrinopoulou et al. 2015). A typical example is right censoring, which occurs when the survival time of interest is only known to be greater than some observed censoring time due to the end of follow up or the occurrence of early withdrawal or competing events.

In recent years, these two statistics concepts have been blended to develop joint models for longitudinal and survival data analysis. Originally, the joint model was introduced to independently address the difference in longitudinal data analysis and survival analysis problems. Joint modelling in longitudinal data analysis was basically meant to adjust for non-ignorable missing data due to informative or outcome related to dropouts, which traditional methods such as linear mixed-effects models cannot handle accurately. While survival analysis was first proposed for Cox's proportional hazards model to deal with time-dependent covariates that are measured intermittently, which may be/ not subject to measurement error, Joint models, have also become popular in medical research. Here, both the longitudinal variable (such as a disease biomarker) and the time-to-event variable (such as the disease-free

survival time), are important outcome variables to evaluate the efficacy of interventions or treatments.

Time-to-event outcomes can be used to censor a longitudinal data and modelling both the repeated and time event outcomes separately, for instance, using time-dependent random effect models and survival model (Barrett and Su 2016), modelling Linear mixed models and Cox models (Hickey et al. 2016). This may sometimes be ineffective and can as well lead to biased size estimates if the two event outcomes are correlated (Ibrahim, Chu, and Chen 2010). The Cox PH model is the commonly used model for survival data (Cox 1972) and serves as part of the techniques that will be used in the study. The Cox model has extended to include the random effects (Vaupel, Manton, and Stallard 1979), multivariate survival times (Hougaard 2000), covariates measurement error (Wulfsohn and Tsiatis 2010), time-dependent covariates (Sweeting and Thompson 2011). The model also incorporates extensions such as longitudinal modelling with outcome-dependent drop-out (Henderson, Diggle, and Dobson 2000), and modelling longitudinal-time event with the inclusion of latent classes (Berlin, Parra, and Williams 2014).

1.2 Motivating Study Data

Impaired Renal function is recognised as one of the risk factors that leads to development of tuberculosis (TB), especially the extra-pulmonary TB (EPTB) type. This type of TB is more pronounced in patients with chronic renal disease (CRD) compared to patients with normal renal function. The progressive decrease of renal failure over months or years occur in stages. Each stage determines the progression of TB patients with impaired renal function from the level of abnormally low to progressively defective glomerular filtration rate. There are no randomised clinical trials that provide guideline or evidence of TB treatment with impaired renal failure.

The treatment guidelines currently available are based on case studies, such as pharmacological guideline of the drugs used and experts' recommendations from the international agencies involved in TB control. TB patients with impaired renal failure experience different side effects such as high blood pressure, anaemia, bones weakness and damage of nerves. However, when the TB progresses, this may eventually lead to impaired renal failure, which necessitates looking for factors that may influence the treatment of an impaired renal failure patient with TB. Such factors are the drug pharmacokinetics to determine the proportion of drug excreted by kidneys and dialysis clearance (both haemodialysis and peritoneal dialysis), which affects the serum levels of drugs and consequently, the toxicity. Furthermore, the toxic severity predicts high blood levels of drugs, the accessibility of alternative effective ways to cure the TB patient with co-infection and possible drug interactions, which may affect therapy.

Although, TB is a rare causal agent of progressive renal failure, it is an important one. It is definitely preventable and easy to treat. Study evidence showing the extent to which TB causes end-stage renal failure worldwide are few. For this reason, there is little information on the contribution that tuberculosis makes to the burden of renal disease. The motivating study data used throughout this thesis pertains to the hospital records of TB patients with impaired renal failure that have undergone treatments from 1/03/2008 to 1/03/2018 from Grey Hospital, King Williams Town, Eastern Cape, South Africa. The medical and clinical interest from the data lies in the long-term performance of treatment of renal failure in TB patients. In this study, the glomerular filtration rate (GFR) biomarker, in which serum creatinine level is used as an indicator of filtration through the glomerular filter considered for the survival rate.

Preliminary analyses were conducted to show the evidence within subject-specific trajectories for longitudinal are important prediction ways for serum creatinine in renal failure. The longitudinal profiles average is showed in Figure 1a (Appendix) and the within subject-specific profiles for some randomly selected TB patients with impaired renal function from our motivating data is illustrated in Figure 1b (Appendix). It was observed that patients with GFR failure status show a relative difference time of development than patients without failure. Hence, these longitudinal profiles have enough information to help the clinician to monitor the risk status progression of each of the patients.

1.3 Research Problem Statement

Tuberculosis with impaired renal function is not different from other forms of disease caused by the Mycobacterium TB complex. The commonest causative agent is the human tubercle bacillus (*M. tuberculosis*) but the bovine tubercle bacillus (*M. bovis*) can also be responsible. However, it is unknown how the renal function dosage adjustments advocated by the guidelines affect the efficiency outcomes for TB patients with chronic renal failure. It is also unclear how this dosage treatment affects the frequency of drug-related side effects in patients with chronic renal failure resulting from the differences between previous treatment and current follow-up. Therefore, it is very important to understand clearly and comprehensibly, the impact of renal function in investigating and assessing the efficiency of using repeated covariates measures with correlated error to predict clinical outcome and the within the subject-specific cumulative effects association structures between a longitudinal biomarker and time-to-event. To identify correlation in the survival endpoint and processes that produced longitudinal measured outcomes, which could improve the efficacy of clinical treatment as well as providing a better insight into many areas of

the clinical effect of interventions contained within the treatment. To tackle these questions, a Bayesian approach was specifically applied to develop a flexible joint longitudinal-survival model to determine where and how the approaches could give value in interpreting the findings.

1.4 Aim of the Research

The study aims to offer an explicit assessment of the gain in efficiency from using repeated measures for covariates with correlated error for a joint model to predict clinical outcome and improve the prediction prognosis.

1.4.1 Objectives of the Research

1. This study seeks to discover the pragmatic application of Bayesian inferences for flexible joint modelling of longitudinal-survival time event outcomes with latent class variables.
2. This study will test the cumulative effects association structures between a longitudinal biomarker and time-to-event.
3. This study will be used to express the association between changes in history in longitudinal-survival outcomes
4. This study will discover correlation summary measures of the trajectory of longitudinal-survival outcomes

1.5 Research Questions

The models will establish a good paradigm for the analysis of repeated and time event with follow-up data that is mainly applicable in two settings:

i. When the focus is on a survival outcome, and wish to account for the effect of endogenous time-dependent covariates measured with error.

ii. When the focus is on the longitudinal outcome and wishes to correct for cumulative effects association.

1.6 Outline of the Thesis structure

Chapter 1 introduced the background of joint models for longitudinal and time-to-event outcomes, the problem of research for the thesis, the aims and objectives of the research. Chapter 2 discussed the overview of the literature reviews on methodological development and clinical application of joint models of longitudinal and time-to-event outcomes over the past two decades. It also described the fundamentals and standard statistical techniques for longitudinal analysis and time-to-event analysis. Chapter 3 discussed the joint model of longitudinal and time-to-even process for repeated covariates measures with a correlated error. Chapter 4 discussed the analyses and interpretation of the results from the joint model analysis. Chapter 5 discussed the conclusion and recommendations in line with the joint model.

CHAPTER TWO

2.0 An Overview on Literature Reviews for Joint Models

2.1 Introduction

This chapter describes the literature available on the fundamentals and standard statistical techniques for longitudinal analysis and time-to-event analysis. For longitudinal data analysis, the missing data mechanisms, missing data imputation must rely on certain missing data assumptions for both ignorable and non-ignorable missing data and the most common models used are discussed. For time-to-event data analysis, the basic quantities, missing data mechanisms, the methods for time-to-event data and diagnostics are described. For joint analysis, different approaches to parameter estimation and the different modelling strategies are discussed. A gap in the research of joint modelling with regard to the use of parametric time-to-event models in joint modelling is identified. This thesis will try to fill in that gap.

2.2 Longitudinal Data Analysis

Longitudinal data are measurements of the samples (or subjects) often collected or measured repeatedly over time in studies like clinical trials or follow-up studies, in which individuals are measured repeatedly. It studies the change in an outcome over time. In longitudinal data analysis, there are two approaches to use: (i.) Marginal models –this refers to the outcomes of the population mean over time and the effects of the covariate on the population mean; (ii.) Mixed-effects models – represents the effects of the covariates on subject-specific mean response trajectories. Once the data structure is described, the key is to distinguish between the parameters of the

model, classified into fixed effects and random effects. In this way, the response or dependent variable is assumed a function of fixed effects, non-observable cluster-specific random effects, and an error term. However, two main classes of models are reviewed - the linear mixed-effects models (LMM) for normally distributed longitudinal data and Generalized linear mixed-effects models, which is an extension of LMM to exponential families.

i. Linear Mixed-Effects Models (LMM)

Linear mixed-effect models also known as Classical mixed models have been a standard approach to analyse normally distributed longitudinal outcome variables. However, the general form of Linear Mixed effect Models assumes that

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i, \qquad (2.4)$$

where Y_i is the outcome vector for subject i with dimension n_i, X_i and Z_i are $n_i \times p$ and $n_i \times q$ matrices of known covariates, $i = 1, \ldots, n$. The columns of Z_i are a subset of the columns in X_i such that $q \leq p$. The error term (ε_i) is a $n_i \times 1$ vector of measurement errors, b_i is a $q \times 1$ vector called subject-specific random effects, and β is a $p \times 1$ vector of regression coefficient called the fixed effects. The random effects assumed that $b_i \sim N(0, D)$ is independent of ε_i and describe the between-subject variability i.e. variance-covariance matrix D and $N = \sum_{i=1}^{n} n_i$ while ε_i explains the within-subject variability and assumed to follow $N(0, \sum_e)$. If b_i is a random intercept and the error terms are assumed to be mutually independent, we can then write $D = \sigma_b^2$ and $\sum_e = \sigma_e^2 I_{n_i}$, where I_{n_i} is the identity matrix of dimension n_i. It follows that $\theta = (\sigma_b^2, \sigma_e^2)$ or reparameterized as $\theta = (\sigma_b^2, \rho)$, where the intra-class correlation coefficient is expressed as $\sigma_b^2 / (\sigma_b^2 + \sigma_e^2)$. Equation (2.4) model implies that,

marginally, Y_i follows a normal distribution with mean $X_i\beta$ and variance-covariance matrix $V_i = Z_i D Z_i^T + \sum_i$. The random effects b_i can be interpreted as residuals to express unit-specific trends deviating from the population mean. The simplest form of $\sum_i$ is $\sigma^2 I_{n_i}$.

ii. Generalized Linear Mixed Effects Models (GLMMs)

Generalized linear mixed-effects models (GLMMs) are an extension of generalized linear models by incorporating random regression coefficients to characterize within-subject correlations in longitudinal or clustered data. GLMMs also extend linear mixed- effects models to a rich class of distributions, which can be generally expressed in the form of exponential families conditional on random coefficients. A distinctive feature of GLMMs is that the fixed effects may no longer have a marginal interpretation. In addition, the computational burden increases substantially in GLMMs.

In GLMM, it is assumed that given a vector of random effects b_i, Y_{ij}, $j = 1,\ldots,n_i$, are independent and follows a distribution in the exponential family with $var[Y_{ij}|b_i] = V\left(E\left[Y_{ij} \mid b_i\right]\right)\ell$, where ℓ is the dispersion parameter, and variance function $V(.)$ is determined by the specific distribution of $Y_{ij}|b_i$. The mean of conditional response u_{ij} is linked to the fixed and random effects through a linear predictor ξ_{ij} for some known link function *g(.)*

$$\mu_{ij} = E\left[Y_{ij} \mid b_i\right],$$
$$g\left(\mu_{ij}\right) = \xi_{ij},$$
$$\xi_{ij} = X_{ij}^T \beta + Z_{ij}^T b_i$$

The distribution of random effects b_i is commonly assumed as $N\left(0,\sum_b\right)$. For instance, model (2.4) is a special case of the generalized linear mixed-effects model with identity link function $g(\mu)=\mu$ and normal distribution $N\left(0,\sigma^2\right)$ for $Y_{ij}|b_i$. In this case $var[Y_{ij}|b_i]=\sigma^2, \ell=\sigma^2, V(.)=1$. Parameters in the generalized linear mixed-effects model can be estimated either by directly maximizing the likelihood with numerical integration, by EM algorithm, or by maximizing an approximation of the log-likelihood function.

For Model assumptions, we assume each subject is associated with a vector of random coefficients (effects) b_i. Let Y_i denote the observation on subject i at occasion $j, j=1,\ldots,n_i, i=1,\ldots,n$. Conditional on the random effects b_i, Y_{ij} are assumed to follow a distribution in the exponential family with probability density function

$$f\left(y_{ij}\,|\,b_i,\beta,\phi\right)=\exp\left\{\left[y_{ij}\theta_{ij}-\alpha\left(\theta_{ij}\right)+c\left(y_{ij}\right)\right]\phi\right\},$$

where θ_{ij} and ϕ are parameters and $\alpha(.)$ and $c(.)$ are known functions. Based on the theory of exponential families, the conditional mean and variance of Y_{ij} are $u_{ij}=E\left(Y_{ij}\,|\,b_i,\beta,\phi\right)=\alpha'\left(\theta_{ij}\right)$ and $v_{ij}=var\left(Y_{ij}\,|\,b_i,\beta,\phi\right)=\alpha''\left(\theta_{ij}\right)/\phi$, respectively. Here we use $\alpha'(.)$ and $\alpha''(.)$ to denote the first and second derivatives of $\alpha(.)$. Normal, binomial, Poisson, exponential, gamma, and inverse Gaussian distributions are special cases of exponential families.

Difference from linear mixed effects models, GLMMs model the mean of Y through a one-to-one continuous differentiable transformation and assume that the transformed mean is characterized by a linear model, i.e. $\xi_{ij}=X_{ij}^T\beta+Z_{ij}^Tb_i$, where ξ_{ij} is the linear

predictor. The random effects b_i are usually assumed to follow a multivariate normal distribution with zero mean and variance-covariance matrix D and are assumed to be independent of the covariates. Therefore, linear mixed effects models are a special case of GLMMs, in which β has a marginal interpretation because $E\left(Y_{ij} \mid \beta\right) = E\left\{E\left(Y_{ij} \mid b_i, \beta\right)\right\} = E\left(x_{ij}^T\beta + z_{ij}^T b_i\right) = x_{ij}^T\beta$ given that $E\left(b_i\right) = 0$. This indicates that the marginal mean of Y is a linear model with respect to β. However, this relationship is not generally true when the link function $g(.)$ is nonlinear. For other distributions in GLMMs, β is generally interpreted as the impact of covariates on the mean response of a specific subject conditional on the random effects.

For GLMMs inference, the likelihood function of $\beta, \phi,$ and D is evaluated by integrating the conditional probability distribution over b_i. Specifically, we have

$$L\left(\beta, \phi, D\right) = \prod_{i=1}^{n} \int f\left(Y_i \mid b_i\right) f\left(b_i\right) db_i$$

$$= \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f\left(Y_i \mid b_i\right) f\left(b_i\right) db_i \qquad (2.5)$$

The maximum likelihood estimates can be obtained by maximizing the equation (2.5). Since the integration is intractable, the Gaussian quadrature approximates through integral as a weighted sum, can be used to approximate the integral.

2.3 Time-to-Event Data Analysis

Survival analysis, or time-to-event data analysis, refers to statistical methods for time-to-event data. An event time, or survival time, is defined as the time from an initial event such as diagnosis of a disease to the occurrence of an event of interest such as death. Time-to-event data arise commonly in clinical trials and other follow-up studies. For example, time to death or treatment failures is a primary clinical

outcome variable to evaluate the effectiveness of a treatment for patients with a terminal disease. For convenience, I used the words event, failure, and death interchangeably.

Survival Function

Assumes that T is a continuous random variable with probability density function (p.d.f.) $f(t)=\lim_{\delta t\to 0}\frac{P(t\le T<t+\delta t)}{\delta t}$ and cumulative distribution function (c.d.f.) $F(t)=\Pr\{T<t\}$, giving the probability that the event has occurred by duration t and δt is the change in time. The survival function is

$$S(t)=\Pr\{T<t\}=1-F(t)=\int_t^{\infty} f(x)dx$$

which gives the probability of being alive just before duration t, or more generally, the probability that the event of interest has not occurred by duration t.

Hazard Function

Hazard function describes the instantaneous rate of occurrence of the event (failure rate) or risk of an even within $[t,t+dt]$ provided that the subject survived to time t (Rubin 2004). This is also referred to as the risk function and expressed as

$$h(t)=\lim_{\delta t\to 0}\frac{\Pr\{t\le T<t+dt\,|\,T\ge t\}}{dt},t>0$$

The numerator of this expression is the conditional probability that the event will occur in the interval $[t,t+dt]$ given that it has not occurred before, and the denominator, dt is the width of the interval. Taking the limit as the width of the interval goes down to zero, we obtain an instantaneous rate of occurrence.

The cumulative hazard function, which is the area under the hazard function up to time, t is defined as

$$H(t) = \int_0^t h(u)\,du$$

Finally, the survival function in terms of the hazard function is expressed as

$$S(t) = \exp\left(-\int_0^t h(u)\,du\right) = \exp\left(-H(t)\right)$$

However, let n be subjects censored in survival data and subject i number of death at time t_i with event $\delta_i = 1$, the likelihood function for both survival and hazard function can be written as:

$$\ell = \prod_{i=1}^{n} \ell_i = \prod_{i=1}^{n} S(t_i)H(t_i)$$

With the equation, the log likelihood can be expressed as:

$$\begin{aligned} log\,\ell &= \sum_{i=1}^{n} log\left(\ell_i\right) \\ &= \sum_{i=1}^{n} log\left(S\left(t_i\right)\right)\delta_i log\left(H\left(t_i\right)\right) \\ &= \sum_{i=1}^{n} \delta_i log\left(H\left(t_i\right)\right) - \hbar\left(t_i\right) \end{aligned}$$

Non-Parametric Methods of Survival Models

Using Non-parametric methods in survival analysis, the main characteristic is that there is no assumptions made about the distribution of survival times. The Kaplan-Meier estimate is common when discussing nonparametric survival methods. The Kaplan-Meier product-limit estimator was proposed by Kaplan and Meier in 1958 (E.L. Kaplan 1958).

The Kaplan-Meier Estimate

The Kaplan-Meier estimate of the Survival function $S(t) = \Pr(T > t)$ of T can be expressed as

$$\hat{S}(t)=\begin{cases}1, & 0\le t<t_1\\ \prod_{t_j\le t}\left(1-\frac{d_j}{\tau_j}\right), & t>t_1,\end{cases} \qquad (2.6)$$

where $0<t_1<\ldots<t_D$ denote the distinct uncensored number of event times t_i (Dirk F. Moore 2016), $\tau_j=\sum_{i=1}^{n} I\left(\tilde{T}_i\ge t_j\right)$ is the number of subjects who are "at risk" just prior to time t_j, and $d_j=\sum_{i=1}^{n} I\left(\tilde{T}_i\le t_j,\delta_i=1\right)$ is the number of uncensored events at t_j. Note that $\hat{S}(t)$ is a proper survival function only if the largest observation time is uncensored (E.L. Kaplan 1958).

This estimate contains no assumed parametric distribution. Nonparametric survival methods are particularly useful when we want to compare the survival curves of two groups, such as an experimental group and control group (Dirk F. Moore 2016). Nonparametric methods was examined as an exploratory analysis, but since this method is not able to generate survival probabilities, other methods are utilized more extensively.

Semi-parametric Methods of Survival Models

The most popular Semiparametric model for survival analysis modelling is the Cox proportional hazard model (Cox 1972). Given that X is the vector covariates and $\lambda\left(t\mid X(\mathrm{t})\right)$ is the conditional hazard of T at time t, the hazard function for the Cox model is expressed as:

$$\lambda\left(t\mid X(\mathrm{t})\right)=\lambda_0(t)\exp\left(\beta^T X(t)\right) \qquad (2.7)$$

where $\beta=(\beta_1,\ldots,\beta_n)^T$ is the vector of regression coefficients and $\lambda_0(t)$ is the baseline hazard function (unspecified). The model (2.7) can be rewritten as:

$$\lambda\left(t \mid X(\mathrm{t})\right) = \int_0^t \exp\left(\beta^T X\left(s\right)\right) d\lambda_0\left(s\right)$$

or

$$S\left(t \mid X(\mathrm{t})\right) = \exp\left[-\int_0^t \exp\left\{\beta^T X\left(s\right)\right\} d\lambda_0\left(s\right)\right]$$

where $\lambda\left(t \mid X(\mathrm{t})\right)$ and $S\left(t \mid X(\mathrm{t})\right)$ are the conditional cumulative hazard function and conditional survival function given the covariate history up to t and $\lambda_0\left(t\right)$ is an unspecified baseline cumulative hazard function.

A proportional hazards model stems from the previous idea of wanting to examine the difference between two survival distributions. This difference can be defined using the parameter, ψ, in what is known as the Lehmann alternative, $S_1\left(t\right) = S_0\left(t\right)^{\psi}$ (Dirk F. Moore 2016). Utilizing the relationship between the survival function and the hazard function we know that $h_1\left(t\right) = \psi h_0\left(t\right)$ and association is known as the proportional hazards assumption (Dirk F. Moore 2016). We can also allow the inclusion of covariates in vector z by letting $\psi = e^{z\beta}$ (Dirk F Moore 2016). There are no assumptions made about the distribution of event times with a proportional hazards model (Anon 2012). The partial log-likelihood function does not require a baseline hazard to be specified (Anon 2012). Instead, the model assumes that covariates act multiplicatively on the hazard rate (Anon 2012)

Partial Likelihood

Cox proportional hazards model is a Semiparametric model that extends the proportional hazards model by using the partial likelihood function (Fox and Weisberg 2011). The partial likelihood allows for a baseline survival distribution to be defined by covariates instead of a specific parametric survival distribution (Dirk F. Moore 2016). A basic representation of the Cox proportional hazards model is,

$$\lambda\left(t \mid X(\mathrm{t})\right) = \lambda_0\left(t\right)\exp\left(\beta^T X\left(t\right)\right)$$

where $\lambda\left(t \mid X(\mathrm{t})\right)$ is an unspecified baseline hazard function, β is a vector of regression coefficients, and $\beta_i^T = \left(\beta_{i1}, \beta_{i2}, ..., \beta_{ip}\right)$ is a vector of covariates (Anon 2012). The common method for the model inference on β is the partial likelihood method to eliminate the infinite-dimensional baseline hazard function and it is expressed as:

$$L\left(\beta\right) = \prod_{i=1}^{p}\left\{\frac{\exp\left(\beta^T X_i\left(\tilde{T}_i\right)\right)}{\sum_{j\in\mathbb{R}\left(\tilde{T}_i\right)}\exp\left(\beta^T X_j\left(\tilde{T}_i\right)\right)}\right\}^{\delta_i}$$

where $R(t) = \left\{i : \tilde{T}_i \geq t\right\}$ is the risk set at time t, and the set of individuals who are at risk or alive prior to time t. By maximizing the partial likelihood estimate of Cox model, a joint likelihood of β and λ_0 can be expressed as:

$$L\left(\beta, \lambda_0\right) = \prod_{i=1}^{p}\left[\left\{d\lambda_0\left(\tilde{T}_i\right)\right\}\exp\left(\beta^T X_i\left(\tilde{T}_i\right)\right)\right]^{\delta_i}\exp\left[-\int_0^{\tilde{T}_i}\exp\left(\beta^T X_i(s)\right)d\lambda_0(s)\right]$$

Despite the Cox proportional hazards model's popularity, its proportionality assumptions are often not satisfied. Thus, there is a need for other models that do not use proportionate assumptions.

Parametric Methods of Survival Models

Parametric methods for survival analysis make use of the assumption that the survival times come from a specific distribution (David Collett 2003; Klein and Moeschberger 1997) such as Exponential, Weibull, Log-Logistic, Log-Normal, Extreme Value and Logistic distribution. Consequently, some distribution are applied to actuarial applications, in which left-truncation and right censoring cases such as Gompertz, Perks, Beard, Makeham, Makeham-Perks and Makeham-Beard

distributions are widely used. These distributions are generally used to model explicitly the effect of variables on survival times and likelihood inference is used for different shapes of the hazard function.

In a right-censoring case, the likelihood of parametric model is given by:

$$L(\theta,\delta_i,t_i)=\prod_{i=1}^{p}\{f(t_i;\theta)\}^{\delta_i}\{S(t_i;\theta)\}^{1-\delta_i}$$

or

$$L(\theta,\delta_i,t_i)=\prod_{i=1}^{p}\{h(t_i;\theta)\}^{\delta_i}\{S(t_i;\theta)\}$$

where n represents the number of individuals, θ indicates the parameters estimates, t_i denotes the follow-up time in i^{th} individual and $\delta_i$is the censoring indicator for the i^{th} individual.

In left truncation cases, the probabilities are changed to conditional probabilities using $\frac{f(t_i;\theta)}{S(A_i;\theta)}$ and $\frac{S(t_i;\theta)}{S(A_i;\theta)}$ to replace the probability density and survival function, respectively. The likelihood for left truncation and right-censoring changes to (Klein and Moeschberger 1997):

$$L(\theta,\delta_i,t_i)=\prod_{i=1}^{p}\left(\frac{f(t_i;\theta)}{S(A_i;\theta)}\right)^{\delta_i}\left(\frac{S(t_i;\theta)}{S(A_i;\theta)}\right)^{1-\delta_i}$$

where A_i is the truncation time for the i^{th} individual. The two likelihoods are differ because the left-truncated and the right-censoring likelihood is divided by the survival of the truncation times.

Parametric survival models are based on a distribution for the hazard function, $h(t)$(Dirk F. Moore 2016). A simple survival distribution is the exponential distribution, which has a constant hazard, $h(t)=\lambda$ (Dirk F. Moore 2016). I can derive the cumulative hazard function as:

$$H(t) = \int_0^t h(u)\,du = \int_0^t \lambda\,du = \lambda t\,|_0^t = \lambda t$$

Consequently, I have a survival function of $S(t) = e^{-\lambda t}$ and probability density function is $f(x) = \lambda e^{-\lambda t}$. Several other distributions can be utilized for a parametric survival model depending on the distribution that best fits the data.

Unlike nonparametric survival models, parametric models do generate a survival probability based on covariates. Parametric models lack the flexibility to capture the shape of the hazard function and patient-specific survival predictions are highly dependent on a correct baseline hazard function (Crowther, Abrams, and Lambert 2012).

Time-Independent Covariates

Assume that X is the time-independent covariates and $S(t \mid X)$ is the survival function subject with covariate X can be predicted from Cox model as:

$$\hat{S}(t \mid X) = \exp\left\{\hat{\lambda}_0(t)\exp\left(\beta^T X\right)\right\} \qquad (2.8)$$

Hence, $\hat{S}(t \mid X)$ can be approximated by a normal distribution with mean $S(t \mid X)$ and variance of $\hat{Var}\left\{\hat{S}(t \mid X)\right\} = \left\{\hat{S}(t \mid X)\right\}^2 \left\{\sum_{\tilde{T}_i \le t} \frac{\delta_i}{\left[\sum_{j \in R(\tilde{T}_i)} \exp\left\{\hat{\beta}^T X_j\left(\tilde{T}_i\right)\right\}\right]^2} + W^T M^{-1} W\right\}$

where

$$W=\sum_{\tilde{T}_i\le t}\delta_i\frac{\sum_{j\in R(\tilde{T}_i)}X_j\left(\tilde{T}_i\right)\exp\left\{\hat{\beta}^T X_j\left(\tilde{T}_i\right)\right\}}{\left[\sum_{j\in R(\tilde{T}_i)}\exp\left\{\hat{\beta}^T X_j\left(\tilde{T}_i\right)\right\}\right]^2}-\hat{\lambda}_0(t),$$

$$M=\sum_{\tilde{T}_i\le t}^{p}\delta_i\left[\frac{\sum_{j\in R(\tilde{T}_i)}X_j\left(\tilde{T}_i\right)X_j\left(\tilde{T}_i\right)^T\exp\left\{\hat{\beta}^T X_j\left(\tilde{T}_i\right)\right\}}{\sum_{j\in R(\tilde{T}_i)}\exp\left\{\hat{\beta}^T X_j\left(\tilde{T}_i\right)\right\}}-\frac{\left[\sum_{j\in R(\tilde{T}_i)}X_j\left(\tilde{T}_i\right)\exp\left\{\hat{\beta}^T X_j\left(\tilde{T}_i\right)\right\}\right]\left[\sum_{j\in R(\tilde{T}_i)}X_j\left(\tilde{T}_i\right)\exp\left\{\hat{\beta}^T X_j\left(\tilde{T}_i\right)\right\}\right]^T}{\left[\sum_{j\in R(\tilde{T}_i)}\exp\left\{\hat{\beta}^T X_j\left(\tilde{T}_i\right)\right\}\right]^2}\right]$$

Time-Dependent Covariates

Assume that $X(t)$ is the time-dependent covariates with conditional survival function $S\left(t\mid X(t)\right)$ is estimated either by:

$$\hat{S}\left(t\mid X(\mathrm{t})\right)=\exp\left[-\int_0^t\exp\left(\hat{\beta}^T X(s)\right)d\hat{\lambda}_0(s)\right]$$

Or product-limit estimate

$$\hat{S}\left(t\mid X(\mathrm{t})\right)=\prod_{s\le t}\left[1-\int_0^t\exp\left\{\hat{\beta}^T X(s)\right\}d\hat{\lambda}_0(s)\right]$$

It is nontrivial issue to obtain an analytical variance of $\hat{S}\left(t\mid X(t)\right)$ and this can be done by the bootstrap method.

Accelerated Failure Time Models

Accelerated Failure Time (AFT) model provides an alternative to Proportional Hazard model for statistical modelling of survival data. AFT model is used in industrial fields and seldom applied in the case of survival data. If the appropriate parametric form of AFT model is used then it offers a potential statistical approach in case of survival data, which is based upon the survival curve rather than the hazard function. The AFT model is also known as the log-location scale model.

Accelerated Failure Time Models with Time-Independent Covariate

The accelerated failure time (AFT) describe the relationship between the survival time T of a time covariate X subject and the baseline time T_0 given as $T_0 = T\exp\left(\beta^T X\right)$, where T_0 is the time-independent of X. The AFT model can then be expressed as:

$$S\left(t \mid X\right) = S_0\left\{t\exp\left(\beta^T X\right)\right\}$$

or

$$\lambda\left(t \mid X\right) = \lambda_0\left\{t\exp\left(\beta^T X\right)\right\}\exp\left(\beta^T X\right)$$

where $S\left(t \mid X\right)$ is the survival function at the time T and $S_0\left\{t\exp\left(\beta^T X\right)\right\}$ is the baseline survival function at the time T. The factor $\exp\left(\beta^T X\right)$ is referred to as the acceleration factor. This factor measures the association in the AFT model and is the ratio of survival times corresponding to any fixed value of survival time. The acceleration factor evaluates the effect of predictor variables on survival time. In AFT, if $\exp\left(\beta^T X\right) > 1$, the effect of covariate is decelerated and if $\exp\left(\beta^T X\right) < 1$, the effect of the covariate is accelerated.

Parametric Accelerated Failure Time Models

In a parametric AFT model, the error term is assumed to follow a parametric distribution. Some common parametric AFT models are discussed briefly below.

a. ***Exponential AFT model***: The model assumes that

$$\log\left(T\right) = q_0 + q^T X + \varepsilon$$

where ε is independent of X and follows the standard extreme-value distribution with survival function $P\left(\varepsilon > z\right) = \exp\left\{-\exp\left(z\right)\right\}$. The model is also equivalent to:

$$S\left(t \mid X\right) = S_0\left\{t\exp\left(\beta^T X\right)\right\}$$

where $S_0\left(t\right) = \exp\left\{-\lambda t\right\}$, $\lambda = \exp\left(-q_0\right)$, and $\beta = -q$

Exponential AFT satisfies the proportional hazard function assumption as:

$$\lambda(t \mid X) = \lambda_0(t)\exp(\beta^T X)$$

with an exponential baseline hazard $\lambda_0(t) = \lambda$

b. ***Weibull AFT model***: the model assumes that

$$\log(T) = q_0 + q^T X + \phi\varepsilon$$

where $\phi > 0$ is a scale parameter, and ε is independent of X and follows the standard extreme-value distribution with survival function $P(\varepsilon > z) = \exp\{-\exp(z)\}$. The Weibull AFT model is expressed as:

$$S(t \mid X) = S_0\{t\exp(\beta^T X)\}$$

where $S_0(t) = \exp\{-\lambda t^{\phi}\}$, $\lambda = \exp(-q_0/\sigma)$, $\phi = 1/\sigma$, and $\beta = -q$

The Weibull AFT satisfies the proportional hazard function assumption as:

$$\lambda(t \mid X) = \lambda_0(t)\exp(\phi\beta^T X)$$

with an exponential baseline hazard $\lambda_0(t) = \lambda\phi t^{\phi-1}$. The Weibull AFT model reduces to the exponential AFT model when $\phi = 1$.

c. ***Log-normal AFT model***: the model assumes that

$$\log(T) = q_0 + q^T X + \varepsilon$$

where ε is independent of X and follows a normal distribution with mean 0 and variance σ^2. The Log-normal AFT model is expressed as:

$$S(t \mid X) = 1 - \gamma\left\{\frac{\log(t) - (q_0 - \beta^T X)}{\sigma}\right\}$$

where $S_0(t) = 1 - \gamma\left\{\frac{\log(t) - q_0}{\sigma}\right\}$ and $\beta = -q$

Semi-parametric Accelerated Failure Time Models

Semi-parametric AFT model assumes the following model

$$Y = q_0 + q^T X + \varepsilon$$

where Y is a known monotone increasing transformation of a survival time and ε is independent of X. To make q_0 identifiable, assume further that $E(\varepsilon) = 0$. Alternatively, the model can be expressed as: $Y = q^T X + \varepsilon$ where $\varepsilon = \omega + q_0$

Accelerated Failure Time Models with Time-Dependent Covariate

An AFT model relates the survival time T of a time-dependent covariate X subject to the baseline time variable T_0 that corresponds to the condition $X = 0$ expressed by

$$T_0 = \int_0^T \exp\left(\beta^T X(s)\right) ds$$

where T_0 is assumed to be independent of X and it can also be expressed as:

$$S(t \mid X(t)) = S_0\left[\int_0^T \exp\left\{\beta^T X(s)\right\} ds\right]$$

where $S(t \mid X(t))$ is the conditional survival function of a subject given $X(s), (0 < s \leq t)$ at the time T, the covariate history up to time t, and $S_0(t)$ is the survival function at the time T_0 (Dupuy 2014; Jin 2016; Lin and Ying 1995; Robins and Tsiatis 1992).

This implies that subjects with covariate X fail on expanded time scale $\int_0^t \exp\{\beta^T X(s)\} ds$. Under this model, the subject covariate history $\{X(s), 0 \le s \le t\}$ is accelerated by a factor of $\exp\{\beta^T X(t)\}$ at time t relate to the baseline. However, the model can be written as:

$$\lambda(t \mid X(t)) = \lambda_0 \left[\int_0^t \exp\{\beta^T X(s)\} ds \right] \exp\{\beta^T X(t)\}$$

where $\lambda(t \mid X(t))$ is the conditional hazard of T at time t given the covariate history up to t and $\lambda_0(t)$ is the baseline hazard function.

2.4 joint model overview

Joint models of longitudinal and survival data have received increased attention from scholars in the past two and half decades. The models are useful for analysing survival models with missing data or measurement errors in time-dependent covariates. It could also be used for longitudinal models with informative dropouts as well as the associated structures between the longitudinal and survival processes via latent variables. In practice, joint model of longitudinal and survival data often occur simultaneously. For instance, in several biomedical researches, patients' information (such as CD4 cell) are frequently collected repeatedly over period with interest on the time to recovery or recurrence of a disease. However, there may be association between the longitudinal trajectories and the time-to-event. Using a separate analysis for the association between the longitudinal trajectories and the time-to-event may lead to biased and inefficient estimates if the two outcomes are correlated. Therefore, joint models are essential to use in order to incorporate all information simultaneously for valid and efficient estimates and inferences.

The motive behind joint modelling in medical research emanated from four broad scientific situations:

i. Ways to improve inference for a longitudinal measurements with an informative dropout
ii. To improve inference for survival measurements with an intermittent and endogenous time-dependent covariates
iii. To study the correlated association structure between the longitudinal trajectories and the time-to-event
iv. To improve the inference efficiency from the external information

From previous studies, several approaches have been developed for modelling longitudinal and survival data simultaneously such as combining the prediction association of longitudinal measurements and time to an event (Wulfsohn and Tsiatis 1997). Joint models with an inclusion of endogenous and time-dependent covariates to model the relationship between the covariate and risk of the event (Rizopoulos, 2011). In some proposed methods, joint modelling reduces biasness in the extended Cox model and the two-stage model and also improves predicted survival probabilities (Hickey et al. 2016). However, a study shows that joint modelling has been used to determine a surrogacy to an event such as cancer biomarkers as an indicator of cancer progression or regression (Elashoff, Li, and Li 2007). A classic approach of joint models setting is the use of mixed-effects model for the longitudinal data and a Cox models or an Accelerated failure time (AFT) for the survival data, with a shared random effects. Most times, the likelihood approach is used for the parameter estimates through EM algorithm.

Past studies have tended to focus on the 'Univariate' joint modelling, in which a single longitudinal and a single survival response are jointly modelled. In practice,

most of the data collected are frequently complex. It is imperative to collect multiple longitudinal outcomes and possibly multiple survival outcomes such as recurrent outcomes from patients to have a good joint modelling analysis. This is referred to as 'multivariate joint modelling'. An increase in the flexibility and predictive power from the extension of the classical Univariate joint model structure to a multivariate framework brings a number of computation intensity challenges.

There have been several attempts to extend joint analysis for a single longitudinal outcome to the situation where multiple outcomes are simultaneously recorded with the event times. A key issue in modelling multivariate longitudinal data and their association with event times is to formulate the joint evolution of these multiple endpoints. In a standard joint models formulation, the underlying current subject-specific value of a biomarker is assumed to associate with the risk of an event happening at the same time t with an association parameter α, which may not be enough to describe the association structure. In the light of this, different studies have proposed an alternative association structure between the longitudinal and time-to-event outcomes (Brown, Ibrahim, and Degruttola 2005; Rizopoulos et al. 2014; Ye, Lin, and Taylor 2008). Of note, the model for longitudinal outcomes needs to take into account two outcomes association; first, the risk of an event allowed to depend on the longitudinal profile slope, and secondly, the risk of an event that depends on the cumulative effect. These data associations are usually characterized by latent variables, either continuous (random effects) or discrete (latent classes), which are also used to link the longitudinal outcomes and survival data.

In the setting of standard Cox model, many studies have proposed the use of weighted cumulative models (Breslow et al. 1983; Thomas 1988). In another study, a parametric time-dependent weight function was included as an extension to Cox

model (Abrahamowicz et al. 2006). In other studies on joint model of survival-longitudinal data, the association structure is often evaluated by either linking the value of longitudinal outcomes into the survival model or by using the shared random-effects models to examine scientifically the association between the survival time and the underlying unobserved longitudinal processes (Henderson et al. 2000; Tsiatis and Davidian 2004a).

2.5 Joint Modelling for Longitudinal and Time-to-Event Data

Joint modelling is an extension of survival data analysis, which combines the prediction association of longitudinal measurements and time to an events (Wulfsohn and Tsiatis 1997). In biostatistics, a longitudinal biomarker measured repeatedly over time may be used to predict occurrence of an event such as death or inception of a disease. Joint modelling plays a significant role and allows the inclusion of endogenous covariates and time-dependent covariates in a model to see relationship between the covariate and the risk of the event (Anon 2012; Rizopoulos 2011a). In some previous studies, joint modelling reduces biasness in the extended Cox model and the two-stage model also improves predicted survival probabilities (Hickey et al. 2016). However, Another study however, shows that joint modelling can be used to determine surrogacy to an event such as cancer biomarkers as an indicator of cancer progression or regression (Elashoff et al. 2007).

2.5.1 Mixed-Effects Models for Longitudinal Sub-model

Let $y_i(t)$ denote the observed value of the longitudinal outcome for the ith individual (i =1,….,n) at time t with a particular time points $t_{ij}, j = 1.....n_i$. Therefore, the mixed-effect model is expressed as:

$$\begin{cases} y_i(t) = x_i(t)^T \beta + z_i(t)^T \mu_i + \varepsilon_i(t), \\ \mu_i \sim N(0, E), \\ \varepsilon_i \sim N(0, \sigma^2) \end{cases}$$

Where β is the regression coefficients of a matrix vector for the fixed effects x_i, z_i is the row vectors matrix for the random effects u_i. *E* denotes the random effects covariance variance matrix, $\varepsilon_{ij}(t)$ represents the error terms and σ as the variance of the error term.

2.5.2 Cox Models for Survival Data Sub-model

Assume that T_i^* represents the true failure time (FT) for the ith patients with C_i censoring time and $T_i = min(T_i^*, C_i)$ denotes the observed FT for ith patients. However, hazard function from the Cox model is assumed to satisfy the relationship:

$$h_i(t) = \lim_{dt \to 0} \frac{P(t \le T^* < t + dt \mid T^* \ge t)}{dt} = h_0(t) \exp(\alpha^{|} w_i), \qquad t > 0$$

where w_i is the associated covariates with the hazard function and $\alpha^{|}$ represents the corresponding regression coefficients vector, $h_0(t)$ denotes the baseline hazard function. It is assumed that the hazard ratio is only determined by the covariates, whose value is fixed during the follow-up, such as age, gender and treatment outcomes.

2.6 Two-Stage Joint model Approach

A common approach used for joint model of longitudinal and survival model is based on two-stage methods, which seems to be simpler in computation. In a study by Henderson et al (2000), different random effects were allowed in the longitudinal

and survival models with the assumption that the within subject-specific random effects are correlated (Henderson et al. 2000). Others studies propose the use of Bayesian approaches for the joint models (Huang, Hu, and Dagne 2014; Ibrahim, Chen, and Sinha 2004). In the literature, various kinds of two-stage approaches have been developed. A naïve two-stage approach has the following stages:

Stage 1: A linear mixed-effect model is fitted for the longitudinal data and the covariates measured/mismeasured are estimated from the fitted model.

Stage 2: The estimates from the first stage is then substituted to fit the survival model separately with unobserved true covariates values.

Modified two-stage approaches have developed to serve as an advantage over the naïve methods for simplicity and ease of implementation with an existing software. The constraint of these approaches is that they may lead to biased estimates and inferences. There are many reasons for this. Firstly, if the truncations results from the events are not included in the longitudinal covariates model parameter estimation, the longitudinal outcomes will produce bias estimate. Secondly, if the estimation of uncertainty in the first stage is not integrated in the second stage for the estimation of survival model hence, the standard errors of the estimates in the survival model may be underestimated. Thirdly, the longitudinal and survival processes information are not fully combined in each model fit to result in the most efficient estimates. In this sense, various modified two-stage approaches have been developed to address the bias estimation issues in the longitudinal model parameters caused by ignoring informative truncation and depending on the strength of the association between the longitudinal and survival processes.

2.6.1 The naïve two-stage method and full likelihood approach

Suppose in a sample of n subjects, let Y_{ij} represents the longitudinal measurements collected on each subject at times $t_{ij}, i = 1,....,n$, $j = 1,....,n_i$, where n_i is the number of measurements on subject i and $Y_i = \left(Y_{i1},...,Y_{in_i}\right)$. Denote T_i to be the event time for subject i, which is subject to C_i censoring assumed to be non-informative, where C_i is independent of T_i given covariates X_i. The observed event time for subject i are $\tilde{T}_i = min\left(T_i, C_i\right)$ and $\Delta_i = I\left(T_i \le C_i\right)$, such that $\Delta_i = 1$ if $T_i \le C_i$ and $\Delta_i = 0$ if $T_i > C_i$. The longitudinal data consist of the measurements of the *i*th subject $y_{ij} = \left\{y_i(t_{ij}), j = 1,2...,n_i\right\}$ taken at time points t_{ij}.

The joint model for longitudinal and survival data is postulated from a Cox hazard model in the form:

$$\lambda_i\left(t \mid \mathrm{M}_i(t), w_i\right) = \lambda_0\left(t\right) exp\left(\gamma^{\mathrm{T}} w_i + \alpha m_i(t)\right) \qquad (1)$$

$$h_i(t) = h_0(t)\exp\left[\alpha^T w_i + m\left\{x_i(t)^T\beta + z_i(t)^T b_i\right\}\right], \qquad t > 0$$

where $\lambda_0(t)$ is the unspecified baseline hazard function and w_i is a baseline covariates vector. Also, $\mathrm{M}_i(t) = \left\{m_i(\mu), 0 \le \mu < t\right\}$ represents the history of the true unobserved longitudinal process up to time point t_{ij}. The longitudinal sub-model is given as:

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t), \\ m_i(t) &= x_i(t)^T\beta + z_i(t)^T\mu_i + \varepsilon_i(t),\ \varepsilon_i(t) \sim N(0, \sigma_\varepsilon^2) \end{aligned} \qquad (2)$$

where x_i and z_i are matrices of the fixed effects and random effects respectively. β is a vector of regression coefficient. The naïve two-stage approach for the joint models is specified as proposed by (Tsiatis and Davidian 2004b). In this method, the longitudinal model is first fitted separately and the fitted values from the longitudinal analysis, $\hat{m}_i(t)$ are incorporated as covariates in the joint model fitting.

Based on the limitations of the existing methods, Huong et al (2018) recently proposed a modified two-stage approach, in which the fitted values of the parameters from the longitudinal model were used and the expected function was approximated for the complete data log-likelihood. In contrast to the use of the partial likelihood method to estimate the coefficients regression for the relative risk model, Huong et al (2018) applied the approximation approach for the full likelihood method. Newton-Raphson algorithm was used for the implementation in the second stage. The two-stage method for the joint model by Huong et al (2018) is described as follow:

Stage 1: Linear mixed-effects model is fitted for the longitudinal process from the longitudinal sub model as:

$$\tilde{y}_i(t_{ij}) = \hat{m}_i(t_{ij}) + \varepsilon_i(t_{ij}) = x_i^T(t_{ij})\beta^* + z_i^T(t_{ij})b_i^* + \varepsilon_i(t_{ij}),\ \varepsilon_i(t_{ij}) \sim N(0, \sigma_\varepsilon^2) \qquad (3)$$

Stage 2: The joint model is then fitted using the estimates from the fitted parameters in stage 1 in the form of:

$$\lambda_i(t) = \lambda_0(t)exp\left(\gamma^T w_i + \alpha\hat{m}_i(t)\right), \hat{m}_i(t) = x_i^T(t)\tilde{\beta} + z_i^T(t)\tilde{b}_i \qquad (4)$$

The advantage of Huong et al (2018) proposed method is the ease of implementation when the standard mixed-effects software is used for the stage 1 and software for survival model for the stage 2.

2.7 A standard Joint model for longitudinal and survival data

Joint models are used to deal with statistical issues that cannot be handled in separate model analysis of longitudinal and survival data. It is a refined approach to model the association between time-dependent covariates and the event of interest when the covariate trajectory is not completely observed and subject-specific to measurement error. Most times, two types of data are collected for a longitudinal study for each subject-specific variable, namely, repeated events of longitudinal measurements during follow-up observation, and process of events in time-dependent covariates. The joint modelling approach assumes that the random effects account for the correlation between the longitudinal repeated measures as well as the association between the longitudinal outcome and the survival events; the random effects are shared between the two processes (Rizopoulos 2010). The joint model is the duo of a survival model for both the categorical and continuous time-to-event process with a mixed-effects model for repeated data. However, the joint model for longitudinal and survival data, therefore, enables one to study the likelihood approach framework, where generalized linear mixed-effects models measure the longitudinal process and the time-dependent event is measured by Cox regression model.

In a sample of n subjects, let Y_{ij} represents the longitudinal measurements collected on each subject at times $t_{ij}, i = 1,....,n$, $j = 1,....,n_i$, where n_i is the number of measurements on subject i and $Y_i = \left(Y_{i1},...,Y_{in_i}\right)$. Denote T_i to be the event time for

subject i, which is subject to C_i censoring assumed to be non-informative, where C_i is independent of T_i given covariates X_i. The observed event time for subject i are $\tilde{T}_i = min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$, such that $\Delta_i = 1$ if $T_i \leq C_i$ and $\Delta_i = 0$ if $T_i > C_i$. In models selection approach, marginal distribution of Y_i would be specified in which generalized linear mixed effects models within-subject correlations are modelled by latent random effects b_i, where b_i or $f(b_i)$ indicates the underlying trajectory/trend in Y_i. However, under the framework of generalized linear mixed effect model, the distribution of Y_{ij} is expressed as:

$$g\{E(Y_{ij} \mid b_i)\} = X_{ij}^T \beta_1 + \tilde{X}_{ij}^T b_i,$$

where $g(.)$ is a link function. In the joint model literature, T_i is usually characterized using parametric or semi-parametric Cox models. A parametric model can be specified as

$$g\{E(T_i \mid b_i)\} = X_i^{'T} \beta_1 + v b_i,$$

where $X_i^{'}$ is a vector of covariates that are associated with the event process, and $g(.)$ is a known link function. Under the Cox regression framework the hazard of T_i is specified as

$$\lambda_i(t \mid b_i) = \lambda_0(t) exp\left(X_i^{'T} \beta_1 + v b_i\right),$$

where $\lambda_0(t)$ is a parametric or completely unspecified baseline hazard function. If $b_i = (b_{0i}, b_{1i})$ contains the random intercept and slope for Y_i, then the model assumes that the subject-specific starting value and time trend in the underlying trajectory of

Y_i affect the event risk. A natural extension is that a function of b_i, e.g., $b_{0i} + b_{1i}t$ is treated as a time-dependent covariate in model.

The joint model can also be specified if assumed to non-zero-mean normal distribution for a shared parameter model and if the parameter is shared by both longitudinal and survival data according to Rizopoulos (2010).

The joint model for a shared parameter model is expressed as:

$$\begin{cases} y_i(t) = x_i(t)^T \beta + z_i(t)^T \mu_i + \varepsilon_i(t), \\ h_i(t) = h_0(t) \exp\left[\alpha^T w_i + m\left\{x_i(t)^T \beta + z_i(t)^T \mu_i\right\}\right], \qquad t > 0 \end{cases}$$

where *m* computes the effect of the repeated biomarkers with the risk of an event. However, it is assumed that the risk of time-dependent outcome is correlated with the true and unobserved value of the repeated biomarkers.

Rizopoulos (2010) showed an extended joint model to allow an additional source of variation at the survival endpoint that cannot be explained by the longitudinal data. Specifically, a separate random variable γ_i can be introduced into the model, so that we have

$$\lambda_i\left(t \mid b_i, \gamma_i\right) = \lambda_0\left(t\right) exp\left(X_i^{'T} \beta_1 + v b_i + \gamma_i\right),$$

Alternatively, the survival model can be specified as:

$$\lambda_i\left(t \mid \gamma_i\right) = \lambda_0\left(t\right) exp\left(X_i^{'T} \beta_2 + \gamma_i\right),$$

And γ_i and b_i are assumed to have a multivariate normal distribution with variance-covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{bb} & \Sigma_{b\mu}^{T} \\ \Sigma_{b\mu} & \sigma_{\mu}^{2} \end{pmatrix}.$$

Joint modelling can be used to predict that the event risk at the given time *t* depends on the mean level of biomarkers at the same time period *t*,

$$M_1 = h_{0k}(t)\exp\left\{\alpha_k{}^{T} w_{ik} + \sum_{P=1}^{P} m_{pk} f_{ip}(t)\right\}$$

It can be used as an extension of model M_1, which is not only the current value but also the slopes of biomarker trajectories at the time *t* related to the hazard,

$$M_2 = h_{0k}(t)\exp\left\{\alpha_k{}^{T} w_{ik} + \sum_{P=1}^{P} m_{pk} f_{ip}(t) + \sum_{P=1}^{P} m_{pk}^{d} f_{ip}^{'}(t)\right\}$$

where $f_{ip}^{'}(t) = \frac{d}{dt} f_{ip}(t)$. Since $f_{ip}^{'}(t)$ denotes the slope of the longitudinal outcome over a period, m_{pk} is held constant.

The third extension of joint modelling considers the prediction of the Model M_3 as a cumulative value of the longitudinal outcome, which relates the survival outcomes with a summary area under the longitudinal profiles of the whole history of the markers.

$$M_3 = h_{0k}(t)\exp\left\{\alpha_k{}^{T} w_{ik} + \sum_{P=1}^{P} m_{pk}^{d} \int_{0}^{t} f_{ip}(s)ds\right\}$$

However, the model is not depending only on the current value of the longitudinal outcome but also on the cumulative value in time *t* calculated by the integral of $f_{ip}(s)$.

Rizopoulos (2010) proposes a combination of all the models expressed to give:

$$M = h_{ik}(t,\theta_s) = h_{0k}(t)\exp\left\{\alpha_k^{|} w_{ik} + \sum_{P_1=1}^{P_1} m_{p_1 k} f_{ip_1}(t) + \sum_{P_2=1}^{P_2} m_{p_2 k}^{d} \int_0^t f_{ip_2}(s)ds\right\}$$

This model assumes that the risk for an event at a particular time *t* is associated with the underlying value of *P*1 biomarkers at a specific time point and the area under the curve for *P*2 biomarkers at the same time point.

2.8 Joint model likelihood

The likelihood function of joint modelling of the observed data

$$L(\theta, b \mid D) = \prod_{i=1}^{N}\left(\prod_{j=1}^{n_i} P\left(y_i(t_{ij}) \mid \theta, b_i\right) P\left(T_i, \delta_i \mid \theta, b_i\right)\right)$$

where

$$P\left(y_i(t_{ij}) \mid \theta, b_i\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n_i}\left[y_i(t_{ij}) - x_i(t)^{|}\beta - z_i(t)^{|}\mu_i\right]^2\right\}$$

and

$$P\left(T_i, \delta_i \mid \theta, b_i\right) = h_0(T_i)\exp\left[\alpha^{|} w_i + m\left\{x_i(t)^{|}\beta + z_i(t)^{|}\mu_i\right\}\right]^{\delta_i} \times \exp\left\{-\int_0^{T_i} h_0(\mathrm{u})\exp\left[\alpha^{|} w_i + m\left\{x_i(t)^{|}\beta + z_i(t)^{|}\mu_i\right\}\right]du\right\}$$

where $P(b_i)$ is the probability density function of random effects and θ is the parameter vector of the joint model. Given the unobserved "true" value of the longitudinal process $m\left\{x_i(t)^{|}\beta + z_i(t)^{|}\mu_i\right\}$, the observed time to event and longitudinal data are assumed to be mutually independent.

Maximum likelihood estimation is one of the standard approaches to obtain the parameters estimates and statistical inference for the joint modelling, in which the integration with respect to random effects is required. Due to the complexity of the joint model, the integration usually turns out to be quite a challenge. The EM

algorithm can be successfully applied to the joint modelling for inference, and the expectation of any function of random effects is evaluated by using an m-points Gauss-Hermite quadrature formula. A Monte Carlo approach is also quite straightforward to assist in obtaining the expectation of functions of random effects or using fully exponential Laplace approximation for these integrations with respect to random effects.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter describes the method for data collection, models specification and the proposed joint models for the study. It entails the detailed account of the research design according to the structure of the objectives of the study. It also outlines the ethical consideration, and how it was addressed. It discusses the challenges around the statistical models used in the joint model of longitudinal and survival data and proposed a method to deal with such challenges. In this chapter, we used the existing models to see the challenges in the model fit and provide an improved model to overcome the challenges.

3.2 Data Collection and Design

Data were collected from Grey Hospital in King Williams Town, Eastern Cape Province, South Africa. In total, 612 tuberculosis (TB) patients, with impaired renal function were sampled out of the 2987 patients on observation study whose data was recorded and stored in computer database files of the hospital. The data was a special consideration applied to the treatment of TB in patients with impaired renal function from March 2008 to March 2018.

The main complaint of the patients admitted into the study were patients with at least 3 months history of TB, patients with marked loss of weight such as 20kg in 3 months, patients with generalized abdominal pains and those with severe fever and night sweats. The background co-infection considered were patients with HPT,

dyslipidemia, newly diagnosed HIV, a CD4 count of 13, and started on AZT/3TC/EFV 2 weeks before the study. The study also considered some clinical examinations like vitals if normal, generalized shotty nodes, bipedal oedema and all other systems if they are normal. The epidemiology of TB in patients with renal disease is seen if there is an increased risk incidence and prevalence of TB in End-stage renal disease (ESRD) and dialysis patients. Also, if there is an increased rate of opportunistic infections (OIs) especially TB in HIV-positive hemodialysis patients versus HIV-negative patients and if there is an increase in mortality of these patients.

3.2.1 Ethical Consideration

Govan Mbeki Research and Development Centre (GMRDC), University of Fort Hare, Alice with ethical reference number QIN111SAZE01 and the Eastern Cape Department of Health, Eastern Cape, South Africa, with reference number-EC_201808_007 approved this research. Subjects were identified with numbers instead of their names and all information regarding the study was kept confidential.

3.3 Joint Model Estimation Methods

In medical research, it is very common to collect data on measurements of longitudinal outcomes such as a continuous response and the time-to-event outcomes of each individual as well as other additional covariates data. Most times, the interest is to capture the longitudinal processes trajectory and/ or to specify the association characteristics between the longitudinal processes and the survival outcomes with other additional covariates. A popular approach is to assume a linear mixed-effects model for longitudinal data (Laird and Ware 1982) and the use of Cox proportional hazards (PH) model for the survival outcomes (Cox 1972). However, in

this thesis, we are addressing two joint modelling challenges that usually occur in analyzing the joint model of longitudinal and survival data.

1. The challenge of the association structures between longitudinal covariates and time-to-event process in examining the within-subject measurement error that could influence the estimation, when the assumption of normality and the mutual independence is violated.
2. The model for longitudinal outcomes taking into account the association of the outcome from the risk of an event that depends on the longitudinal profile slope and the risk of an event that depends on the cumulative effect.

Longitudinal outcomes are usually collected with time, and is used as time-dependent covariates when modelling survival outcomes. To apply the Cox PH model with time-dependent covariates, complete understanding of the true covariates account is required for each individual. Generally, time-dependent covariates are often measured at irregular intervals or intermittently with error. This complicates and often leads to estimation biases in the survival model if mis-measured values are substituted for true covariates in the proportional hazards model (Prentice 1982). The primitive idea is to simply assign or attribute the unknown value to ordinary least square estimate, which also show biased inference in the estimation process.

3.4. Joint Model with Correlated Measurement Error for Repeated Covariates

Many research methodologies have proposed a reduction in the bias estimate for the parameter inference association structures between longitudinal covariates and time-to-event process. This is because of the failure time due to the within-subject measurement error that could influence the estimation when the assumption of

normality and mutual independence is violated (Bartlett and Keogh 2018; Lash et al. 2014) or for the available longitudinal outcomes collected intermittently only (Van Den Houten et al. 2019). In order to reduce the bias of the primitive approach, regression calibration is used to replace the unknown covariates by available observed data regression, and may still give biased estimates (Spiegelman, Rosner, and Logan 2000; Wang et al. 1998; Wang, Lin, and Guttierrez 1999). It may, therefore, yield incorrect results for a measurement error that is large and/or large relative risk of interest (Wang, Wang, and Wang 2000; Wulfsohn and Tsisatis 1997). Also, a corrected score approach was proposed in order to reduce the estimate bias to deal with the time-independent covariates of interest like random effects, which usually leads to consistent and robust estimators (Kosmidis 2014; Kosmidis and Firth 2010). Likelihood approach has been consistently and efficiently used with the assumption of normality for both random effects and the measurement error (Faucett and Thomas 1996; Henderson et al. 2000; Rizopoulos 2011b; Wulfsohn and Tsisatis 1997). Likelihood approaches can be very intensive computationally due to the integration of the random effects. A conditional score estimation using the generalized linear model was proposed in a study to provide adequate statistic for the unknown covariates, and also used conditional estimating equation for the parameter inference (Stefanski and Carroll 1987). In another study, a conditional score estimator for Cox proportional model was proposed for the parameter to avoid any underlying random effects assumptions (Tsiatis and Davidian 2001). Song and Wang (2014) improved the method to show the consistency and robustness of the estimates.

When modelling the joint longitudinal-survival data, the normality assumption is considered on measurement error and unobserved random effects. In the literature,

the normality assumption is taken for granted in the sense that within-subject random errors are normally distributed and mutually independent. For distributional normality assumption, the measurement error can, to some extent, still correlate but the random effects may not exactly account for all the underlying association. It is also noted in the literature that misspecification of the covariance structures could contribute to statistical inference error (Daniels and Zhao 2003; Henderson et al. 2000; Tsiatis and Davidian 2004b; Wang, Carroll, and Liang 1996).

3.4.1 The Proposed Estimation Method for Covariate with Correlated Error

This chapter examined the within-subject measurement error that could influence estimation when the assumption of normality and mutual independence is violated by using the conditional score (CS) approach proposed by (Tsiatis and Davidian 2001). Given the potential issues of assumption violation, an estimating equation approach based on generalized conditional score (GCS) is proposed to relax parametric distributional assumptions for random effects and is relatively easy to implement. The main interest is to jointly model the time-dependent biomarkers and event times, in which a longitudinal model is used to characterize the biomarker underlying (unobservable) trajectory, which is incorporated as a latent, time-dependent covariate in the survival model to predict failure times for Cox models with intermittently measured Time-dependent covariates. Then, the proposed method is compared with other existing methods (such as Idea and naïve regression (NR) methods) to determine the best performing method in examining the within-subject measurement error that could influence estimation when the assumption of normality and mutual independence is violated.

3.4.2 Model definition

For subject i let $Y_i(t)$ denote the underlying, smooth trajectory of biomarker, T_i the event time, and $Z_i(t)$ a set of possibly time-varying covariates. Because the biomarker is intermittently measured at time points t_{ij} and there are intra-subject measurement errors, $Y_i(t)$ is not directly observable; instead, measurements of $W_i(t_{ij})$ are available such that

$$W_i(t_{ij}) = Y_i(t_{ij}) + \varepsilon_i(t_{ij}) \qquad (1)$$

where $\varepsilon_i(t_{ij})$ are measurement errors. In reality, the event time T_i may be subject to right censoring so we observe $\bar{T}_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \le C_i)$, where C_i is the censoring time. Measurements of $W_i(t_{ij})$ will then be available at time points $t_{ij} \le \bar{T}_i$, $j = 1,..,n_i$.

The longitudinal model focuses on characterizing the change in $Y_i(t)$ overtime. If the change can be described by a polynomial function or splines of time t, then $Y_i(t)$ is specified as

$$Y_i(t) = \rho(t)^T b_i, \qquad (2)$$

where b_i is a vector of subject-specific random effects. A simple example is $Y_i(t) = \rho_{0i} + \rho_{1i}t$, or more flexibly $Y_i(t) = \rho_{0i} + \rho_{1i}t + ... + \rho_{pi}t^p, b_i = (b_{0i},..,b_{pi})^T$. Model (2) assumes the trajectory of $Y_i(t)$ is determined by a relatively small set of time-fixed random effects, which may not fully account for within-subject variation overtime. A mean-zero stochastic process $U_i(t)$ is thus added to the model such that

$$Y_i(t) = \rho(t)^T b_i + U_i(t) \qquad (3)$$

The process $U_i(t)$ is usually assumed to be independent of b_i and covariates $Z_i(t)$. Examples of $U_i(t)$ include integrated Ornstein-Uhlenbeck (IOU) process and stationary Gaussian process. In fact, $U_i(t)$ captures biological fluctuations around the smooth trend $\rho(t)^T b_i$ and induces an additional within-subject autocorrelation structure on top of that by b_i. Choosing in between (2) and (3) reflects the belief whether the “inherent," dominant time trend in the biomarker is associated with T or the biological fluctuations are important features we should capture as well when characterizing the survival model.

In (2) and (3), random effects b_i are commonly assumed to be normally distributed and the mean and covariance matrix may depend on $Z_i(t)$. Measurement errors $\varepsilon_i(t_{ij}) \sim N(0, \sigma^2)$ are independent of b_i and $U_i(t)$ for all $t \geq 0$. Under (3), $\varepsilon_i(t_{ij})$ takes into account measurement error as well as local, transient biological variation that is unlikely to carry over across j, so the independence assumption for $\varepsilon_i(t_{ij})$ at different t_{ij} is reasonable. Model (3) reduces to (2) when the stochastic process $U_i(t)$ is absorbed into $\varepsilon_i(t_{ij})$ so that $\varepsilon_i(t_{ij})$ contains both measurement error and local biological fluctuations. In this case, a covariance structure may be necessary for $\varepsilon_i(t_{ij})$ to characterize within-subject correlation over time. If time-period is too long, the within-subject correlation due to biological variation will be negligible, or if measurement error is in a larger scale than biological variation, the independence assumption would still hold approximately.

A Cox proportional hazards model is used to specify the interrelationship between $Y_i(t)$, T_i, and $Z_i(t)$:

$$\lambda_i(t) = \lim_{dt \to 0} \frac{P\{t \le T_i < t + dt \mid T_i \ge t, \bar{Y}_i(t), Z_i(t)\}}{dt} = \lambda_0(t) \exp\{\gamma Y_i(t) + \eta^T Z_i(t)\} \qquad (4)$$

where $\bar{Y}_i(t) = \{Y_i(\mu), 0 \le \mu \le t\}$ is the history of the biomarker up to time t. This specification implies that given covariates and history $\bar{Y}_i(t)$, the biomarker is associated with the event risk through its current value $Y_i(t)$. Alternative specifications are possible, for example, $\lambda_i(t) = \lambda_0(t) \exp\{\gamma b_{1i} + \eta^T Z_i(t)\}$, where b_{1i} is the random slope if $Y_i(t)$. This formulation is applied to circumstances where it is believed that, given $\bar{Y}_i(t)$ and covariates, the main force that drives the event process is the underlying constant rate of change in $Y_i(t)$. The Cox model proposed shows that the censoring is non-informative, the random effects b_i and failure time are both independent of the longitudinal measurements observed.

Many researchers using standard statistics software like R and SAS have studied parameter estimation and statistical inference for the joint model for the linear mixed effects with different potential covariance structures for within-subjects. In many of these studies, ε_i are assumed to be mutually independent with constant variance i.e. $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, may not be true all the time. In this study, we examined the mutually independent assumption violation influence of ε_i on the statistical parameter estimation and inference of survival without any distributional assumption made on random effects b_i.

3.4.3 Conditional Score Estimator

A conditional score estimator (CS) was first proposed by (Tsiatis and Davidian 2001). It is an estimating-equation based method, that condition on a “sufficient statistic” for random effects b_i in respect to unknown parameters in the Cox model

when the underlying time-dependent covariate follows a linear mixed effect model. Consider a random intercept and slope model for $Y_i(t)$:

$$Y_i(t) = b_{0i} + b_{1i}t. \quad (5)$$

Assume measurement errors $\varepsilon_i(t_{ij})$ in (1) are i.i.d. normal variables $N(0,\sigma^2)$ given random effects b_i, covariates $Z_i(t)$, $\bar{Y}_i(t_{ij})$ and that the biomarker is measured at t_{ij} and subject i is at risk at t_{ij}. Let $\hat{Y}_i(t)$ be the ordinary least squares estimator for $Y_i(t)$ based on measurements up to t. In the case of (5), $\hat{Y}_i(t) = (1,t)\hat{b}_i$, where $\hat{b}_i$ is the ordinary least squares estimates intercept and slope for subject i. Note that $\hat{b}_i$ is defined for subjects with at least two measurements by time t. Let $J_i(t)$ be the at-risk indicator for subjects with at least two measurements at time t, such that $J_i(t) = I(\bar{T}_i \geq t, t_{i2} \leq t)$. Under the assumption that censoring or seeing a measurement at time t given b_i and biomarker data prior to t are not dependent on measurement errors prior to t, it can be shown that

$$\hat{Y}_i(t) \mid J_i(t) = 1, \tau_i(t), b_i, Z_i(t) \sim N(Y_i(t), \sigma^2\theta_i(t)),$$

where $\tau_i(t) = \{t_{ij}, t_{ij} < t\}$ and $\sigma^2\theta_i(t)$ is the variance of predicted value $\hat{Y}_i(t)$. In the case of (5),

$$\theta_i(t) = {}^{1}\!\Big/ n_{i,t} + (t - \bar{t}_{i,t})^2 \Big/ \sum_{j=1}^{n_{i,t}} (t_{ij} - \bar{t}_{i,t})^2,$$

where $n_{i,t}$ is the number of time points in $\tau_i(t)$ and $\bar{t}_{i,t}$ is the mean of the $n_{i,t}$ time points.

The following derivation of conditional score estimating equations assumes σ^2 is known. Extension to the case of unknown σ^2 is discussed next.

Define the counting process $dN_i(t) = I(t \le \bar{T}_i < t + dt, \Delta_i = 1, t_{i2} \le t)$. The conditional density for the joint distribution of $\{dN_i(t) = r, \hat{Y}_i(t) = y\}$ is

$$P\{dN_i(t) = r, \hat{Y}_i(t) = y \mid J_i(t) = 1, \tau_i(t), b_i, \mathrm{Z}_i(t)\}$$
$$= P\{dN_i(t) = r, \hat{Y}_i(t) = y, J_i(t) = 1, \tau_i(t), b_i, \mathrm{Z}_i(t)\} \times P\{\hat{Y}_i(t) = y \mid J_i(t) = 1, \tau_i(t), b_i, \mathrm{Z}_i(t)\} \qquad (6)$$

The first component from (6) is a Bernoulli variable with probability given as: $\lambda_0(t)dt \exp\{\gamma Y_i(t) + \eta^T Z_i(t)\}$, and the second is a normal variable $N(Y_i(t), \sigma^2\theta_i(t))$. After simplifications it can be shown that to order dt, (6) is equal to

$$\exp\left[Y_i(t)\left\{\frac{\gamma\sigma^2\theta_i(t)dN_i(t) + \hat{Y}_i(t)}{\sigma^2\theta_i(t)}\right\}\right] \times \frac{\left\{\lambda_0(t)\exp\{\eta^T Z_i(t)\}dt\right\}^{dN_i(t)}}{\{2\pi\sigma^2\theta_i(t)\}^{1/2}} \exp\left\{-\frac{\hat{Y}_i^2(t) + Y_i^2(t)}{2\sigma^2\theta_i(t)}\right\},$$

Which, indicates that a "sufficient statistic" for b_i is $S_i(t, \gamma, \sigma^2) = \gamma\sigma^2\theta_i(t)dN_i(t) + \hat{Y}_i(t)$. Conditional on $S_i(t, \gamma, \sigma^2)$, it can be shown that

$$\lim_{dt \to 0} P\{dN_i(t) = 1 \mid S_i(t, \gamma, \sigma^2), \mathrm{Z}_i(t), \tau_i(t), J_i(t)\} / dt$$
$$= \lambda_0(t)\exp\{\gamma S_i(t, \gamma, \sigma^2) - \gamma^2\sigma^2\theta_i(t)/2 + \eta^T Z_i(t)\}J_i(t)$$
$$= \lambda_0(t)E^*_{oi}(t, \gamma, \eta, \sigma^2).$$

This suggests that the conditional density of $dN(t) = \sum_{i=1}^{n} dN_i(t)$ given $S_i(t, \gamma, \sigma^2), \mathrm{Z}_i(t), \tau_i(t), J_i(t)$, $i = 1, .., n$, is $\lambda_0(t)E^*_{oi}(t, \gamma, \eta, \sigma^2)$, where $E^*_o(t, \gamma, \eta, \sigma^2) = \sum_{i=1}^{n} E^*_{oi}(t, \gamma, \eta, \sigma^2)$. Therefore, a natural estimator for $\lambda_0(t)dt$ is given by

$$\lambda_0(t)dt = dN(t) / E^*_o(t, \gamma, \eta, \sigma^2) \qquad (7)$$

The parameters γ and η can be estimated by solving the following estimating equation, which is an analogue to score equations of partial likelihood:

$$\sum_{i=1}^{n} \int \{S_i(t, \gamma, \sigma^2), \mathrm{Z}_i(t)\}^T \{dN_i(t) - E^*_{oi}(t, \gamma, \eta, \sigma^2)\lambda_0(t)dt\} = 0$$

Substitution of $\hat{\lambda}_0(t)$ in (7) for $\lambda_0(t)$ yields the conditional score estimating equation

$$\sum_{i=1}^{n}\int[\{S_i(t,\gamma,\sigma^2),\mathrm{Z}_i(t)\}^T-\frac{E_i^*(t,\gamma,\eta,\sigma^2)}{E_{oi}^*(t,\gamma,\eta,\sigma^2)}]dN_i(t)=0, \qquad (8)$$

where

$$E_i^*(t,\gamma,\eta,\sigma^2)=\sum_{i=1}^{n}\{S_i(t,\gamma,\sigma^2),\mathrm{Z}_i(t)\}^T E_{oi}^*(t,\gamma,\eta,\sigma^2).$$

When $\sigma^2=0$, the estimating equation (8) reduces to the partial likelihood score equations.

The above derivation assumes that σ^2 is known.

According to the proposed work of (Tsiatis and Davidian 2001), a within-subject error was assumed to be normally distributed and mutually independent, violating the independent assumption may result to a bias estimate, in which the variance of the least square fit will have changed. From the assumption that the random effects b_i and the error within-subject ε_i for the i subject are normally distributed with $N(0,\sigma^2 I_{n_i})$ is given as $E(\hat{\mathrm{Z}}_i(t)_{ls})=Z_i(t)$ and $\mathrm{Var}(\hat{\mathrm{Z}}_i(t)_{ls})=\sigma^2\bar{t}^T(D_{i,t}^T,D_{i,t})^{-1}\bar{t}$, which will help in the statistical inference of conditional intensity for survival analysis when using the conditional score approach framework. However, in most cases σ^2 needs to be estimated from the data. A natural estimator of σ^2 is the pooled residual sum of squares from the least squares fit over subjects with at least three measurements:

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}I(n_i>2)SSE_i}{\sum_{i=1}^{n}I(n_i>2)(n_i-2)},$$

where SSE_i is the residual sum of squares for the n_i measurements on subject i. It can be shown that $\hat{\sigma}^2$ is a consistent estimator of σ^2 under the assumption that the measurement errors $\varepsilon_i(t_{ij})$ are i.i.d. $N(0,\sigma^2)$ random variables given random effects

b_i, covariates $Z_i(t)$, the biomarker history prior to t_{ij}, and $J_i(t_{ij}) = 1$. When $\hat{\sigma}^2$ is substituted for σ^2, solving the conditional score equation (8) yields $(\hat{\gamma}, \hat{\eta})$. It can be shown that $(\hat{\gamma}, \hat{\eta})$ is consistent and asymptotically normal. The usual sandwich approach may be used to derive their standard errors. Compared to the likelihood approaches, the conditional score method is relatively easier to implement. Although the derivation is discussed in a simple situation where a linear time trend is assumed for each subject, the method can be extended to general polynomial and spline trajectories.

If the mutually independent assumption has been violated with covariance Σ_i, the variance formula above for $\hat{Z}_i(t)_{ls}$ is no longer valid. The underlying variance is given as

$$\mathrm{Var}(\hat{Z}_i(t)_{ls}) = \bar{t}^T (D_{i,t}^T, D_{i,t})^{-1} D_{i,t}^T \Sigma_{i,t} D_{i,t} (D_{i,t}^T, D_{i,t})^{-1} \bar{t}, \qquad (9)$$

where the covariance matrix of the measurement errors for i^{th} subject at time t is $\Sigma_{i,t} = Cov(\varepsilon_{ij} : t_{ij} \le t)$ and assumes that the measurement errors are mutually independent, the above algorithm can be expressed as

$$\mathrm{Var}(\hat{Z}_i(t)_{ls}) = \sigma^2 \bar{t}^T (D_{i,t}^T, D_{i,t})^{-1} \bar{t}, \qquad (10)$$

where

$$\ddot{\sigma}^2 = E(\hat{\sigma}^2) \frac{\sum_{i=1}^{n} I(n_i > 2)(\mathrm{tr}(\Sigma_i) - \mathrm{tr}((D_i^T D_i)^{-1} D_i^T \Sigma_i D_i))}{\sum_{i=1}^{n} I(n_i > 2)(n_i - 2)},$$

3.4.4 Inference on Pooled Estimator

If I define $W_i = D_i b_i + \varepsilon_i$, where D_i is the $(n_i \times 1)$ matrix with ℓ^{th} column t_i^{ℓ} where $t_i = (t_{i1}, t_{i2}, .., t_{in_i})^T$ and $\ell = 0,1,..,n_i$ and $\hat{\sigma}^2$ is the estimating equation solution.

$$\sum_{i=1}^{n} I(n_i > 2)((W_i - D_i\hat{b}_i)^T (W_i - D_i\hat{b}_i) - (n_i - 2)\sigma^2) = 0,$$

where $\hat{b}_i = (D_i^T D_i)^{-1} D_i^T W_i = b_i + (D_i^T D_i)^{-1} D_i^T \varepsilon_i$. as

$(W_i - D_i\hat{b}_i)^T (W_i - D_i\hat{b}_i) = \varepsilon_i^T (I_{n_i} - D_i (D_i^T D_i)^{-1} D_i^T)\varepsilon_i$, gives the unbiased estimating equation for σ^2 and we note that

$$\begin{aligned}
&E(I(n_i > 2)((W_i - D_i\hat{b}_i)^T (W_i - D_i\hat{b}_i) - (n_i - 2)\sigma^2)) \\
&= E(E[I(n_i > 2)((W_i - D_i\hat{b}_i)^T (W_i - D_i\hat{b}_i) - (n_i - 2)\sigma^2) \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i]) \\
&= E(E[I(n_i > 2)(\varepsilon_i^T (I_{n_i} - D_i (D_i^T D_i)^{-1} D_i^T)\varepsilon_i - (n_i - 2)\sigma^2) \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i])
\end{aligned}$$

However, since we have $E(\varepsilon_i \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i) = 0$ and $\mathrm{Var}(\varepsilon_i \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i) = \sigma^2 I_{n_i}$,

$$\begin{aligned}
&E[I(n_i > 2)(\varepsilon_i^T (I_{n_i} - D_i (D_i^T D_i)^{-1} D_i^T)\varepsilon_i) \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i] \\
&= I(n_i > 2)(E[\varepsilon_i^T (I_{n_i} - D_i (D_i^T D_i)^{-1} D_i^T)\varepsilon_i) \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i]) \\
&= I(n_i > 2)(\mathrm{tr}\{(I_{n_i} - D_i (D_i^T D_i)^{-1} D_i^T) E(\varepsilon\varepsilon^T) \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i)\}) \\
&= I(n_i > 2)\,\mathrm{tr}\{(I_{n_i} - D_i (D_i^T D_i)^{-1} D_i^T)\sigma^2 I_{n_i}\} \\
&= I(n_i > 2)(n_i - tr\{D_i (D_i^T D_i)^{-1} D_i^T\})\sigma^2\} \\
&= I(n_i > 2)(n_i - 2)\sigma^2,
\end{aligned}$$

Since, $\hat{\sigma}^2$ is uniform estimate of σ^2 and shows that the internal conditional expectation is zero; hence, the estimating equation is unbiased.

However, if $E(\varepsilon_i \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i) = 0$ and $\mathrm{Var}(\varepsilon_i \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i) = \Sigma_i$, then,

$$\begin{aligned}
&E[I(n_i > 2)(\varepsilon_i^T (I_{n_i} - D_i (D_i^T D_i)^{-1} D_i^T)\varepsilon_i) \mid T_i, C_i, b_i, \mathrm{Z}_i, t_i, n_i] \\
&= I(n_i > 2)(tr\{\Sigma_i\} - tr\{(D_i^T D_i)^{-1} D_i^T \Sigma_i D_i\})
\end{aligned}$$

Hence, since $\Sigma_i \neq \sigma^2 I_{n_i}$ therefore $\hat{\sigma}^2$ is often a biased estimate and shows different interpretation. The influence of misspecification of covariance of error on the

inference of the survival parameters may lead to a larger bias for the conditional score estimator. This shows that when the mutually independent assumption is violated, it may result in the misuse of least square estimator variance of $Z_i(t)_{ls}$, which causes biased inference for survival parameters using the above conditional score approach. To rectify this erroneous variance estimator, one can substitute it by equation (9). In lieu of this, generalized least square (GLS) estimate of unknown underlying variance for survival parameters which is proposed in this thesis and assumes to have smaller variance than ordinary least square (OLS) estimate and referred to as generalized conditional score estimator.

3.4.5 Generalized Conditional Score Estimator

Let $\hat{Z}_i(t)_{gls}$ represents the GLS estimate of $Z_i(t)$ for measurement errors for i^{th} subject at time t, thus we have $Z_i(t)_{gls} = \bar{t}^T (D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t})^{-1} D_{i,t}^T \Sigma_{i,t}^{-1} W_{i,t}$. This GLS estimate $\hat{Z}_i(t)_{gls}$ is assumed to follow a multivariate normal distribution with mean $Z_i(t)$ and variance $\bar{t}^T (D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t})^{-1} \bar{t}$ and it is denoted by $\sigma^2_{\hat{X}_i(t)_{gls}}$. However, to show that the conditional score on $Z_i(t) = 1$ and the full expression for sufficient statistic is

$$S_i(t, \gamma, \Sigma_{i,t}) = \gamma \sigma^2_{\hat{X}_i(t)_{gls}} dN_i(t) + \hat{Z}_i(t)_{gls},$$

where $dN_i(t)$ is the counting process increase of $N_i(t) = I(\delta_i = 1, U_i \le t, t_{i2} \le t)$. Likewise, the conditional intensity function is expressed as

$$\lim_{dt \to 0} dt^{-1} P(dN_i(t) = 1 \mid S_i(t, \gamma, \Sigma_{i,t}), Y_i(t)$$

$$= \lambda_0(t) \exp\left(\gamma S_i(t, \gamma, \Sigma_{i,t}) - \frac{1}{2} \gamma^2 \sigma^2_{\hat{X}_i(t)_{gls}} + \eta^T Z_i \right) Y_i(t), \qquad (11)$$

The conditional intensity function of $dN(t) = \Sigma_{i-1}^n dN_i(t)$ is expressed as

$$\{S_i(t,\gamma,\Sigma_{i,t}), \mathrm{Z}_i, \bar{t}_i(t), Y_i(t), i=1,..,n\} = \lambda_0(t)E_0(t,\gamma,\eta,\Sigma_t)$$

where $\Sigma_t = \{\Sigma_{i,t} : i=1,..,n\}$ and $E_0(t,\gamma,\eta,\Sigma_t) = \sum_{i=1}^{n} E_{0i}(t,\gamma,\eta,\Sigma_{i,t})$;

$E_{0i}(t,\gamma,\eta,\Sigma_{i,t}) = \exp\left(\gamma S_i(t,\gamma,\Sigma_{i,t}) - \frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}} + \eta^T \mathrm{Z}_i\right)Y_i(t)$, indicates a fair estimate of $\lambda_0(t)dt$, which is expressed as $\lambda_0(t)dt = dN(t)/E_0(t,\gamma,\eta,\Sigma_t)$.

Let $\Upsilon_i(t,\psi_k)$ represents a deterministic function (predictably-locally bounded process) and I denote $\psi_k = (\gamma,\eta)^T$, therefore, intensity function of i^{th} subject is denoted as $\lambda_i(t,\psi_k)$ to underline the parameter function of ψ_k. An estimate of ψ_k can be defined as a solution to the equation using the M-estimator proposed by Andersen et al. (1993, p433) as

$$\begin{aligned} V_1(\psi_k) &= \sum_{i=1}^{n}\int \Upsilon_i(t,\psi_k)\{dN_i(t) - \lambda_i(t,\psi_k)dt\} \\ &= \sum_{i=1}^{n}\int \Upsilon_i(t,\psi_k)\{dN_i(t) - E_{0i}(t,\gamma,\eta,\Sigma_{i,t})\lambda_i(t)dt\} = 0, \end{aligned} \qquad (12)$$

When I substitute $\hat{\lambda}_0(t)dt$ for $\lambda_0(t)dt$ the equation (12) may be expressed as

$$V_1(\psi_k) = \sum_{i=1}^{n}\int \Upsilon_i(t,\psi_k)\left\{dN_i(t) - \frac{E_{0i}(t,\gamma,\eta,\Sigma_{i,t})}{E_0(t,\gamma,\eta,\Sigma_t)}dN(t)\right\} = 0$$

and define $E_1(t,\gamma,\eta,\Sigma_t) = \Sigma_{i=1}^{n}E_{1i}(t,\gamma,\eta,\Sigma_{i,t})$ with $E_{1i}(t,\gamma,\eta,\Sigma_{i,t}) = \Upsilon_i(t,\psi_k)E_{0i}(t,\gamma,\eta,\Sigma_{i,t})$, subsequently, the estimating equation can be given as

$$V_1(\psi_k) = \sum_{i=1}^{n}\int\left\{\Upsilon_i(t,\psi_k) - \frac{E_{0i}(t,\gamma,\eta,\Sigma_{i,t})}{E_0(t,\gamma,\eta,\Sigma_t)}\right\}dN_i(t) = 0. \qquad (13)$$

From what was proposed by Andersen et al. (1993, p433) on M-estimator and Maximum likelihood estimator, we, therefore, aimed for two classes of $\Upsilon_i(t,\psi_k)$. The first class is $\Upsilon_i(t,\psi_k) = (S_i(t,\gamma,\Sigma_{i,t}), \mathrm{Z}_i)^T$, which is referred to as ME estimation and the

second class is $\Upsilon_i(t,\psi_k) = (S_i(t,\gamma,\Sigma_{i,t}) - \gamma\sigma^2_{\hat{X}_i(t)_{gls}}, \mathrm{Z}_i)^T$, referred to as MLE estimation. The main interest is to estimate a statistical inference for $\psi_k = (\gamma,\eta)^T$, which is a straightforward application to the proposed generalized conditional score framework. Newton-Raphson numerical methods are expected methods used to obtain the solution to the equation (13), which in fact may have various roots when a primitive estimator is used as a starting value and may be a pragmatic approach for locating consistent root. We used simulation studies to ascertain the proposed processes.

3.4.6 Inference on Conditional Intensity and Sufficient Statistics

A detailed inference for determining and verifying the sufficient statistics and conditional intensity function process as proposed in our models are given as

$$\begin{cases} W_i(t_{ij}) = b_{io} + b_{i1}t_{ij} + b_{i2}t_{ij}^2 + .. + b_{iq}t_{ij}^q + \varepsilon_{ij}, \\ \lambda_i(t) = \lambda_0(t)\exp\left(\gamma X_i(t) + \eta^T \mathrm{Z}_i\right) \quad i = 1,2,..,n; \quad j = 1,2,..,m_i, \end{cases}$$

where $b_i = (b_{i0}, b_{i1}, .., b_{iq})^T$, $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i}, .., \varepsilon_{m_i i})^T \sim N(0,\Sigma_i)$ and GLS estimator of $X_i(t)$ is expressed as

$$X_i(t)_{gls} = \bar{t}^T (D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t})^{-1} D_{i,t}^T \Sigma_{i,t}^{-1} W_{i,t},$$

The mean and variance of the generalized least square estimator are given as

$$\begin{aligned} E(\hat{X}_i(t)_{gls} \mid b_i) &= \bar{t}^T (D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t})^{-1} D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t} b_i = \bar{t}^T b_i, \\ \mathrm{Var}(\hat{X}_i(t)_{gls} \mid b_i) &= \bar{t}^T (D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t})^{-1} D_{i,t}^T \Sigma_{i,t}^{-1} \Sigma_{i,t} \Sigma_{i,t}^{-1} D_{i,t} (D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t})^{-1} \bar{t} \\ &= \bar{t}^T (D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t})^{-1} \bar{t} = \sigma^2_{\hat{X}_i(t)_{gls}} \end{aligned}$$

Let assume that $\Sigma_{i,t}^{-1}$ or $\Sigma_{i,t}$ is known, hence the conditional estimator of $\hat{X}_i(t)_{gls}$ on $\{b_i, \bar{t}_i(t), Y_i(t) = 1, Z_i\}$ is normally distributed with mean of $X_i(t) = \Sigma_{t=0}^q b_{it} t^t$ and variance $\bar{t}^T (D_{i,t}^T \Sigma_{i,t}^{-1} D_{i,t})^{-1} \bar{t}$. Therefore, we have $\{dN_i(t) = \kappa, \hat{X}_i(t)_{gls} = x \mid Y_i(t) = 1, b_i, W_i(t_{ij})s$ taken up to include time t at $\bar{t}_i(t)$ is

$$= P\left(dN_i(t)=\kappa, \hat{X}_i(t)_{gls}=x, b_i, Z_i, \bar{t_i}(t)\right) P\left(\hat{X}_i(t)_{gls}=x \mid Y_i(t)=1, b_i, Z_i, \bar{t_i}(t)\right)$$

$$= \left[\lambda_0(t)dt \exp\left(\gamma X_i(t)+\eta^T \mathrm{Z}_i\right)\right]^{\kappa} \left[1-\lambda_0(t)dt \exp\left(\gamma X_i(t)+\eta^T \mathrm{Z}_i\right)\right]^{1-\kappa} \frac{\exp\left(\frac{(x-X_i(t))^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}\right)}{\left(2\sigma^2_{\hat{X}_i(t)_{gls}}\right)^{1/2}}.$$

The likelihood for conditional intensity of $\{dN_i(t)=\kappa, \hat{X}_i(t)_{gls}\}$ given $\{Y_i(t)=1, b_i, Z_i, \bar{t_i}(t)\}$ to the order of dt is expressed as

$$\left[\lambda_0(t)dt \exp\left(\gamma X_i(t)+\eta^T \mathrm{Z}_i\right)\right]^{dN_i(t)} \frac{1}{\left(2\sigma^2_{\hat{X}_i(t)_{gls}}\right)^{1/2}} \exp\left(-\frac{\hat{X}_i(t)_{gls}-2\hat{X}_i(t)_{gls}X_i(t)+X_i(t)^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}\right)$$

$$= \exp\left(X_i(t)\left(\gamma dN_i(t)+\frac{\hat{X}_i(t)_{gls}}{\sigma^2_{\hat{X}_i(t)_{gls}}}\right)\right) \frac{\left[\lambda_0(t)dt \exp\left(\eta^T \mathrm{Z}_i\right)\right]^{dN_i(t)}}{2\pi\sigma^2_{\hat{X}_i(t)_{gls}}} \exp\left(-\frac{\hat{X}_i(t)^2{}_{gls}+X_i(t)^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}\right). \quad (3.14)$$

When I have conditional intensity on $Y_i(t)=1$, complete sufficient statistic (S_i) for b_i becomes

$$S_i(t,\gamma,\Sigma_{i,t}) = \gamma\sigma^2_{\hat{X}_i(t)_{gls}} dN_i(t)+\hat{X}_i(t)_{gls}$$

However, equation (14) becomes

$$\frac{\left[\lambda_0(t)dt \exp\left(\eta^T \mathrm{Z}_i\right)\right]^{dN_i(t)}}{\left(2\sigma^2_{\hat{X}_i(t)_{gls}}\right)^{1/2}} \exp\left(-\frac{S_i^2-2S_i\gamma\sigma^2_{\hat{X}_i(t)_{gls}}dN_i(t)+\gamma^2\sigma^4_{\hat{X}_{i,t}}dN_i(t)+X_i(t)^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}+\frac{X_i(t)S_i}{\sigma^2_{\hat{X}_i(t)_{gls}}}\right)$$

$$= \frac{\left[\lambda_0(t)dt \exp\left(\eta^T \mathrm{Z}_i\right)\right]^{dN_i(t)}}{\left(2\pi\sigma^2_{\hat{X}_i(t)_{gls}}\right)^{1/2}} \exp\left(-\frac{S_i^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}+S_i\gamma dN_i(t)-\frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}}dN_i(t)+\frac{2X_i(t)S_i-X_i(t)^2}{\sigma^2_{\hat{X}_i(t)_{gls}}}\right)$$

$$= \left[\lambda_0(t)dt \exp\left(\eta^T \mathrm{Z}_i\right)\right]^{dN_i(t)} \exp\left(-\frac{S_i^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}+S_i\gamma dN_i(t)-\frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}}dN_i(t)\right)\vartheta_0$$

$$= P\left(dN_i(t), \hat{X}_i(t)_{gls} \mid Y_i(t)=1, b_i, Z_i, \bar{t_i}(t)\right),$$

where $\vartheta_0 = \exp\left(\frac{2X_i(t)S_i-X_i(t)^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}\right)\left(2\pi\sigma^2_{\hat{X}_i(t)_{gls}}\right)^{-1/2}$. Therefore, we have

$$P\left(dN_i(t)=1 \mid S_i(t,\gamma,\Sigma_{i,t}), Z_i, \bar{t}_i(t), Y_i(t)=1\right)$$
$$= \frac{P\left(dN_i(t)=1, S_i \mid Z_i, \bar{t}_i(t), Y_i(t)=1\right)}{P\left(dN_i(t)=1, S_i \mid Z_i, \bar{t}_i(t), Y_i(t)=1\right)+P\left(dN_i(t)=0, S_i \mid Z_i, \bar{t}_i(t), Y_i(t)=1\right)}.$$

The numerator (n_u) and denominator (d_e) for the above equation is expressed as

$$n_u = \lambda_0(t)dt\exp\left(\eta^T Z_i\right)\exp\left(-\frac{S_i^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}+S_i\gamma-\frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}}\right)\int \vartheta_0 P\left(b_i \mid ...\right)db_i,$$

$$d_e = n_u + \exp\left(-\frac{S_i^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}\right)\int \vartheta_0 P\left(b_i \mid ...\right)db_i,$$

The density function for random effects b_i is denoted as the probability of $(b_i \mid ...)$. Hence, I have

$$P\left(dN_i(t)=1 \mid S_i(t,\gamma,\Sigma_{i,t}), Z_i, \bar{t}_i(t), Y_i(t)=1\right)$$
$$\approx \frac{\lambda_0(t)dt\exp\left(\eta^T Z_i\right)\exp\left(-\frac{S_i^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}+S_i\gamma-\frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}}\right)}{\lambda_0(t)dt\exp\left(\eta^T Z_i\right)\exp\left(-\frac{S_i^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}+S_i\gamma-\frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}}\right)+\exp\left(-\frac{S_i^2}{2\sigma^2_{\hat{X}_i(t)_{gls}}}\right)}$$
$$\approx \lambda_0(t)dt\exp\left(\eta^T Z_i\right)\exp\left(S_i\gamma-\frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}}\right)=\lambda_0(t)\exp\left(S_i\gamma-\frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}}+\left(\eta^T Z_i\right)\right)dt.$$

Therefore, the conditional intensity function is given as

$$\lambda_i\left(t \mid S_i(t,\gamma,\Sigma_{i,t})\right) \approx \lambda_0(t)\exp\left(S_i\gamma(t,\gamma,\Sigma_{i,t})-\frac{1}{2}\gamma^2\sigma^2_{\hat{X}_i(t)_{gls}}+\left(\eta^T Z_i\right)\right)Y_i(t).$$

3.4.7 Covariance Function Estimation

It is a well-known fact that in the proposed generalized conditional score method, it requires the full understanding of within-subject covariance from the survival estimating equation (13), which is a function of covariance matrices $\Sigma_{i,t}s$ and are usually unknown with a practicable inference. Let assume that ψ_c represents the parameters within-subject covariance and $V_2(\psi_2)=0$ denotes the corresponding

estimating equation, the inference of the statistical estimation for within-subject covariance Σ_i can exclude the survival parameters $\psi_k = (\gamma, \eta)^T$, which is based on the information on the measurements derived from the longitudinal process. However, if we denote $\psi = (\gamma, \eta, \psi_c)^T$, then parameters estimating equations ψ can be expressed by

$$V(\psi) = \begin{pmatrix} V_1(\psi_k \psi_c) \\ V_2(\psi_c) \end{pmatrix} = 0 \qquad (15)$$

where $V_1(\psi_k \psi_c)$ is used to relax the required covariance $\Sigma_{i,t} s$ understanding in equation (13). I obtain the estimator for $\hat{\psi}_c$ through two ways, firstly, it is obtained according to equation $V_2(\psi_c) = 0$ and secondly, by replacing the parameters of the covariance by $\hat{\psi}_c$ in $V_1(\psi_k \psi_c) = 0$ for the survival inference. A method of modified Cholesky Decomposition approach is used to estimate the within-subject covariance for the longitudinal outcomes modelled with linear mixed effects $W_i(t_i) = b_{io} + b_{i1} t_i + b_{i2} t_i^2 + .. + b_{iq} t_i^q + \varepsilon_i(t)$ model and assumed to be from multivariate normal distribution with $N(0, \Sigma_i)$.

3.4.8 Cholesky Decomposition Modification

I adopted a modified Cholesky decomposition to model the positive definite covariance matrix Σ_i in order to equally obtain $L_i \Sigma_i L^T = D_i$ or $\Sigma_i^{-1} = L_i^T D_i^{-T} L_i$, where L_i a lower triangular matrix with diagonal entries equal to 1 and D_i is a unit diagonal matrix. These lower diagonal entries, L_i are said to be negatives of the autoregressive coefficients (θ_{ijl}) from the autoregressive model $W_{ij} = X_{ij} + \Sigma_{l=1}^{j-1} \theta_{ijl} (W_{il} - X_{il}) + \varepsilon_{ij}$, where X_i is the underlying values from longitudinal outcomes W_i and D_i (diagonal entries) are the variance prediction denoted by

$\sigma^2_{ij} = \text{var}(\varepsilon_{ij} \mid b_i)$. I also denote the model for unconstrained coefficients θ_{ijl} and $\log\sigma^2_{ij}$ as

$$\theta_{ijl} = \alpha_{ijl}^T \omega, \ \log\sigma^2_{ij} = \rho_{ij}^T \varpi.$$

Hence, the within-subject covariance parameter Σ_i is denoted as $\psi_c = (\omega, \varpi)^T$, where α_{ijl} and ρ_{ij} are vectors of covariates, which may comprise the baseline covariates and time based polynomial terms as well as the interaction between them. Similarly, random effects is treated as a nuisance parameter when using the generalized conditional score approach in order of sample size and can be possibly, a high dimension, thus MLE of ψ_c may result to a biased estimation. Because of this, we tend to use restricted log-likelihood for ψ_c express as

$$L_{nq}(\psi_c) = -\frac{1}{2}\sum_{i=1}^{n}\log|\Sigma_i(\psi_c)| - \frac{1}{2}\sum_{i=1}^{n}\log|D_i^T\Sigma_i^{-1}(\psi_c)D_i$$
$$-\frac{1}{2}\sum_{i=1}^{n}\left(\left(W_i - D_i\bar{b}_i\right)^T\Sigma_i^{-1}(\psi_c)\left(W_i - D_i\bar{b}_i\right)\right), \qquad (16)$$

Where $\bar{b}_i = \left(D_i^T\Sigma_i^{-1}(\psi_c)D_i\right)^{-1}D_i^T\Sigma_i^{-1}(\psi_c)W_i$, $\Sigma_i = Cov(\varepsilon_i)$ and $W_i = (W_{1i},..,W_{im_i})^T$.

I assume that $W_{ij} = (W_{1i},..,W_{im_i})^T$, which denotes j^{th} measurement for i subjects and X_i represents covariates measurements for i subjects. A modified Cholesky decomposition method is used to capture the within-subject covariance for a positive definite and symmetric matrix, assumed that observed data from different subjects are independent. The modified characteristics of log-likelihood function after multiplying by -2 *w.r.t.* ψ_c.

$$\ell_{nq}(\psi) = \sum_{i=1}^{n}\log|\Sigma_i| + \sum_{i=1}^{n}\log\left|X_i^T\Sigma_i^{-1}X_i\right|$$
$$+\sum_{i=1}^{n}\left(W_i^T\Sigma_i^{-1}W_i - W_i^T\Sigma_i^{-1}X_i\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^T\Sigma_i^{-1}W_i\right)$$

Then the derivatives are as follows

$$\frac{\partial \ell_{nq}}{\partial \omega_s} = 2\sum_{i=1}^{n} \text{Trace}\left(\left(X_i^T \Sigma_i^{-1} X_i \right)^{-1} X_i^T \frac{\partial L_i^T}{\partial \omega_s} H_i^{-1} L_i X_i \right) + 2\sum_{i=1}^{n} W_i^T D_i^T \frac{\partial L_i^T}{\partial \omega_s} H_i^{-1} L_i D_i W_i,$$

$$\frac{\partial \ell_{nq}}{\partial \varpi_s} = \sum_{i=1}^{n} \text{Trace}\left(R_{i\ell} G_i \right) - \sum_{i=1}^{n} W_i^T D_i^T L_i^T H_i^{-1} R_{i\ell} L_i D_i W_i,$$

where Trace (.) is the trace function of a matrix, $R_{i\ell} = diag(r_{1i\ell}, .., r_{i\ell m_i})$, $D_i = I_{m_i} - X_i \left(X_i^T \Sigma_i^{-1} X_i \right)^{-1} X_i^T \Sigma_i^{-1}$ and $G_i = I_{m_i} - L_i X_i \left(X_i^T \Sigma_i^{-1} X_i \right)^{-1} X_i^T L_i^T H_i^{-1} = L_i D_i L_i^{-1}$, I_{m_i} is the identity matrix.

Also we have $L_i X_i = 0$ and $G_i D_i X_i = 0$ and the properties are

$$\begin{aligned} E\left(W_i W_i^T \middle| \alpha_i \right) &= Var\left(W_i \middle| \alpha_i \right) + E\left(W_i \middle| \alpha_i \right) E\left(W_i \middle| \alpha_i \right)^T \\ &= \Sigma_i + X_i \alpha_i \alpha_i^T X_i^T. \end{aligned}$$

I denote $\frac{\partial \ell_{nq_i}}{2\partial \omega_s} = \Psi_{i\omega_s}$ and $\frac{\partial \ell_{nq_i}}{2\partial \varpi_s} = \Psi_{i\varpi_s}$, then we have the following

$$\begin{aligned} E\left(\Psi_{i\omega_s} \middle| \alpha_i \right) &= \text{Trace}\left(\left(X_i^T \Sigma_i^{-1} X_i \right)^{-1} X_i^T \frac{\partial L_i^T}{\partial \omega_s} H_i^{-1} L_i X_i \right) + \text{Trace}\left(D_i^T \frac{\partial L_i^T}{\partial \omega_s} H_i^{-1} L_i D_i E\left(W_i W_i^T \middle| \alpha_i \right) \right) \\ &= \text{Trace}\left(\left(X_i^T \Sigma_i^{-1} X_i \right)^{-1} X_i^T \frac{\partial L_i^T}{\partial \omega_s} H_i^{-1} L_i X_i \right) + \text{Trace}\left(D_i^T \frac{\partial L_i^T}{\partial \omega_s} H_i^{-1} L_i D_i \Sigma_i \right) \\ &= \text{Trace}\left(\frac{\partial L_i^T}{\partial \omega_s} H_i^{-1} L_i \Sigma_i \right) \\ &= \text{Trace}\left(\frac{\partial L_i^T}{\partial \omega_s} H_i^{-1} \right) = 0, \end{aligned}$$

$$\begin{aligned} 2E\left(\Psi_{i\varpi_s} \middle| \alpha_i \right) &= \text{Trace}\left(R_{i\ell} G_i \right) - \text{Trace}\left(D_i^T L_i^T H_i^{-1} R_{i\ell} L_i D_i E\left(W_i^T W_i \middle| \alpha_i \right) \right) \\ &= \text{Trace}\left(R_{i\ell} G_i \right) - \text{Trace}\left(D_i^T L_i^T H_i^{-1} R_{i\ell} L_i D_i \Sigma_i \right) \\ &= \text{Trace}\left(R_{i\ell} \right) - \text{Trace}\left(L_i^T H_i^{-1} R_{i\ell} L_i \Sigma_i \right) \\ &= \text{Trace}\left(R_{i\ell} \right) - \text{Trace}\left(R_{i\ell} L_i \Sigma_i L_i^T H_i^{-1} \right) \\ &= \text{Trace}\left(R_{i\ell} \right) - \text{Trace}\left(R_{i\ell} \right) = 0 \end{aligned}$$

$$\therefore \quad E\left(\Psi_{i\omega_s} \right) = E\left(\Psi_{i\omega_s} \middle| \alpha_i \right) = 0$$

$$E\left(\Psi_{i\varpi_s} \right) = E\left(\Psi_{i\varpi_s} \middle| \alpha_i \right) = 0$$

For the second derivatives, I denote $\Psi_{i\omega_s}$ and $\Psi_{i\varpi_s}$ as the following

$$\Psi_{\omega_s} = \frac{1}{n}\sum_{i=1}^{n}\left[\text{Trace}\left(\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^T\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}L_iX_i\right)+W_i^TD_i^T\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}L_iD_iW_i\right],$$

$$\Psi_{\varpi_s} = \frac{1}{2n}\sum_{i=1}^{n}\left[\text{Trace}\left(R_{i\ell}G_i\right)-W_i^TD_i^TL_i^TH_i^{-1}R_{i\ell}L_iD_iW_i\right].$$

Derivative with respect to ω_s and ϖ_s

$$\frac{\partial\Psi_{\omega_s}}{\partial\omega_{s'}} = \frac{1}{n}\sum_{i=1}^{n}\left[\text{Trace}\left(\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^T\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}\frac{\partial L_i}{\partial\omega_{s'}}X_i\right)\right.$$
$$-\text{Trace}\left(\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^T\left(\frac{\partial L_i^T}{\partial\omega_{s'}}H_i^{-1}L_i+H_i^{-1}L_i^T\frac{\partial L_i}{\partial\omega_{s'}}\right)X_i\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^T\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}L_iX_i\right)$$
$$\left.+W_i^T\frac{\partial D_i^T}{\partial\omega_{s'}}\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}L_iD_iW_i+W_i^TD_i^T\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}\left(\frac{\partial L_i}{\partial\omega_{s'}}D_i+L_i\frac{\partial D_i}{\partial\omega_{s'}}\right)W_i\right],$$

$$\frac{\partial\Psi_{\omega_s}}{\partial\varpi_\ell} = \frac{\partial\Psi_{\varpi_\ell}}{\partial\varpi_\ell}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left[W_i^TD_i^T\left(L_i^TH_i^{-1}\frac{\partial L_i}{\partial\omega_s}+\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}L_i\right)X_i\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^TL_i^TH_i^{-1}R_{i\ell}L_iD_iW_i\right.$$
$$\left.-\text{Trace}\left(\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^T\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}G_iR_{i\ell}L_iX_i\right)-H_i^{-1}D_i^T\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}R_{i\ell}L_iD_iW_i\right],$$

$$\frac{\partial\Psi_{\varpi_\ell}}{\partial\varpi_{\ell'}} = \frac{1}{2n}\sum_{i=1}^{n}\left[\text{Trace}\left(R_{i\ell}L_iX_i\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^TL_i^TH_i^{-1}R_{i\ell}G_i\right)\right.$$
$$\left.-2W_i^T\frac{\partial D_i^T}{\partial\omega_s}L_i^TH_i^{-1}R_{i\ell}L_iD_iW_i+W_i^TD_i^TL_i^TH_i^{-1}R_{i\ell}R_{i\ell}L_iD_iW_i\right],$$

Thereafter, I have

$$\frac{\partial D_i}{\partial\omega_{s'}} = -X_i\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^T\left(\frac{\partial L_i^T}{\partial\omega_s}H_i^{-1}L_i+H_i^{-1}L_i^T\frac{\partial L_i}{\partial\omega_{s'}}\right)L_i,$$

$$\frac{\partial D_i}{\partial\varpi_{\ell'}} = X_i\left(X_i^T\Sigma_i^{-1}X_i\right)^{-1}X_i^TH_i^{-1}L_iR_{i\ell}L_iD_i.$$

In order for the restricted log-likelihood to serve as a linear regression model, restricted log-likelihood was maximized with respect to ψ_c to obtain a restricted maximum likelihood estimate (REML) and it requires numerical approaches to estimate. However, one can use Newton-Raphson algorithm for the numerical

methods but complexity of restricted log-likelihood function as well as Choleskey regression model decomposition and the high-dimensional within-subject random effects, our try-out experiment indicates that Newton-Raphson algorithm performances are good enough to capture the $\psi_c = (\omega, \varpi)^T$ in a restricted maximum likelihood estimate (REML). Alternatively, Nelder-Mead algorithm was used to generate a restricted maximum likelihood estimate (REML) for $\hat{\psi}_c$ and the standard deviation of $\hat{\psi}_c$ was calculated by the sandwich rule. I let the first derivative to be equal to zero in order to show if they are unbiased estimator under the assumption that different subjects from the observed data are mutually independent. The REML is said to be consistent with asymptotic normal distribution under some certain conditions, i.e. if $\Psi_\omega(\psi) = (\Psi_{\omega_1}, .., \Psi_{\omega_d})^T$ and $\Psi_\varpi(\psi) = (\Psi_{\varpi_1}, .., \Psi_{\varpi_g})^T$ have the unbiased estimating equations of

$$V(\psi) = \begin{pmatrix} \Psi_\omega(\psi) \\ \Psi_\varpi(\psi) \end{pmatrix} = 0$$

The sandwich rule may be used to compute the standard deviation of the estimators. I simulated data to demonstrate such properties and the results are shown in chapter four of this thesis.

3.5 Joint Model with Cumulative Association Structure Effects

There have been several attempts to extend joint analysis for a single longitudinal outcome to the situation where multiple outcomes are simultaneously recorded with the event times. A key issue in modelling multivariate longitudinal data and their association with event times is in formulating the joint evolution of the multiple endpoints. In a standard joint models formulation, the underlying current subject-specific value of a biomarker is assumed to associate with the risk of an event

happening at the same time t with an association parameter α, which may not be enough to describe the association structure. In l light of this, different studies have proposed an alternative association structure between the longitudinal and time-to-event outcomes (Brown et al. 2005; Rizopoulos et al. 2014; Ye et al. 2008). , It has been noted that the model for longitudinal outcomes needs to take into account two associated outcomes. First, the risk of an event that depends on the longitudinal profile slope, and second, the risk of an event to depend on the cumulative effect. These data associations are usually characterized by latent variables, either continuous (random effects) or discrete (latent classes), which are also used to link the longitudinal outcomes and survival data.

In the setting of standard Cox model, many studies have proposed the use of weighted cumulative models (Breslow et al. 1983; Thomas 1988). In another study, a parametric time-dependent weight function was included as an extension to Cox model (Abrahamowicz et al. 2006). In many studies, joint model of survival-longitudinal data is often evaluated by either linking them, putting the value of longitudinal outcomes into the survival model or by using the shared random-effects models to examine scientifically the association between the survival time and the underlying unobserved longitudinal processes (Henderson et al. 2000; Tsiatis and Davidian 2004a).

In most cases, statistical inference is required to implement the survival model, when the complete knowledge of the longitudinal process is not completely available. In recent researches, it is observed that past and present longitudinal processes may informatively affect the survival model. Rate of change of the disease would be important if there would be a change in the risk factor. The analysis should not only be based on disease survival rate to the current risk factors at even time but also

pertains to the past period. If the level of longitudinal process on the survival risk is not accounted for by the past information in the analysis of the longitudinal data, the analysis may leads to under estimation or bias estimate.

3.5.1 The Proposed Methods for the Estimation Challenges

This chapter considers extensions of the joint models to multidimensional longitudinal and/or survival data. The approaches covered here are useful for studies that collect information on more than one longitudinal outcome or event on each participant (repeated measurements of multi-covariates patient-reported outcomes were used to predict patient survival). In the course of this study, we proposed to introduce the cumulative information longitudinal process into the survival model instead of using the standard default present value models to examine the association between the event time and longitudinal process. A weighted moment integration of trajectory information was used to observe the impact of longitudinal process on the survival time process and it is specified as

$$\int_0^t \phi(t-v) n_i(v)\, dv \qquad (17)$$

where $n_i(.)$ is the longitudinal process trajectory, v represents the time point of past longitudinal process and $\phi(.)$ is the weighted function. However, the longitudinal profile levels at different past times may have different influence on the present survival risks. The weight of underlying longitudinal process for the present levels may be constant over time but might get weaker as time lag increases. The longitudinal process may have positive effect or negative effect on the survival risk.

Integration has been a method to solve the problem arising from the statistical inference in the survival model with time-varying coefficient $\phi(t)$ for $n_i(t)$. Many well-established statistical inference approaches have proposed survival model such as piecewise function for coefficient parameterization, linear polynomial method, B-splines methods to describe the underlying time-varying coefficient. However, for this study, a double exponential function is proposed for the distribution of weighted curve $\phi(t-v)$. The weighted function through an exponential function can be given as

$$\begin{aligned} \phi(t-v) &= be^{-a(t-v)}, \qquad 0 \le v \le t, a > 0, \text{ and } b \in R, \\ \phi(v) &= be^{-a(v)}, \qquad 0 \le v \le t, a > 0, \text{ and } b \in R, \end{aligned} \tag{17.1}$$

The reason for choosing exponential function for the effects of weighted function $\phi(.)$ is that, the past effects of longitudinal process on the current risk of survival rate may have the tendency to be reducing as time lag increases.

However, the joint model stated may not be appropriate for more complex association between the outcomes of longitudinal and time-to-event processes (Sylvestre and Abrahamowicz 2009). Therefore, a specific alternative model (a truncated weighted skewed-normal) that accounts for a cumulative effect for longitudinal outcomes that includes the integral of the longitudinal trajectory, was proposed (Brown 2009; Rizopoulos 2012), and this may increase the statistical power analysis (Abrahamowicz et al. 2006). This extension of joint model of the longitudinal outcome, which relates the survival outcomes with a summary area under the longitudinal profiles of the whole history of the markers as:

$$h_i(t) = h_{0k}(t)\exp\left\{\alpha_k{}^T w_{ik}(t) + \gamma\int_0^t f_{ip}(s,b_i)ds\right\} \qquad (17.2)$$

where $\int_0^t f_{ip}(s,b_i)ds$ is the summary area under the longitudinal profile bounded by $f_{ip}(s,b_i)$.

3.5.2 The Model formulation

The framework for a single longitudinal process, in which linear mixed-effects models was proposed with polynomial function of time for longitudinal responses. This framework is extended in this study to statistical inferences on multivariate generalized linear mixed-effects model to describe the distribution of different types of longitudinal processes. I first looked at a class of model of random effects variables to model the association between longitudinal and survival data. Suppose there are K response variables, for $k = 1,..,K$, let Y_{ik} be a $(n_{ik} \times 1)$ vector of repeated responses for the $K\text{th}$ outcome collected on i subject at any time point $t_{ijk}, i = 1,..,n, j = 1,...,n_{ik}$. However, it is possible that the measurement t_{ijk} differs from subject to subject. In this proposed model, the assumption is that given random effects b_{ik}, the distribution of Y_{ik} is a member of exponential families whose linear predictor is given as

$$h_k\left\{E\left(Y_{ik}(t) \mid b_{ik}\right)\right\} = f_{ik}(t) \qquad (18)$$

where $h_k(.)$ is a known monotonic link function, $Y_{ik}(t)$ the $k\text{th}$ longitudinal response measured at time t on subject i, and $f_{ik}(t)$ a time-dependent function to describe

the true, underlying longitudinal profile for response k. Within the mixed effects model framework, a natural choice for $Y_{ik}(t)$ is

$$f_{ik}(t) = X_{ik}^{T}(t)\beta_k + Z_{ik}(t)^{T} b_{ik}, \qquad (19)$$

where β_k is a vector of fixed effects, and $X_{ik}(t)$ and $Z_{ik}(t)$ are covariates measured at time t. To allow flexible trajectories in the longitudinal responses, the components in $X_{ik}(t)$ and $Z_{ik}(t)$ can be further classified into time-dependent covariates, and the latter may contain spline functions to characterize the time trend.

The standard Cox proportional hazards model is proposed for survival data to investigate the association of event time with covariates and a cumulative condition to account for the trajectory longitudinal process information is assumed as

$$\int_{t_l}^{t} \phi(t-v) n_i(v)\, dv = \int_{t_l}^{t_0(t)} \phi(v)\, n_i(t-v) dv \qquad (20)$$

where $t_l \in \{0, t-c, \max(0, t-c)\}$ represents the assumption that the cumulative information is computed if the baseline or last period of time $t-c$, which not earlier than the baseline subsequently is $t_0(t) \in \{t, c, \min(t, c)\}$. Therefore, with the assumption of generalized linear mixed effects (19) for multivariate longitudinal responses, equation (20) describes the longitudinal responses cumulative information, which, can be expanded as

$$
\begin{aligned}
&\int_0^{t_0(t)} \phi(v)\, n_i(t-v)dv = \int_0^{t_0(t)} \phi(v)\left(\sum_{\ell=0}^{p} \alpha_{\ell i}(t-v)^{\ell}\right)dv + X_i^T\beta\int_0^{t_0(t)}\phi(v)dv + +Z_i^T b\int_0^{t_0(t)}\phi(v)dv \\
&= \sum_{\ell=0}^{p}\alpha_{\ell i}\left(\int_0^{t_0(t)}\phi(v)(t-v)^{\ell}\,dv\right) + X_i^T\beta\int_0^{t_0(t)}\phi(v)dv + Z_i^T b\int_0^{t_0(t)}\phi(v)dv \\
&= \sum_{\ell=0}^{p}\alpha_{\ell i}\left(\int_0^{t_0(t)}\phi(v)\left(\sum_{q=1}^{\ell} C_\ell^q t^{\ell-q}(-1)^q v^q\right)dv\right) + X_i^T\beta\int_0^{t_0(t)}\phi(v)dv + Z_i^T b\int_0^{t_0(t)}\phi(v)dv \\
&= \sum_{\ell=0}^{p}\alpha_{\ell i}\left(\sum_{q=1}^{\ell} C_\ell^q t^{\ell-q}(-1)^q \int_0^{t_0(t)}\phi(v)v^q dv\right) + X_i^T\beta\int_0^{t_0(t)}\phi(v)dv + Z_i^T b\int_0^{t_0(t)}\phi(v)dv \\
&= \sum_{q=0}^{p}\sum_{\ell=q}^{p}\alpha_{\ell i} t^{\ell-q}(-1)^q C_\ell^q\int_0^{t_0(t)}\phi(v)v^q dv + X_i^T\beta\int_0^{t_0(t)}\phi(v)dv + Z_i^T b\int_0^{t_0(t)}\phi(v)dv \\
&= m_i(t)\int_0^{t_0(t)}\phi(v)dv + \sum_{q=1}^{p}\frac{(-1)^q}{q!}m_i^{(q)}(t)\int_0^{t_0(t)}\phi(v)v^q dv
\end{aligned}
\tag{21}
$$

where $(t-v)^{\ell} = \Sigma_{q=0}^{\ell} C_\ell^q t^{\ell-q}(-1)^q v^q$, $C_\ell^q = \frac{\ell!}{(\ell-q)!q!}$ and $m_i^{(q)}(t)$ is the derivative of $m_i(t)$ w.r.t. time and $m_i(t) = X_i^T\beta + Z_i^T b$. From the equation (21), by considering the cumulative information of longitudinal process, the coefficient of the current value becomes $m_i(t) = \int_0^{t_0(t)}\phi(v)dv$ in the Cox proportional hazard model. However, the coefficient of the current value of longitudinal regression process in the Cox proportional hazard model and the cumulative term from equation (20) is subsequently as

$$
\begin{aligned}
&\int_0^{t_0(t)}\phi(u)\, m_i(t-u)du \\
&= \int_0^{t_0(t)}\alpha e^{-au}\left(\sum_{\ell=0}^{q}\alpha_{\ell i}(t-u)^{\ell} + X_i^T\beta + Z_i b\right) \\
&= \alpha\left(\sum_{\ell=0}^{q}\alpha_{\ell i}\int_0^{t_0(t)} e^{-au}(t-u)^{\ell}\,du + X_i^T\beta\int_0^{t_0(t)} e^{-au}du + Z_i b\int_0^{t_0(t)} e^{-au}du\right).
\end{aligned}
$$

A general formulation of Cox proportional model to characterize its association with the longitudinal outcomes can be expressed as

$$h\left(t \mid H_i(t), W_i, u_i\right) = h_0(t)\exp\left[W_i^T\gamma + \sum_{k=1}^{K} m_{ik}\left\{f_{ik}(t), u_i; \alpha_k\right\}\right], \qquad (22)$$

where $H_i(t) = \left\{f_{ik}(s), 0 \le s < t, 1 \le k \le K\right\}$ denotes the history of underlying longitudinal processes up to time t, W_i is a vector of time-fixed covariates with regression coefficients γ, $m_{ik}(.)$ specifies which components of $f_{ik}(t)$ are associated with the hazard at time t, u_i is a frailty on subject i, and α_k is a vector of outcome specific regression coefficients in function m_{ik}. Some examples of $m_{ik}(.)$ are given as

$$m_{ik}\left\{f_{ik}(t), u_i; \alpha_k\right\} = \sum_{g=0}^{q} \alpha_{gk} \frac{d^g f_{ik}(t)}{dt^g}, \text{with} \frac{d^0 f_{ik}(t)}{dt^0} = f_{ik}(t) \qquad (22.1)$$

$$m_{ik}\left\{f_{ik}(t), u_i; \alpha_k\right\} = \alpha_k^T b_{ik}, \qquad (22.2)$$

$$m_{ik}\left\{f_{ik}(t), u_i; \alpha_k\right\} = u_i \qquad (22.3)$$

Equation (22.1) in its simplest form (*q = 0*) models the effect of the underlying trajectory of the *kth* longitudinal outcome on event times. This parameterization is frequently used in joint models with interest being to evaluate impact of longitudinal biomarkers on the risk of clinically meaningful milestones. When q > 0, the formulation assumes that not only the longitudinal outcome at time t, but also its slope and curvature at time t, affects the event risk. To facilitate model parsimony and interpretability, values of q no greater than 2 are recommended. Such parameterisation is often used in connection with spline functions when modelling $f_{ik}(t)$, which guarantees tractable quantification of the derivatives. Equation (22.2) represents a commonly used strategy for longitudinal measurements with non-ignorable missing data caused by some terminating events such as death or dropout. This parameterization assumes that the risk of death or dropout at time t depends on deviation of the *ith* subject from the overall mean (e.g., deviation of

subject-specific slope from the population slope). In contrast to the previous two parameterizations, formulation (22.3) does not directly include any components in $f_{ik}(t)$ to characterize its association with the event risk, but rather assumes that the dependence between the longitudinal and survival endpoints is modelled by the joint distribution (e.g., multivariate normal distribution) between a frailty u_i and the random effects b_{ik}. This is a more flexible framework, which allows extra variation at the survival endpoint that cannot be explained by longitudinal data. The model reduces to (22.2) if u_i can be expressed as a linear function of b_{ik}.

3.5.3 The weights function

In equation (23), it was assumed that all the biomarker previous values from 0 to time t are of equal significance for studies with smaller follow up time, but for studies with many biomarkers, it is not reasonable to assume that the baseline values have the same importance as values with a longer time of follow-up. To account for possible complex biomarker overtime, an extension of the association between the biomarker values measured at a different time point and risk estimates require a flexible approach (Fisher and Lin 1999; Thomas 1988). In the majority of humans biomarker measured, it is expected that the more recent measures may be more pertinent for hazard estimation at time t compared to values measured that are far away from the same time point and this is where a weight function is required that is a decreasing function of time.

The expectation of this weight function has been seen in a study (Vacek 1997), in which normal and exponential cumulative distribution functions were applied to assign heavier weights to earlier values. The normal density function is applied to

assign the maximum weights to measures at period $s, \forall s \le t$ and is defined as (Abrahamowicz et al. 2006):

$$\omega(t-s)_* = \exp\left\{-\frac{(t-s)^2}{2\sigma^2}\right\},$$

where $\omega(.)$ is the appropriate weight function selected for different weights at different time points $(t-s)_* = t-s$ when $t > s$ and otherwise zero, s is the time prior to or equal to t and $t-s$ is the elapsed time since exposure. However, in this section, I used an approach for cumulative effect, weighted normal effects and weighted skewed-normal effects and truncated weighted skewed-normal effects distributions for flexibility of the model to estimate the necessary parameters from the dataset directly. I specified the differential weights in the formulation of cumulative effect as:

$$h_i(t) = h_0(t)\exp\left\{\alpha^T w_i(t) + \gamma\int_0^t \omega(t-s)_* f_t(s, b_i)ds\right\},$$

I use Area under the density curve (AUC) to calculate the time intervals of specific interest, in order to facilitate the relative importance of different biomarker history. The relative AUC is defined as:

$$\tau AUC_{\{t,(t-s)_*\}} = F(t-s)_* - F(t)$$

3.5.4 Statistical inference

Statistical inference for the joint models is developed through Maximum likelihood estimation (MLE) framework. It follows the conventional assumptions for longitudinal process and time to event process assumed mutually independent for random effects and the mutual covariates. The independent measurements value is taken by pre-specifying the visit process and the underlying censoring process is assumed

non-informative. In addition, the random effects as well as the observed data are assumed mutually independent. The likelihood function for the joint models is given as

$$\begin{aligned}\ell(\theta;data) &= \int P\left(T_i,\delta_i,Y_i :_{i=\overline{1,m}},\theta\right)P\left(b_i :_{i=\overline{1,m}}|\,\theta\right)db_1db_2...db_m \\ &= \int\left(\prod_{i=1}^{m}P\left(T_i,\delta_i,Y_i\,|\,b_i,\theta\right)\right)\left(\prod_{i=1}^{m}P\left(b_i\,|\,\theta\right)\right)db_1db_2...db_m \\ &= \prod_{i=1}^{m}\int P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)db_i\end{aligned}$$

where θ is the parameters vector of joint models, the argument $data$ represents observed data of all the individuals from both longitudinal and time to event processes, and the log-likelihood function is given as

$$\ell\left(\theta;data\right)=\sum_{i=1}^{m}\log\left(\int P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)db_i\right). \qquad (25)$$

The MLE approach is used for the statistical inference to develop the joint models for multivariate longitudinal process and time to event process, then, the parameters estimate proposed to minimize the log-likelihood function from equation (25). The estimations for the joint models are obtained by solving the equation (25).

$$\begin{aligned}\frac{\partial\ell\left(\theta;data\right)}{\partial\theta} &= \sum_{i=1}^{m}\frac{\frac{\partial}{\partial\theta}\left(\int P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)db_i\right)}{\int P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)db_i} \\ &= \sum_{i=1}^{m}\frac{\int\frac{\partial}{\partial\theta}\left(P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)\right)db_i}{\int P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)db_i} \\ &= \sum_{i=1}^{m}\frac{\int\frac{\partial}{\partial\theta}\left(\log\left(P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)\right)\right)P\left(T_i,\delta_i,Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)db_i}{\int P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)db_i} \\ &= \sum_{i=1}^{m}\frac{\int\frac{\partial}{\partial\theta}\left(\log\left(P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)\right)\right)P\left(T_i,\delta_i,Y_i,b_i\,|\,\theta\right)db_i}{\int P\left(T_i,\delta_i,Y_i\,|\,\theta\right)} \\ &= \sum_{i=1}^{m}\int\frac{\partial}{\partial\theta}\left(\log\left(P\left(T_i,\delta_i\,|\,b_i,\theta\right)P\left(Y_i\,|\,b_i,\theta\right)P\left(b_i\,|\,\theta\right)\right)\right)P\left(b_i\,|\,T_i,\delta_i,Y_i,\theta\right)db_i,\end{aligned}$$

It is assumed that the derivative change conditions of integral are satisfied, as it is one of the case study in the models.

With MLE approach, we proposed to compute the covariance matrix for the estimates using Fisher information and the negative of Hessian matrix is used in the Fisher information estimate, which is expressed as

$$\begin{aligned}\frac{\partial^2 \ell(\theta, data)}{\partial\theta^T \partial\theta} &= \sum_{i=1}^{m}\int\left(\frac{\partial^2}{\partial\theta^T\partial\theta}\log P(T_i,\delta_i,Y_i,b_i\mid\theta)\right)P(b_i\mid T_i,\delta_i,Y_i,\theta)\,db_i \\ &+\sum_{i=1}^{m}\int\frac{\partial}{\partial\theta}\log P(T_i,\delta_i,Y_i,b_i\mid\theta)\frac{\partial}{\partial\theta^T}\log P(T_i,\delta_i,Y_i,b_i\mid\theta)P(b_i\mid T_i,\delta_i,Y_i,\theta)\,db_i \\ &-\sum_{i=1}^{m}\left(\int\frac{\partial}{\partial\theta}\log P(T_i,\delta_i,Y_i,b_i\mid\theta)P(b_i\mid T_i,\delta_i,Y_i,\theta)\,db_i\right. \\ &\left.\int\frac{\partial}{\partial\theta^T}\log P(T_i,\delta_i,Y_i,b_i\mid\theta)P(b_i\mid T_i,\delta_i,Y_i,\theta)\,db_i\right) \\ &= \sum_{i=1}^{m}E\left(\frac{\partial^2}{\partial\theta^T\partial\theta}\left(\log\left(P(T_i,\delta_i\mid b_i,\theta)P(Y_i\mid b_i,\theta)P(b_i\mid\theta)\right)\right)\mid T_i,\delta_i,Y_i\right) \\ &+\sum_{i=1}^{m}Var\left(\frac{\partial}{\partial\theta}\left(\log\left(P(T_i,\delta_i\mid b_i,\theta)P(Y_i\mid b_i,\theta)P(b_i\mid\theta)\right)\right)\mid T_i,\delta_i,Y_i\right).\end{aligned}$$

Therefore, the covariance matrix for the estimate parameters is approximately expressed as

$$\hat{C}(\hat{\theta}) = \left(\frac{\partial^2 \ell(\theta, data)}{\partial\theta^T\partial\theta}\right)^{-1}$$

Hence, the parameter for the statistical inference is obtained by $C(\hat{\theta})$. The EM-algorithm was used to compute MLE of parameter estimator and full exponential Laplace approximation was used for the posterior expectations of random effects functions. The simulation study proves and verifies the applicability of the approach.

3.5.5 Parameter Estimation Procedure

Parameter estimation can be done via a maximum likelihood method when the number of latent class is given. The likelihood function is written out based on the assumption longitudinal and survival endpoints are independent given the latent class membership. Let θ denote a vector of all parameters that appear in the model. The likelihood is given as

$$\ell(\theta; data) = \sum_{i=1}^{m} \log\left(\int P(T_i, \delta_i \mid b_i, \theta) P(Y_i \mid b_i, \theta) P(b_i \mid \theta) db_i\right)$$

where each component's density function can be written out based on equation (24). Note that the density $P(Y_i \mid b_i, \theta)$ is a product of the multivariate normal density of transformed $\bar{Y}_i$ and Jacobian of Beta transformation. A penalized likelihood approach can be used when the baseline hazard is estimated by splines. Maximum likelihood estimates are obtained via an expectation-maximization (EM) algorithm or some Newton Raphson type of algorithm to maximize the likelihood using the gradient and Hessian matrix. The EM algorithm relies on computation of the posterior probability of class membership in each E-step conditional on the observed data and current parameter estimates. The posterior class membership of subject can be determined using the class with the highest posterior probability. The assumption of conditional independence between longitudinal and survival endpoints can be tested using a residual analysis. In addition, the number of latent classes may be selected using the Bayes Information Criterion (BIC), which has been shown to work reasonably well compared to other criteria when determining the number of components in mixture models.

Parameter estimation is done by EM-algorithm by minimizing the likelihood function of the observed data, $\left(T_i, \delta_i, Y_i : i = \overline{1, m}\right)$, which is performed by iteration of E-step for

the expected log-likelihood of the complete data, $\left(T_i, \delta_i, Y_i, b_i : i = \overline{1,m}\right)$, which is a conditional on the observed data to estimate the current parameters. M-step iteration was also performed on the new parameter estimates by minimizing the log-likelihood expectation. One-step Newton-Raphson method is proposed for survival parameters and other parameters and each M-step iteration for estimating equations are

$$\sum_{i=1}^{m} \int \frac{\partial}{\partial \theta} \left(\log P\left(T_i, \delta_i, Y_i, b_i \mid \theta\right) \right) P\left(b_i \mid T_i, \delta_i, Y_i, \theta_0\right) db_i = 0$$

where θ_0 represents the current parameters estimates and new parameters expected to be estimated for the next iteration, until it converges. However, the expected functions for random effects conditional on the observed data for the current parameters estimates θ_0 are required and given as

$$\begin{aligned} E\left(h(b_i) \mid T_i, \delta_i, Y_i, \theta_0\right) &= \int h(b_i) P\left(b_i \mid T_i, \delta_i, Y_i, \theta_0\right) db_i \\ &= \frac{\int h(b_i) P\left(b_i, T_i, \delta_i, Y_i \mid \theta_0\right)}{P\left(T_i, \delta_i, Y_i \mid \theta_0\right)} db_i \\ &= \frac{\int h(b_i) P\left(b_i, T_i, \delta_i, Y_i \mid \theta_0\right) db_i}{\int P\left(b_i, T_i, \delta_i, Y_i \mid \theta_0\right) db_i} \\ &= \frac{\int h(b_i) P\left(T_i, \delta_i \mid b_i, \theta_0\right) P\left(Y_i \mid b_i, \theta_0\right) P\left(b_i \mid \theta_0\right) db_i}{\int P\left(T_i, \delta_i \mid b_i, \theta_0\right) P\left(Y_i \mid b_i, \theta_0\right) P\left(b_i \mid \theta_0\right) db_i} \end{aligned} \tag{26}$$

The expected functions are on the assumptions that (i) the observed data from different subject are mutually independent and (ii) the random effects and common covariates for both longitudinal and time point process are independent. The Gaussian Hermite quadrature method are proposed in some studies for the joint model inference (Henderson et al. 2000; Wulfsohn and Tsisatis 1997) but quite computationally intensive when the random effects dimension is increasingly large

and ordinary computer has difficulty to deal with such computational burden. Also, an Adaptive Gaussian Hermite quadrature method was proposed to consequently improve Gaussian Hermite quadrature method (Liu and Pierce 1994). Another straightforward method proposed by Henderson et al. (2000) and Hsieh et al. (2006), is Monte Carlo approach, which is also computationally intensive and with Monte Carlo error.

However, a fully exponential Laplace approximation would likely be proposed to deal with posterior expectations. From the equation (25), a second order approximation to the expectation $E\left(h(b_i)\,|\,T_i,\delta_i,Y_i\right)$ can be obtained using a fully exponential function to approximate the moment-generating function $E\left(\exp(h(b_i))\right)$, which has a positive integrand and the result is differentiated as shown below:

Let $M(s)=E\left(\exp(sh(b_i))\,|\,T_i,\delta_i,Y_i\right)$, $\partial M(s)/\partial s=E\left(h(b_i)\exp\left(sh(b_i)\right)|\,T_i,\delta_i,Y_i\right)$ and $\partial M(s)/\partial s\,|_{s=0}=E\left(h(b_i)\,|\,T_i,\delta_i,Y_i\right)$, therefore, I approximate $M(s)$ to obtain the posterior expectation as

$$E\left(\exp(sh(b_i))\,|\,T_i,\delta_i,Y_i\right)=\frac{\int e^{sh(b_i)}P\left(b_i,T_i,\delta_i,Y_i\right)db_i}{\int P\left(b_i,T_i,\delta_i,Y_i\right)db_i}$$
$$=\frac{\int\exp\left(sh(b_i)+\log P\left(b_i,T_i,\delta_i,Y_i\right)\right)db_i}{\int\exp\left(\log P\left(b_i,T_i,\delta_i,Y_i\right)\right)db_i}$$

where $G\left(b_i\right)=\log P\left(b_i,T_i,\delta_i,Y_i\right)$, $G^*\left(b_i\right)=sh\left(b_i\right)+\log P\left(b_i,T_i,\delta_i,Y_i\right)=sh\left(b_i\right)+G\left(b_i\right)$, and $\hat{b}_i=\operatorname{argmax}\log P\left(b_i,T_i,\delta_i,Y_i\right)$, in which

$$G'\left(\hat{b}_i\right)=0,$$
$$G\left(b_i\right)\approx G\left(\hat{b}_i\right)+\frac{1}{2}\left(b_i-\hat{b}_i\right)^T G''\left(\hat{b}_i\right)\left(b_i-\hat{b}_i\right),$$
$$G^*\left(b_i\right)\approx sh\left(\hat{b}_i\right)+G\left(\hat{b}_i\right)+\left(sh'\left(\hat{b}_i\right)\right)^T\left(b_i-\hat{b}_i\right)+\frac{1}{2}\left(b_i-\hat{b}_i\right)^T\left(sh''\left(\hat{b}_i\right)+G''\left(\hat{b}_i\right)\right)\left(b_i-\hat{b}_i\right)$$

where $G'(.)$ and $G''(.)$ represents the 1st and 2nd derivatives w.r.t b_i and similarly $sh'(.)$ and $sh''(.)$. The expression becomes

$$\begin{aligned}\int e^{G(b_i)}db_i &\approx \int e^{G(\hat{b}_i)+\frac{1}{2}(b_i-\hat{b}_i)^T G''(\hat{b}_i)(b_i-\hat{b}_i)}db_i \\ &= e^{G(\hat{b}_i)}\int e^{-\frac{1}{2}(b_i-\hat{b}_i)^T(-G''(\hat{b}_i))(b_i-\hat{b}_i)}db_i \\ &= e^{G(\hat{b}_i)}(2\pi)^{p/2}\left|-G''(\hat{b}_i)\right|^{-1/2},\end{aligned}$$

$$\begin{aligned}\int e^{G^*(b_i)}db_i &\approx \int e^{sh(\hat{b}_i)+G(\hat{b}_i)+(sh'(\hat{b}_i))^T(b_i-\hat{b}_i)+\frac{1}{2}(b_i-\hat{b}_i)^T(sh''(\hat{b}_i)+G''(\hat{b}_i))(b_i-\hat{b}_i)}db_i \\ &= e^{sh(\hat{b}_i)}e^{G(\hat{b}_i)}\int e^{-\frac{1}{2}H_1^T(-sh''(\hat{b}_i)-G''(\hat{b}_i))H_1+\frac{1}{2}(sh'(\hat{b}_i))^T(-sh''(\hat{b}_i)-G''(\hat{b}_i))^{-1}(sh'(\hat{b}_i))}db_i \\ &= e^{sh(\hat{b}_i)}e^{G(\hat{b}_i)}e^{\frac{s^2}{2}(h'(\hat{b}_i))^T(-sh''(\hat{b}_i)-G''(\hat{b}_i))^{-1}(sh'(\hat{b}_i))}(2\pi)^{p/2}\left|-sh''(\hat{b}_i)-G''(\hat{b}_i)\right|^{-\frac{1}{2}},\end{aligned}$$

where $H_1=(b_i-\hat{b}_i)-(-sh''(\hat{b}_i)-G''(\hat{b}_i))^{-1}sh'(\hat{b}_i)$ and the expectation is given as

$$E\left(\exp(h(b_i)s)\,|\,T_i,\delta_i,Y_i\right)\approx\frac{e^{sh(\hat{b}_i)+\frac{s^2}{2}(h'(\hat{b}_i))^T(-sh''(\hat{b}_i)-G''(\hat{b}_i))^{-1}(sh'(\hat{b}_i))}\left|-sh''(\hat{b}_i)-G''(\hat{b}_i)\right|^{-\frac{1}{2}}}{\left|-G''(\hat{b}_i)\right|^{-\frac{1}{2}}}$$

Differentiating the expected value becomes

$$\begin{aligned}&\frac{\partial E\left(\exp(h(b_i)s)\,|\,T_i,\delta_i,Y_i\right)}{\partial s} \\ &\approx\frac{e^{H_2(s)}H_2'(s)\left|-sh''(\hat{b}_i)-G''(\hat{b}_i)\right|^{-\frac{1}{2}}}{\left|-G''(\hat{b}_i)\right|^{-\frac{1}{2}}} \\ &+\frac{e^{H_2(s)}(-\frac{1}{2})\left|-sh''(\hat{b}_i)-G''(\hat{b}_i)\right|^{-\frac{1}{2}}\text{Trace}\left(\left(-sh''(\hat{b}_i)-G''(\hat{b}_i)\right)^{-1}\left(-G''(\hat{b}_i)\right)\right)}{\left|-G''(\hat{b}_i)\right|^{-\frac{1}{2}}},\end{aligned}$$

Since $H_2(s) = sh\left(\hat{b}_i\right) + \frac{s^2}{2}\left(h'\left(\hat{b}_i\right)\right)^T\left(-sh''\left(\hat{b}_i\right) - G''\left(\hat{b}_i\right)\right)^{-1}\left(h'\left(\hat{b}_i\right)\right)$, and $H_2'(s) = {}^{\partial H_2(s)}\!/_{\partial s}$,

therefore, I have

$$\frac{\partial E\left(\exp(h(b_i)s)\,|\,T_i,\delta_i,Y_i\right)}{\partial s}\Big|_{s=0}$$

$$\approx h(b_i) - \frac{1}{2}\text{Trace}\left(\left(-G''\left(\hat{b}_i\right)\right)^{-1}\left(-h''\left(\hat{b}_i\right)\right)\right),$$

$$E\left(h(b_i)\,|\,T_i,\delta_i,Y_i\right) \approx h(\hat{b}_i) - \frac{1}{2}\text{Trace}\left(\left(-G''\left(\hat{b}_i\right)\right)^{-1}\left(-h''\left(\hat{b}_i\right)\right)\right).$$

This is the Laplace approximation of the 2nd order used in the study.

CHAPTER FOUR

4.1 ANALYSIS OF THE ESTIMATIONS RESULTS

4.2 Joint model with correlated error:

4.2.1. Simulation study I: Distributions with AR(1) within-subject covariance

It is not an easy task to develop a statistical inference for positive definite and high dimensional covariance matrix parameters, particularly for the within-subject covariance matrices in this study. However, for linear mixed-effects models, there are available plug-in packages to obtain within-subject covariance estimators, such as *lme, lme4, nlme* by R software package. The value of the unknown covariance in the survival estimating equations was substituted with the output of estimators obtained from *lme4* results, which is fairly unbiased and consistent with right fitted model and large number of longitudinal outcomes as established from the simulation outcome results. The *lme4* package is embedded with the assumption that the random effects are from multivariate normal distributions but can also be generated from other distributions as illustrated with the simulation studies in each of the scenarios. Using bias analysis settings in each simulation scenarios and three different underlying distributions were considered for comparison: (i) normal distribution as explained in the bias analysis; (ii) mixture normal with 0.5 mix distribution generated by setting *sep*=3, chose upper triangular matrix to yield $\mathrm{cov}(b_i)$ and standard deviations of b_i in the case of normal but $D_{12} = 0.043$, and (iii) skew-normal distribution with skewness coefficients of -0.2 and 0.91 for b_{i0} and b_{1i} respectively. The details on how random effects were generated from each

of the distributions (i-iii) can be found in Appendix A. The same mean and covariance was used to generate each of the random effects b_i as norm case unless specified. Random effects distribution and the measurement error was set to be normally distributed with $\mu = 0$ and AR(1) of covariance structure with $\sigma^2 = 0.7$. The correlation parameter of autoregressive of lag 1 was conducted in the following scenario: scenario1: when the correlation is negative ($\rho = -0.65$), scenario 2: when the correlation is weakly positive ($\rho = 0.3$); scenario 3: when the correlation is positive ($\rho = 0.55$), and scenario 4: when the correlation is strongly positive ($\rho = 0.8$).

Each of the correlation with $t_i \in \{0, 2, 4, 8, 14, 24, 32, 40, 48, 56, 64, 72, 80\}$ was generated in order to make the autoregressive of covariance structure more identifiable. The simulations were conducted with Monte Carlo 500 datasets and $n = 500$ for each datasets with $\gamma = -1$ and $\eta = 0$. The inference of covariance parameters for survival and autoregressive was illustrated using four different approaches: the ideal (I), conditional score (CS), and naïve regression (NR) and the proposed generalized conditional score (GCS) summarized in Table 1-4. The ideal results are obtained by using the exact values of $X_i(t)$ in partial likelihood function to obtain the estimators of likelihood and standard deviation for the parameters of survival, which is a benchmark of perfect inference performance. The MEst estimations was used because MLE approach is very similar in performance with MEst approach, and relative bias is calculated through percentage of differences in mean of Monte Carlo estimates and true values. Figure 1A (Appendix F) presents the subject-specific longitudinal responses in the simulation study.

The results from the simulation study indicate that CS method overestimate the Cox model coefficients (γ and η) when the error terms are negatively correlated and

gives a significant bias for a strong correlation. For instance, the relative bias (RB) for γ estimator is more than 50% when the correlation is negative, $\rho = -0.65$ and it tends to reduce the coefficients of the regression model when the errors are correlated positively, while the larger correlation tends to larger bias as summarized in Table 1.

Table 1: Scenario I: Simulation results for three random effects distributions with ρ(-0.65) and σ²(0.7) of AR(1) within-subject covariance.

Methods	Normal			Mixture			Skew-normal		
	Mean	RB (%)	SD	Mean	RB (%)	SD	Mean	RB (%)	SD
$\gamma(-1)$									
Ideal	-1.03	0.43	0.07	-1.04	0.25	0.11	-1.03	1.01	0.09
CS	-1.01	56.1	0.15	-1.07	28.6	0.23	-1.01	62.3	0.16
GCS	-0.94	2.1	0.11	-0.97	1.02	0.08	-0.93	1.15	0.13
NR	-0.89	14.3	0.08	-0.86	11.6	0.06	-0.89	13.1	0.08
AR(1):									
σ^2**-CS**	0.813	31.61	0.039	0.802	29.09	0.041	0.822	31.62	0.040
$\sigma^2(0.7)$**-GCS**	0.634	0.52	0.027	0.632	0.51	0.029	0.590	0.53	0.028
$\rho(-0.65)$**-GCS**	-0.651	0.31	0.014	-0.653	0.04	0.015	-0.654	0.30	0.015

SD, Monte Carlo standard deviation; RB, percentage of estimated relative bias. Methods: Ideal; CS, conditional score; NR, naive regression and GCS, generalized conditional score.

However, when the correlation is strong, that is, $\rho = 0.8$ the conditional score approach has a reduced performance compared to that of naïve approach, and have a slightly larger variation compared to the naïve approach. The naïve approach is used to attenuate the coefficients of regression.

Table 2: Scenario II: Simulation results for three random effects distributions with ρ(0.3) and σ² (0.7) of AR(1) within-subject covariance.

Methods	Normal			Mixture			Skew-normal		
	Mean	RB (%)	SD	Mean	RB (%)	SD	Mean	RB (%)	SD
$\gamma(-1)$									
Ideal	-1.01	0.52	0.08	-1.02	0.40	0.062	-1.00	0.12	0.081
CS	-0.89	8.6	0.104	-9.26	5.3	0.105	-0.91	9.4	0.120
GCS	-1.03	2.4	0.13	-1.04	1.3	0.109	-1.04	1.7	0.127
NR	-0.78	13.7	0.067	-0.88	11.5	0.075	-0.79	16.3	0.071
AR(1):									
σ^2**-CS**	0.831	11.7	0.014	0.837	11.9	0.013	0.835	11.2	0.013
$\sigma^2(0.7)$**-GCS**	0.602	0.5	0.019	0.603	0.7	0.018	0.595	0.5	0.019
$\rho(0.3)$**-GCS**	0.652	0.8	0.021	0.653	1.2	0.021	0.645	2.5	0.021

SD, Monte Carlo standard deviation; RB, percentage of estimated relative bias. Methods: Ideal; CS, conditional score; naive regression and GCS, generalized conditional score.

Meanwhile, In all the three different distribution approaches for random effects, it was noted that the mixture distribution for random effects were observed to have smaller bias in terms of parameter association of γ .

Table 3: Scenario III: Simulation results for three random effects distributions with ρ(0.55) and σ² (0.7) of AR(1) within-subject covariance.

Methods	Normal			Mixture			Skew-normal		
	Mean	RB (%)	SD	Mean	RB (%)	SD	Mean	RB (%)	SD
$\gamma(-1)$									
Ideal	-1.06	0.74	0.06	-1.02	0.31	0.07	-1.01	0.21	0.06
CS	-0.84	26.8	0.07	-0.93	18.5	0.15	-0.84	27.8	0.07
GCS	-1.05	4.3	0.11	-1.03	3.1	0.11	-1.02	1.3	0.13
NR	-0.83	28.8	0.05	-0.91	22.1	0.08	-0.82	30.1	0.05
AR(1):									
σ^2-CS	0.544	28.9	0.022	0.546	28.5	0.022	0.542	29.1	0.022
$\sigma^2(0.7)$-GCS	0.716	1.01	0.037	0.716	1.01	0.036	0.692	2.6	0.035
$\rho(0.55)$-GCS	0.614	0.91	0.032	0.614	0.8	0.031	0.593	2.6	0.031

SD, Monte Carlo standard deviation; RB, percentage of estimated relative bias. Methods: Ideal; CS, conditional score; RC, regression calibration; NR, naive regression and GCS, generalized conditional score.

In regards to the GCS approach proposed, it improves the inference for survival parameters with nearly unbiased estimates in spite of the underlying random effects distribution. It is observed that the mean of the Monte Carlo estimates are very close compared to the ideal approach performances of inference for AR(1) random error covariance as illustrated in Tables 1–4.

Table 4: Scenario IV: Simulation results for three random effects distributions with ρ(0.8) and σ² (0.7) of AR(1) within-subject covariance.

Methods	Normal			Mixture			Skew-normal		
	Mean	RB (%)	SD	Mean	RB (%)	SD	Mean	RB (%)	SD
$\gamma(-1)$									
Ideal	-1.06	0.71	0.06	-1.02	0.31	0.05	-1.01	0.21	0.06
CS	-0.75	36.6	0.05	-0.86	26.1	0.05	-0.74	47.5	0.06
GCS	-1.18	8.3	0.28	-1.15	5.7	0.22	-1.09	1.1	0.25
NR	-0.78	33.7	0.05	-0.85	25.8	0.04	-0.77	44.9	0.05
AR(1):									
σ^2-CS	0.376	66.9	0.019	0.378	66.5	0.041	0.374	67.2	0.009
$\sigma^2(0.7)$-GCS	0.732	4.7	0.075	0.727	3.8	0.029	0.689	4.7	0.064
$\rho(0.8)$-GCS	0.867	0.9	0.034	0.865	0.7	0.015	0.850	2.5	0.034

SD, Monte Carlo standard deviation; RB, percentage of estimated relative bias. Methods: Ideal; CS, conditional score; NR naive regression and GCS, generalized conditional score.

The GCS approach considers within-subject covariance and the corresponding survival estimators of survival part tend to have a larger variation and the approach can substantially reduce bias for the random effects.

4.2.2. Simulation study II: Cholesky decomposition of within-subject covariance

With the same scenarios as in simulation I, another simulation study was carried out for the underlying within-subject covariance using modified Cholesky decomposition instead of AR(1) structure but measurement errors for the underlying covariance remained the same. Scenarios of three different distribution of random effects were investigated again and each scenario was simulated with 500 Monte Carlo datasets and $n = 500$ in the orthogonal polynomial functions with a degree of q and d both equal to 4, respectively, and with underlying values $\omega = (0, -2, 2, -1)'$ and $\varpi = (-0.5, 1, 1.5, -1)'$. Both MLE and MEst estimations approaches similar to those aforementioned were investigated. It was noted that when the number of reasonable longitudinal outcomes are comparatively large, e.g. $8 \le n_i \le 16$, both MLE and MEst estimations of GCS method performed similarly and give almost unbiased estimates and credible statistical inference. However, if n_i are not adequately large due to conditional intensity and covariance complexity, the GCS method of MLE estimate have a tendency of reliable performance compared to MEst estimates.

Table 5 summarizes the results from the simulation performed to demonstrate the performance of the estimation approaches of scenarios with smaller possible n_i of $3 \le n_i \le 11$. The GCS-CD was used for the Cholesky decomposition approach to capture the measurement errors covariance and MEst approach of GCS for survival parameter in the form $K_i(t, \psi_s) = (S_i(t, \gamma, \Sigma_{i,t}), Z_i)'$ as it appears in the estimating equations, and it was proposed that $K_i(t, \psi_s) = (S_i(t, \gamma, \Sigma_{i,t}) - \gamma\sigma^2_{\hat{X}_i(t)gls}, Z_i)'$ to account for the adjusting variation term $\sigma^2_{\hat{X}_i(t)gls}$ in respect to γ estimating equation. The results presented in the Table 5 indicate that both conditional score (CS) and

naïve approach led to substantial bias, in which the CS method performance reduces to performance of naïve approach and both show very poor probability coverage. It was also observed that MEst estimates from GCS approach overestimate the survival parameters slightly but MLE estimates of GCS approach show a good performance statistically with almost unbiased estimates and the probability coverage show that the approach is valid. It was also noted that the random effects variation, $\text{var}(\varepsilon_{ij})$ is variant, while invariant in the autoregressive AR(1) approach from the last simulation on distributions with AR(1) within-subject covariance. However, the proposed REML approach shows a good statistical inference performance about the covariance parameters $\theta_c = (\omega, \varpi)'$.

Table 5: Scenario I: Simulation results for Cholesky decomposition of within-subject covariance

Methods	Normal			Mixture			Skew-normal		
	Mean	RB (%)	SD	Mean	RB (%)	SD	Mean	RB (%)	SD
$\gamma(-1)$									
Ideal	-1.06	0.72	0.06	-1.02	0.3	0.06	-1.04	0.5	0.06
CS	-0.22	79.3	0.04	-0.47	72.5	0.05	-0.24	77.8	0.05
GCS-CD	-1.23	13.5	0.24	-1.06	5.9	0.09	-1.22	14.7	0.24
GCS-CD(*mle*)	-1.10	2.1	0.11	-1.01	0.7	0.08	-1.01	0.9	0.09
NR	-0.38	71.8	0.04	-0.52	65.3	0.04	-0.31	71.8	0.04
MCDI:									
$\zeta_0(0)$	0.002	-	0.001	0.001	-	0.002	0.000	-	0.001
$\zeta_0(-2)$	-2.009	0.2	0.032	-2.002	0.0	0.031	-2.001	0.0	0.042
$\zeta_0(2)$	1.998	0.2	0.031	2.001	0.2	0.023	2.000	0.1	0.034
$\zeta_0(-1)$	-0.988	0.4	0.046	-1.002	0.1	0.034	-1.002	0.1	0.048
MCDI:									
$\zeta_0(0)$	-0.623	0.4	0.015	-0.510	0.4	0.021	-0.613	0.6	0.021
$\zeta_0(-2)$	0.999	1.3	0.121	0.998	1.5	0.113	0.999	1.2	0.119
$\zeta_0(2)$	1.568	0.2	0.128	1.505	0.6	0.134	1.508	0.3	0.139
$\zeta_0(-1)$	-1.016	0.6	0.134	-1.008	0.9	0.126	-1.011	1.1	0.133

SD, Monte Carlo standard deviation; RB(%), percentage of estimated relative bias. Methods: Ideal; CS, conditional score; NR, naive regression; GCS-CD, the MEst output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance; GCS-CD(mle), the MLE output of generalized conditional score estimation with Cholesky decomposition of within-subject covariance.

In many numerical algorithms, a good initial value can largely improve the algorithm performance. In the simulation study, the naïve estimators are used for the starting points for CS and GCS approaches for the survival inference and the initial values for $(\omega, \varpi)'$ can be randomly generated values near the underlying values as starting points, say within the range of ±0.5. In practice, without the knowledge of underlying

values, we can firstly set zeroes as the starting point, and then use the converged values with some slight modifications as the new starting points, which will usually converge to the consistent root. In the meantime, the outputs with several different initial values can be used to check the convergence for the real data application.

A joint mean-covariance model was applied with a modified Cholesky decomposition to estimate a reversible linear equation, which was used to depict the longitudinal outcomes and covariates in the model with subject identification and observation time point.

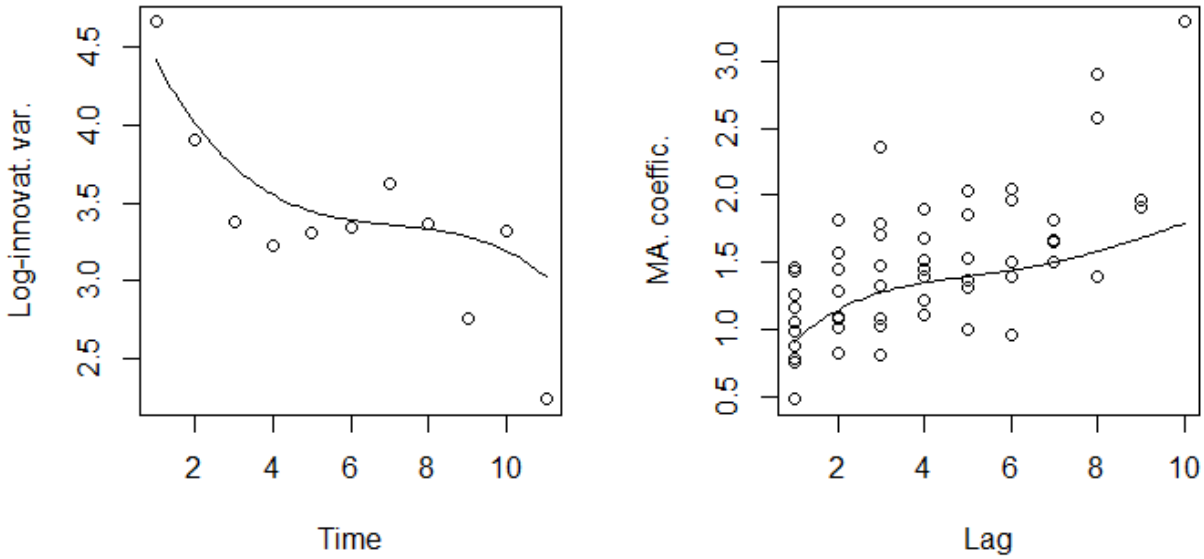


Figure 1: Modified Cholesky Decomposition model fits for log variances (left) and AR(1) coefficients (right) for simulation study

The fitted curve for the real data (Figure 1) shows that the fitted polynomial function curvature shape is captured well and indicate a good fit for autoregressive coefficients (AR(1)) in examining the AR(1) coefficient versus time lag between the measurements and the fitted curve.

4.2.3 Real Data Application

The proposed method is applied to TB-ESRD dataset, which includes a total number of 612 TB patients with impaired renal function. Each patient was assigned to a

treatment regime to compare the treatment effects and the survival end-point was a CD4 count ≥50% decline, which may lead to chronic TB or death of the patients. With time to progression to the survival end-point and other covariates such as age, weight, gender, CD4 count, and drug use were collected on each patient for about every 3 months. It was previously noted that CD4 biomarker is a good surrogate for treatment effect but may be subject to a considerable measurement error. The real data was used to illustrate the joint modelling in covariance within-subject in a longitudinal analysis of a balanced data correspond to the simulation conducted for balanced data, assigned to the $\log_{10} CD4$ count of TB patients with impaired renal.

It was observed that in both simulation and real dataset, the variation of $\log_{10} CD4$ seems to rise with respect to time with a slight reduction in $\log_{10} CD4$ on the last measurement in a real dataset. It was assumed that $X_i(t) = b_{io} + b_{i1}t$ denotes the underlying log of CD4 count for i subject at time t and observed measurements is represented by $W_i(t) = X_i(t) + \varepsilon_i(t)$, where $\varepsilon_i(t)$ can be time-independent or time-dependent. In the analysis of data application, death or chronic TB progression was considered as a censored event and the hazard function was analysed using a Cox proportional hazard model.

CD4 count was transformed using logarithm in order to make it easier to deal with a big number and the longitudinal $\log_{10} CD4$ covariate was pre-analysed with the R command of *nlme* and *lme4* with REML approach, which was used to fit the within-subject covariance with AR(1), independent, and compound Symmetry structures. We observed that from the *nlme* fitted of the linear mixed model for AR(1) within-subject covariance indicated with lowest BIC. From the analysis of the data, longitudinal $\log_{10} CD4$ was modelled with AR(1) within-subject covariance and data-

driven method modified Cholesky decomposition was used for the general case. Thus, three different starting point was randomly chosen to check the estimating convergence and the real data indicated that all the three point converged towards the same estimating values of $(\omega, \varpi)'$.

The analysis of the survival part of the real data suggested that the GCS approach was not significant, which is supported by the study of Song et al. (2002) which suggests the CD4 count and treatment effect was not significant and was proposed as a surrogate marker. Therefore, the treatment covariate was excluded from the final proportional hazard model. $\log_{10} CD4$ was included with other baseline covariates of interest into the final PH model such as age, BMI, and gender, where BMI from the analysis indicates that it was not significant in the final output from the model with the longitudinal $\log_{10} CD4$ covariate, age and gender. The statistical Inference for real data analysis results for the CS and GCS approaches, and for the GCS approach, the within-subject covariance was captured by AR(1) with lme R command and by the data-driven method of the modified Cholesky decomposition. Both MEst and MLE estimations are presented for each approach, except for the GCS with Cholesky decomposition covariance, which is presented in Table 6. The standard deviations were estimated with sandwich rule as mentioned in the simulation study. The covariate age has a positive significant effect on the hazard rate, which indicate that the older the TB patients, the higher the risk of death. The results also indicated that gender has no significant effect on the hazard rate, which shows that gender does not increase the risk of death among the TB patients. However, it was kept in the model to illustrate its insignificance.

Table 6: Statistical Inference for survival analysis of real data results for the CS and GCS approaches

	CS	GCS-AR(1)	GCS-CD

Covariates	MEst	MLE	MEst	MLE	MLE
$\log_{10} CD4$	-3.124 (0.231)	-2.052 (0.228)	-2.282 (0.221)	-2.269 (0.211)	-2.442 (0.255)*
Age	0.036 (0.009)	0.036 (0.009)	0.036 (0.009)	0.036 (0.009)	0.036 (0.009)*
Gender	0.418 (0.213)	0.419 (0.212)	0.426 (0.206)	0.426 (0.205)	0.314 (0.292)

Values inside the parenthesis are the standard deviation, * indicates the significance of covariate of 95% Wald confidence interval.

The simulation studies showed MEst method for Cholesky decomposition might overestimate the parameters a bit, so the MLE approach for Cholesky decomposition was used exclusively to estimate the application of real datasets. The comparison results from the three approaches showed that GCS with Cholesky decomposition (GCS-CD) provided the more substantial coefficients of $\log_{10} CD4$ compared to CS approach, which is the same from the simulation results and attenuate the coefficients of regression if there is any positive correlation among the within-subject error exist. At the same time, the results indicate the estimates of GCS-AR(1) covariance are almost the same as the CS approach but differ and stronger in GCS-CD (modified Cholesky decomposition) for the longitudinal coefficient of $\log_{10} CD4$, which suggests that the capturing of the covariance within-subject are more accurate using the Cholesky decomposition approach than simple AR(1). Figure 3 shows the scatter plot of $\log_{10} CD4$ against time, with the six covariates measurement visibly symmetric and normally distributed. It is reasonable to assume that ε_i is normally distributed, that is $\varepsilon_i \sim N(0, \Sigma_i)$. A joint mean-covariance model was applied with a modified Cholesky decomposition to estimate a reversible linear equation, which was used to depict the longitudinal outcomes and covariates in the model with subject identification and observation time point.

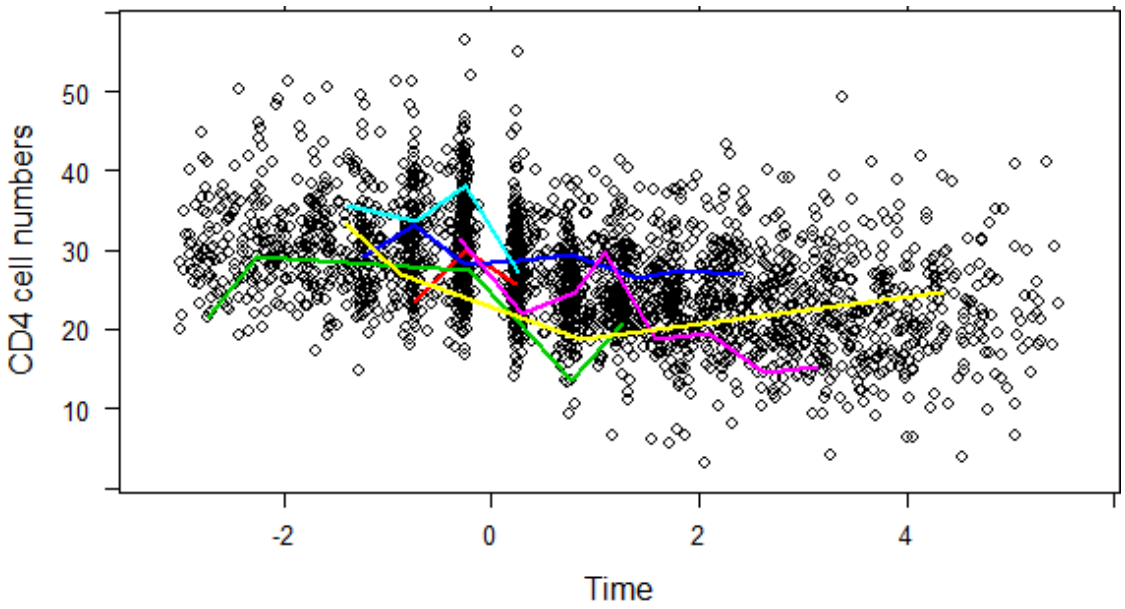


Figure 2: shows the scatter plot of CD4 cell against time

The fitted curve for the real data shows that the fitted polynomial function curvature shape is well-captured and indicates a good fit for autoregressive coefficients (AR(1)) in examining the AR(1) coefficient versus time lag between the measurements and the fitted curve (Figure 4).

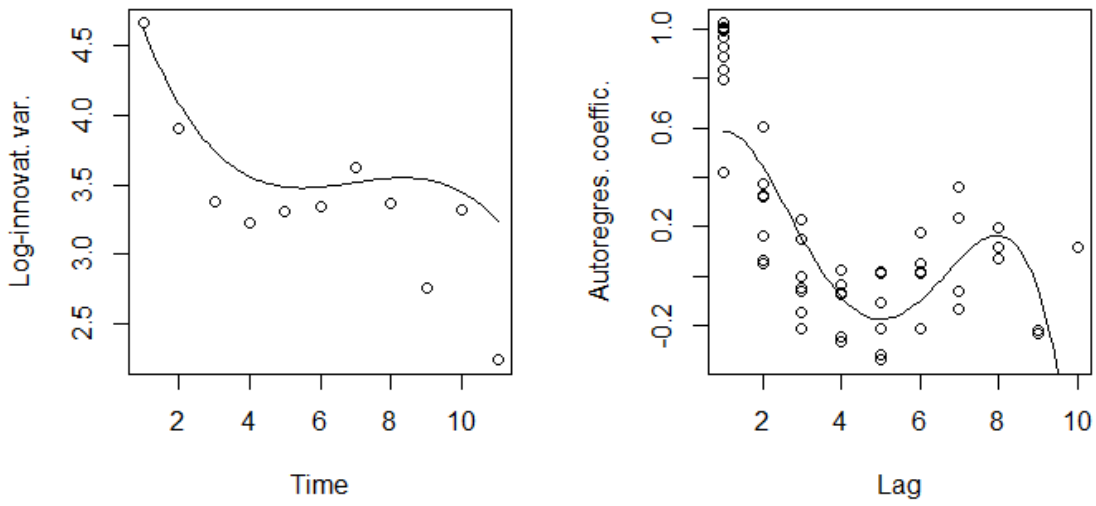


Figure 3: Real data Modified Cholesky Decomposition model fits for log variances (left) and AR(1) coefficients (right)

The MLE approach for modified Cholesky decomposition based on the covariance within-subject was fitted with other covariates and the longitudinal response of $\log_{10} CD4$ in a similar way to compare the fitted models using the log-likelihood of

the estimates. The covariance matrix of the fitted model produced the fitted curves with 95% confidence interval using the sandwich bootstrap method. The mean fitted curve with log-variance, AR(1) and the 95% confidence interval in Figure 4 showed that there is decreasing association of log-variance fitted with respect to time and curve shape of AR(1) fitted coefficient with a time lag.

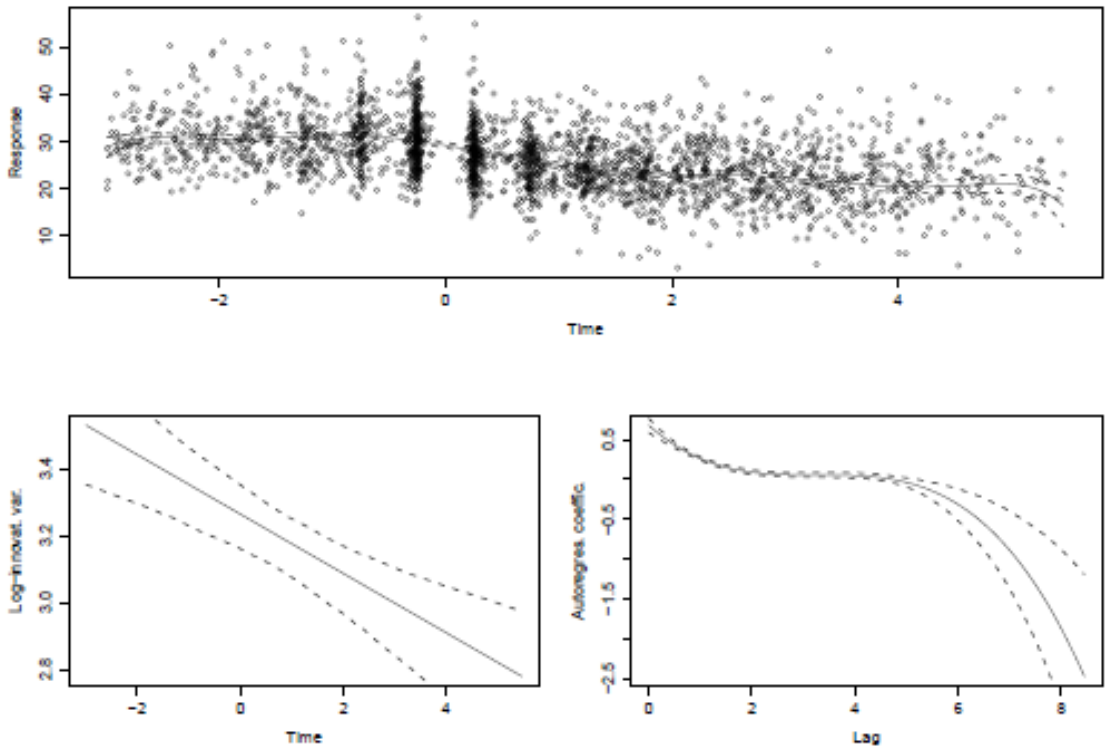


Figure 4: Modified Cholesky Decomposition model fits for mean against time (top), log variances vs. time (left) and AR(1) coefficients vs. time (right) and their corresponding 95% C.I.

4.3. Analysis Results on Joint Model with Cumulative Effects Association Structures

4.3.1 Simulation study

A simulation study was conducted to evaluate the performance of the proposed methodology. Given the complexity and associated long computational time of the

models, we assumed and simulated a continuous longitudinal outcome in the form of:

$$y_i(t) = f_i(t, b_i) + \varepsilon_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})Z_n(t, \lambda_1) + (\beta_2 + b_{i2})Z_n(t, \lambda_2) + (\beta_3 + b_{i3})Z_n(t, \lambda_3) + \varepsilon_i(t),$$

Natural cubic splines were assumed for both the fixed and random effect part of the model, with $b = (b_{0i}, b_{1i}, b_{2i}, b_{3i}) \sim MN(0, P_{diag})$, $\{Z_n(t, \lambda_k) : k = 1, 2, 3\}$ and $\varepsilon_i(t) \sim N(0, \sigma^2)$ as previously defined. The time derived from a uniform distribution was simulated. For the survival part, adjustment was made only for the treatment group for simple survival, expressed as:

$$h_i(t) = h_0(t) \exp\left\{\alpha_1 Trtgroup_i + \gamma \int_0^t \omega(t-s)_* \, f_t(s, b_i) ds\right\},$$

where $\omega(t-s)_*$ was specified as a standard normal distribution. The baseline risk was simulated from a Weibull distribution $h_0(t) = \varphi t^{\varphi - 1}$, given $\varphi = 0.8445$. A censoring rate approximately 80% and a uniform censoring distribution was selected for censoring time with $\mu_c = 2.5$. The results of simulation for the performance of the proposed methodology seems to be good with relatively small bias (Table 1&2). The root mean square estimate (RMSE), with the exclusion of D_4 parameter has the variance-covariance parameter for the random effects correspond to the third and final interval of the cubic spline. This may be due to the number of insufficient repeated outcomes in the interval.

Analysis outputs for the simulated dataset similar to the renal serum trial data are summarized in Table 7. This include the classical joint models' inference with default current values setting, and current values denoted by JM. The model performance of the treatment effects and the change of serum creatinine with time indicated that both approaches give similar results. The results indicated that the treatment has a

positive effect on the longitudinal profile of serum creatinine biomarker. The higher serum creatinine biomarker suggests a lower hazard rate of death of patients with ESRD. The results from the two joint models suggest that treatment of serum creatinine may prolong survival of patients' with end-stage renal disease (ESRD). However, the cumulative analysis indicates that the effect of the current value of survival model shows that the level of serum creatinine (sCr) biomarker effect on the hazard rate of death may be time-varying, and requires further investigation. As expected, the shapes of the integration area were a bit different from the models with both current and cumulative values of longitudinal process for the survival analysis. The simulation results also indicated the proposed estimation approach works better for a cumulative effect, in which the estimations are almost unbiased and the Root Mean Square Error shows the best accurate estimate of the cumulative effect.

Table 7: Analysis outputs for the simulated dataset similar to renal serum trial data

		JM with Current values			JM with Cumulative effects		
	Coefficient	**Value**	**Bias**	**RMSE**	**Value**	**Bias**	**RMSE**
Longitudinal Process	Intercept	0.561	0.052	0.065	0.544	0.044	0.032
	ns(years, 3)1	0.684	-0.205	0.231	0.625	-0.199	0.188
	ns(years, 3)2	0.886	0.113	0.225	0.799	0.107	0.206
	ns(years, 3)3	0.712	0.501	0.607	0.701	0.522	0.556
	ns(years, 3)4	0.751	0.498	0.561	0.762	0.468	0.520
	σ_c	0.274	0.000	0.001	0.216	0.000	0.006
Survival Process	Treatment group	-0.079	0.052	0.228	-0.065	0.051	0.219
	Parameter association (γ)	2.616	-0.002	0.389	2.409	-0.001	0.305
	Scale parameter (σ)	1.003	0.089	0.399	1.000	0.066	0.328
	D[1, 1]	1.000	0.005	0.040	1.001	0.002	0.021
	D[2, 2]	1.092	-0.209	0.251	1.005	-0.214	0.208
	D[3, 3]	1.036	0.483	0.492	1.018	0.469	0.442
	D[4, 4]	0.992	1.073	1.289	0.973	1.123	1.247
	D[5, 5]	0.899	2.542	2.583	0.884	2.520	2.594

The survival density functions of the estimation of the parameters indicated in Figure 5 shows the same shape with that of the real data application. The simulation results

suggest that the estimations from the joint model under my proposed framework are always roughly unbiased and efficient regardless of the value of $b > 0$.

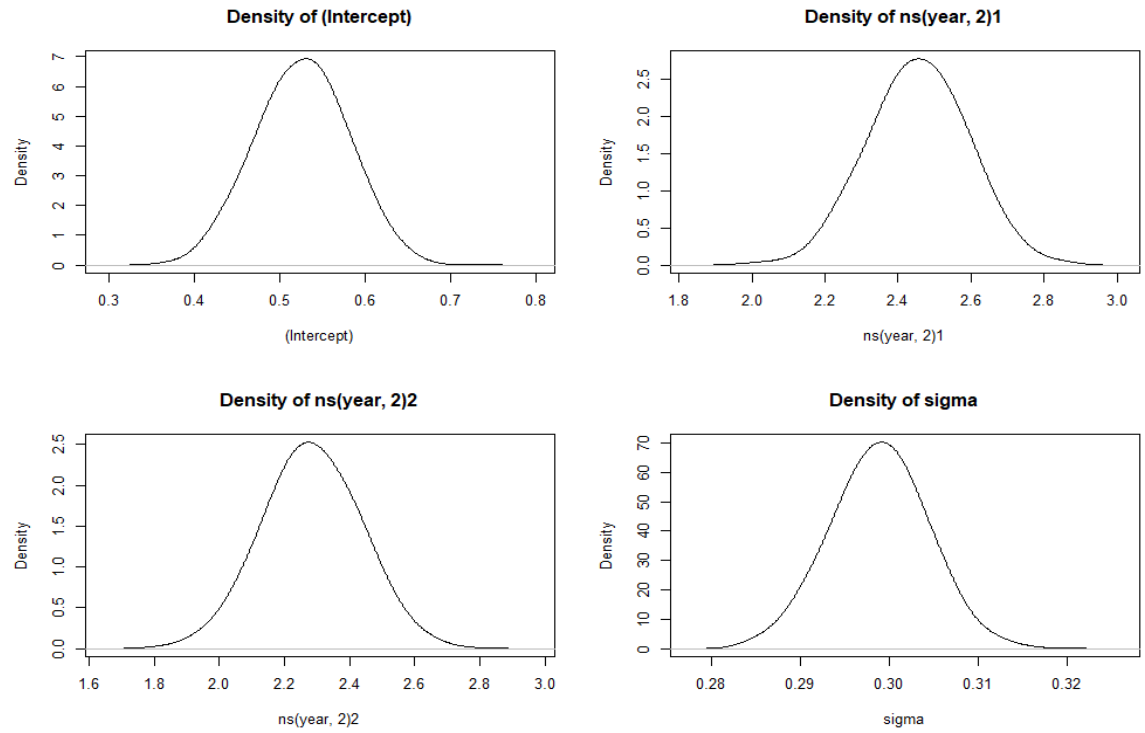


Figure 5: Density functions of estimators of survival parameters from Monte Carlo simulation with the underlying parameter setting and the hazard model

4.3.2. Application to Real Dataset

The proposed methodology was applied to analyse the dataset introduced in chapter 1 of this thesis. Every new patient enrolled in the study cohort every year and followed-up over some time. Patient's clinical information was recorded through annual follow-up protocol measurements. The year of entry was used as a baseline for each new individual and the urinary albumin test results were collected within at least a year of entry into the study. You may have a kidney disease or life-threatening kidney failure when serum creatinine level is high, or if two or more levels are high. The normal serum creatinine range is 0.6–1.1 mg/dL in women and 0.7–1.3 mg/dL in men. This test compares creatinine in blood and urine, and

development of kidney disease is dependent on many risk factors such as age, existences of diabetes, high blood pressure, heart disease, family member history of kidney disease, BMI, smoking status and obesity. A very high serum creatinine plays a very important role in a higher rate of renal disease. In addition, longer time of diabetes in combination with higher sCr values results in a higher renal failure and thus higher risk of End-stage renal disease (ESRD).

Of the 612 TB patients with impaired renal function considered in this study consisting of both longitudinal and survival outcomes on clinical information biomarker of renal disease, only 383 patients were considered as per the inclusion criteria in the analysis presented herein. The baseline characteristics for all the patients included in the study are described in Table 8. The average age and diabetes duration at baseline calculated as time at the diagnosis were 49.4 years (SE = 0.194) and 11.1 years (SE = 0.030) respectively, and the mean baseline blood pressure 124.1 mmHg (SE = 1.62).

Table 8: Baseline Characteristics of TB patients with renal disease

Variable	Mean (SE); N = 2531	Mean (SE); N = 383
Age (years)	49.4 (0.194)	50.21(.528)
Serum creatinine	3.71(0.108)	3.29(0.237)
Urinary albumin	3.39(0.010))	3.51(0.022)
Blood pressure	124.06(1.620))	122.33(2.927)
Diabetes	11.06(0.030)	10.79(0.048)
Family history	3.27(0.017)	3.05(0.045)
status2	.48(0.010)	.57(0.025)

Figures 6 and 7 plots were used to describe the survival and longitudinal outcomes that illustrate the Kaplan-Meier estimate of Treatment success for the two treatment groups, and the subject-specific of longitudinal trajectories for serum creatinine biomarker of patients with and without an endpoint.

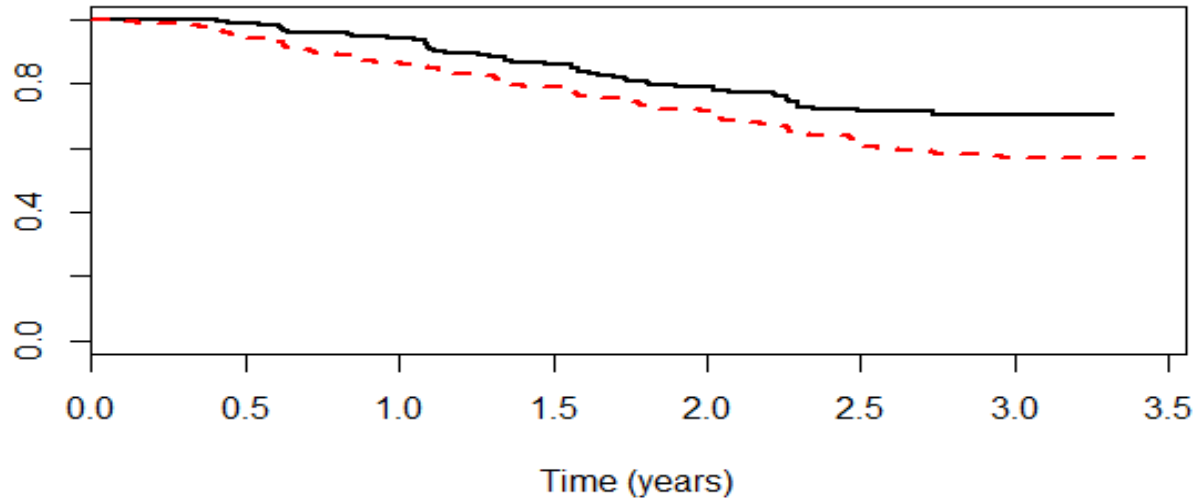


Figure 6: Kaplan-Meier estimator of treatment success probabilities for the two treatment groups

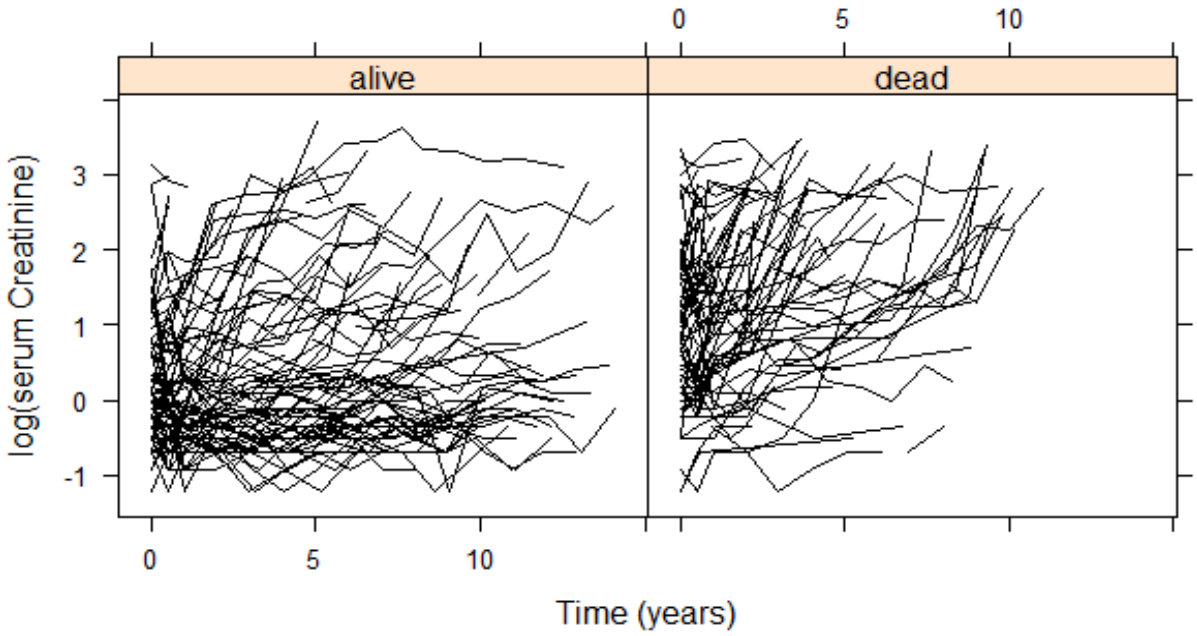


Figure 7: Subject-specific longitudinal trajectories for log serum creatinine for patients with and without endpoint

The natural cubic splines was applied to both the fixed and random effects component of the longitudinal model to establish the evidence of non-linearity of the longitudinal outcomes. We thus have;

$$f_i(t,b_i) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})Z_n(t,\lambda_1) + (\beta_2 + b_{i2})Z_n(t,\lambda_2) + \varepsilon_i(t),$$

The outcome from the longitudinal model includes time as a linear effect for the square root of the biomarker in both fixed and random effects as:

$$\vartheta[E\{y_i(t)\,|\,b_i\}] = \mathrm{Z}_n(t,b_i) = \beta_0 + b_{i0} + (\beta_1 + b_{i1}) \times \text{time},$$

The logit link was used for the albumin and the random effects follow a multivariate normal distribution. Penalized B-splines were used in the model for the square root of serum biomarker outcome. The shape of the *log(serum creatinine)* endpoint appears to be non-linear for many patients. In the longitudinal model, we controlled for the baseline age at entry, low or high level of sCr, blood pressure, diabetes duration and family history was controlled and expressed as

$$y_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})\mathrm{Z}_n(t,\lambda_1) + (\beta_2 + b_{i2})\mathrm{Z}_n(t,\lambda_2) + \beta_3 \mathrm{Age}_i + \beta_3 \mathrm{Diabetes}_i + \beta_3 \mathrm{sCr}_i + \beta_3 \mathrm{BP}_i + \varepsilon_i(t)$$

where $\{\mathrm{Z}_n(t,\lambda_i), i = 1,2\}$ represents the matrix of B-spline for the cubic natural spline of time at 50th percentile with one interior knot for the time to follow-up, $\varepsilon_i(t) \sim N(0, P_i)$ and $b_i \sim N(0, M)$ with $P_i = \sigma_c^2 I_{n_i}$ and M are unstructured variance-covariance matrix.

In the model for Cox regression, the weighted integral of the subject-specific linear estimate of the mixed model $f_t(s,b_i)$ was included in the linear estimate of the relative risk model, which represents the average subject-specific of *log(serum creatinine)* level. In the model, the treatment effect and age, was controlled and also their interaction in the model. Apart from the cumulative effect, two weight functions option were defined in fitting the proposed model . Both the probability density function for the normal distribution and skewed normal distribution were specified to fit the weighted cumulative effect and also fit an unweighted cumulative effect model in the form

$$h_i(t) = h_0(t)\exp\left\{\alpha_1 \text{Drug}_i + \alpha_2 \text{Age}_i + \alpha_3(\text{Drug}_i \times \text{Age}_i) + \gamma \int_0^t \omega(t-s)_* \, f_t(s, b_i)ds\right\}.$$

All the analyses in this study were run in R package.

DIC values were used to select the best-performed model and the results from the model for the current value, cumulative effect and weighted cumulative model have similar values, but the truncated skewed-weighted cumulative effect model performed best with DIC of 7454.173 (Table 9).

Table 9: Candidates model selection

Model	Df.	LPML	DIC	pD
Current value	1184	-4064.73	7819.52	1148.712
Cumulative effect	1185	-3976.741	7621.856	1147.112
Weighted normal effect	1186	-3966.625	7621.947	1153.718
Skewed-Weighted normal	1184	-3966.86	7603.106	1145.709
Truncated Skewed weighted normal	1184	-3911.524	7454.173	1127.597

The results from the models are summarized in Table 10–12. The cumulative effect with weighted description using truncated skewed normal density function was observed to be the best fit between the two weighted models for *log(serum creatinine)*. In the regression coefficients, there are some slight changes in the sub-models, in which 1 unit increase in the value of *log(serum creatinine)* levels is strongly associated with 4.3-fold (95% CI: [3.6-5.1]) increase of the risk of death event under the current value. For the cumulative effect model parameterization, a unit increase in the area under the longitudinal outcome profile equals 1.3 fold (95% CI: [1.2-1.3]) increase of the death risk and for the model with Weighted normal effect, a unit increase for the fold *log(serum creatinine)* corresponds to 5.1 fold (95% CI: [3.8-7.4]) increase in the risk of death. The models with weighted functions, the normal density function for the scale parameter, is observed to be 0.08 (95% CI: [0.04-0.14]). This means that measurement before time t for the serum creatinine within the last 2months and 4days are associated with the event risk at the same

time t, and the measurement after the current value is suggested to be irrelevant in the estimation of risk of an event at time t. The summaries are given in Table 10 and 11.

Table 10: Relative risk model with penalized-spline-approximated baseline risk function for the parameter estimates and 95% credible intervals under the joint modelling analysis for Log(serum Creatinine) longitudinal outcome Event Process

	Current value			Cumulative effect		
Parameters	**LogHazard (SE)**	**95% CI**	**P-value**	**LogHazard (SE)**	**95% CI**	**P-value**
Drug1	0.81 (0.1474)	(-0.61-2.39)	0.317	0.47 (0.1194)	(-1.05-2.06)	0.481
Age	0.06 (0.0014)	(0.04-0.07)	<0.001	0.05 (0.0014)	(0.03-0.07)	<0.001
Drug1:age	-0.01 (0.0027)	(-0.04-0.01)	0.408	-0.004 (0.0022)	(-0.03-0.03)	0.756
Association parameter	1.45 (0.0062)	(1.29-1.63)	<0.001	0.22 (0.0018)	(0.19-0.26)	<0.001
	Weighted normal effect					
	LogHazard (SE)	**95% CI**	**P-value**			
Drug1	-0.18 (0.1403)	(-2.02-1.63)	0.788			
Age	0.06 (0.0017)	(0.03-0.08)	<0.001			
Drug1:age	0.007 (0.0028)	(-0.03-0.04)	0.580			
Ass:(Intercept)	1.63 (0.0273)	(1.34-2.00)	<0.001			
Ass:ns(year, 2)1	0.59 (0.0057)	(0.47-0.72)	<0.001			
Ass:ns(year, 2)2	0.22 (0.0098)	(0.07-0.39)	0.003			
	Skewed Weighted normal			**Truncated skewed Weighted normal**		
	LogHazard (SE)	**95% CI**	**P-value**	**LogHazard (SE)**	**95% CI**	**P-value**
Drug1	0.86 (0.1027)	(-0.44-2.11)	0.199	0.64 (0.0971)	(-0.86-1.89)	0.306
Age	0.06 (0.0013)	(0.04-0.07)	<0.001	0.06 (0.0012)	(0.04-0.07)	<0.001
Drug1:age	-0.02 (0.0018)	(-0.04-0.01)	0.265	-0.01 (0.0019)	(-0.04-0.02)	0.390
Association parameter	2.85 (0.0165)	(2.52-3.21)	<0.001	1.52 (0.0178)	(1.30-2.02)	<0.001
	Weight function					
Scale parameter	0.08 (0.0040)	(0.04-0.14)	<0.001	0.08 0.0015)	(0.05-0.14)	<0.001
Shape parameter				6.23 (0.2261)	(0.79-9.82)	<0.001

A difference in the log-scale for serum creatinine corresponds to a ratio in the original scale and hence exp(Assoct) gives the corresponding hazard ratio for a doubling of serum bilirubin.

In Table 6, the cumulative-effect parameter estimation suggests that a unit increase in the longitudinal profile of diabetes corresponds to a 1.8 fold (95% CI: [1.4 – 2.5]) increase of the risk. This is slightly higher than the 1.5-fold (95% CI: [1.1 – 3.1]) increase of the risk estimated for the truncated weighted skewed normal parameter estimation function (best-fit model). The scale parameter estimation for both weighted normal and weighted skewed normal density function is 3.05 (95% CI:

[1.02 – 5.11]), which indicates that the diabetes responses within the last 9.15 years before t are risk associated event with almost the same time with maximum follow-up year (15 years). Therefore, the results suggest that both the baseline levels of the biomarker and the longitudinal outcome of the biomarker are strongly associated with the hazard of the risk event.

Table 11: Relative risk model with penalized-spline-approximated baseline risk function for the parameter estimates and 95% credible intervals under the joint modelling analysis for diabetes longitudinal outcome: Event Process

	Current value		Cumulative effect	
Parameters	**LogHazard (95% CI)**	**P-value**	**LogHazard (95% CI)**	**P-value**
Drug1	0.61 (-0.31 - 1.12)	0.274	0.62 (-0.21 - 1.32)	0.341
Age	0.03 (0.01 - 0.07)	<0.001	0.03 (0.02 - 0.07)	<0.001
Drug1:age	-0.04 (-0.02 -0.01)	0.321	-0.03 (-0.03 - 0.02)	0.699
Association parameter	0.58 (0.31 - 0.90)	<0.001	0.45 (0.29 - 0.76)	<0.001
	Weighted normal effect			
	LogHazard (95% CI)	**P-value**		
Drug1	-0.17 (-2.05 - 1.33)	0.544		
Age	0.04 (0.01 - 0.09)	<0.001		
Drug1:age	0.05 (-0.02 - 0.08)	0.422		
Association parameter	0.77 (0.22 - 1.19)	<0.001		
	Skewed Weighted normal		**Truncated skewed Weighted normal**	
	LogHazard (95% CI)	**P-value**	**LogHazard (95% CI)**	**P-value**
Drug1	0.45 (-0.14 - 1.61)	0.092	0.44 (-0.16 - 1.59)	0.106
Age	0.07 (0.04 - 0.08)	<0.001	0.08 (0.04 - 0.09)	<0.001
Drug1:age	-0.01 (-0.04 - 0.02)	0.073	-0.01 (-0.04 - 0.02)	0.090
Association parameter	0.73 (0.32 - 2.12)	<0.001	0.41 (0.10 - 1.14)	<0.001
		Weight function		
Scale parameter	3.05 (1.03 - 5.11)	<0.001	3.05 (1.02 - 5.11)	<0.001
Shape parameter			4.15 (0.66 - 5.27)	<0.001

A difference in the log-scale for serum creatinine corresponds to a ratio in the original scale and hence exp(Assoct) gives the corresponding hazard ratio for a doubling of serum bilirubin.

Using normal density weight function, parameterization is more significant when a random intercepts and random slopes structure is assumed for the longitudinal sub model, where a random-effect is used for the subject-specific deviations from the average intercept and slope. However, in this setting from Table 7, the parameter estimation indicates that patients at a higher level for the longitudinal outcome at baseline (intercept) or a steeper increase in the longitudinal trajectories (slope) are more likely to have the risk event.

Figure 8 shows that the comparison of estimated weight functions for *log(serum creatinine)*, indicating the degree of flexibility in the estimated weight function, with different rates of decreasing and different periods of relevance for the biomarkers. As seen from Figure 8, it was observed that the best model fit (truncated weighted skewed normal density) shows a maximal weight for values slightly earlier than the weighted skewed normal density.

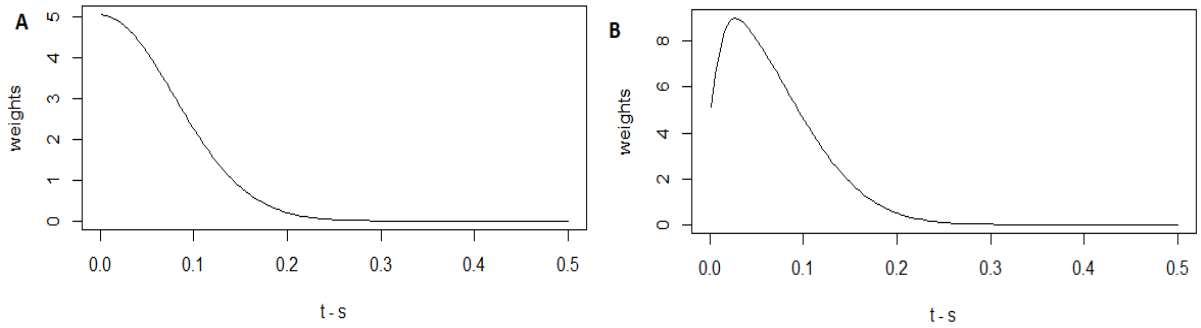


Figure 8: Comparison of Weight functions estimation. A-Weighted skewed normal density and B- Truncated Weighted Skewed normal density

Figure 9 shows the calculated values for the relative Area under the curve (AUC) for each of the estimated parameterization: current value, cumulative, weighted skewed normal and weighted truncated skewed normal functions, which show the relative importance of each of the biomarkers over several specific intervals of the follow up period.

Table 12: Longitudinal Process for parameter estimates and 95% credible intervals with D[i, j] element of the covariance matrix for the random effects under the joint modelling analysis for diabetes longitudinal outcome

	Current value		Cumulative effect	
	Coeff. (95% CI)	**P-value**	**Coeff. (95% CI)**	**P-value**
Intercept	15.67 (12.42 - 17.63)	<0.001	15.67 (12.45 - 17.68)	<0.001
Year	0.56 (050 - 0.71)	<0.001	0.56 (052 - 0.73)	<0.001
σ_c	1.32 (1.21 - 1.57)	<0.001	1.32 (1.98 - 1.55)	<0.001
D[1, 1]	14.33 (13.01 - 16.77)	<0.001	14.33 (13.01 - 16.77)	<0.001
D[2, 1]	3.42 (1.54 - 5.10)	<0.001	3.38 (1.04 - 5.03)	<0.001
D[2, 2]	2.08 (1.09 - 3.22)	<0.001	2.06 (1.11 - 3.25)	<0.001

	Weighted normal effect	
	Coeff. (95% CI)	**P-value**
Intercept	15.64 (12.54 - 17.73)	<0.001
ns(year, 2)1	0.55 (048 - 0.75)	<0.001
σ_c	1.31 (1.20 - 1.59)	<0.001
D[1, 1]	14.31 (13.01 - 16.75)	<0.001
D[2, 1]	3.40 (1.53 - 5.14)	<0.001
D[2, 2]	2.09 (1.11 - 3.25)	<0.001

	Skewed Weighted normal		Truncated skewed Weighted normal	
	Coeff. (95% CI)	**P-value**	**Coeff. (95% CI)**	**P-value**
Intercept	15.63 (12.41 - 17.59)	<0.001	15.62 (12.40 - 17.59)	<0.001
Year	0.54 (049 - 0.70)	<0.001	0.53 (049 - 0.71)	<0.001
σ_c	1.31 (1.19 - 1.58)	<0.001	1.32 (1.19 - 1.58)	<0.001
D[1, 1]	14.30 (12.88 - 16.91)	<0.001	14.30 (12.88 - 16.91)	<0.001
D[2, 1]	3.41 (1.55 - 5.12)	<0.001	3.40 (1.54 - 5.12)	<0.001
D[2, 2]	2.07 (1.11 - 3.23)	<0.001	2.07 (1.12 - 3.24)	<0.001

It was observed that the interval from 0 to 2.0 accounts for more than 40% density function of the relative biomarker and that the function for all the parameterization is closely more equally distributed across the entire period.

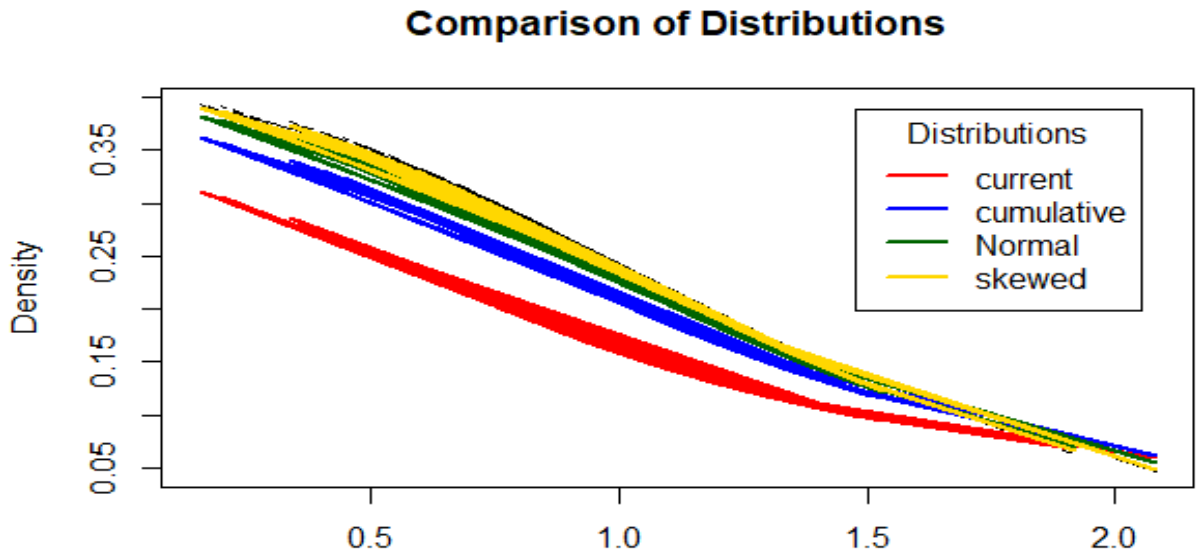


Figure 9: rAUC for Current, Cumulative, Normal and Skewed normal density function for the biomarker

The density functions plots for the estimation of survival parameters is illustrated in Figure 10. The five scenarios parameterization for baseline hazards indicate the survival parameters estimations obtained from the joint likelihood of longitudinal and survival outcomes maximization may not be an asymptotic normal distribution in some cases but they may be well-represented with some skewed distribution or distributions with a heavy tail. This supports the best- fit model of the skewed normal density function.

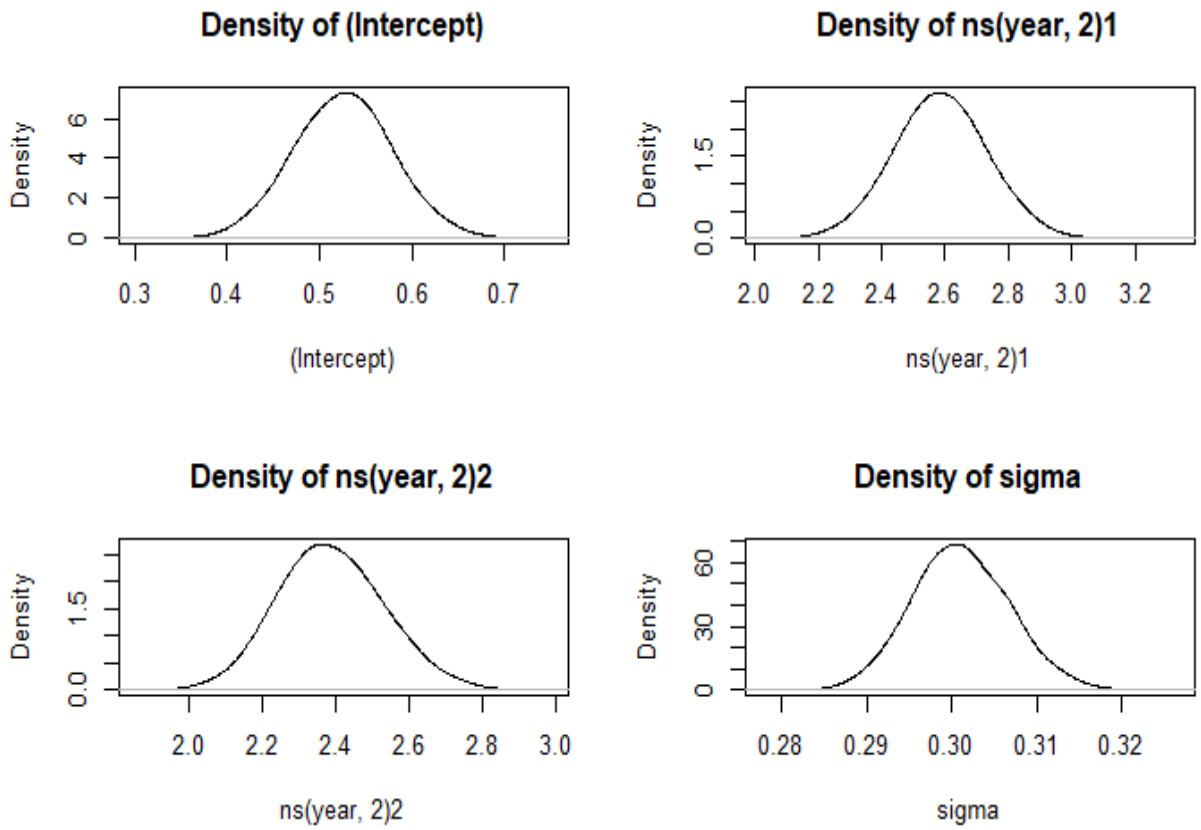


Figure 10: Density functions of estimators of survival parameters from Monte Carlo simulation with the underlying parameter setting and the hazard model for Weighted truncated skewed normal density.

CHAPTER FIVE

5.1 CONCLUSION AND RECOMMENDATION

5.2 Joint model for Correlated Measurement Error for Repeated Covariates

In joint modelling of longitudinal-survival measurements through linear mixed models and Cox proportional model, the error measurements are normally assumed independently and identically distributed from a normal distribution. The usual assumption is that random error is normally distributed but the mutual independence may not always be true. The simulation studies in this thesis indicate that the independence assumption can be violated if the random error is biased. Bias may be introduced on the Cox regression parameters for both longitudinal $\log_{10} CD4$ and baseline covariates if the inference for the survival regression parameters is influenced by the independence assumption violation. Most times, the likelihood approach may be computationally intensive due to the joint model complexity in terms of integration demand for the random effects and survival function.

Thus, the conditional score estimate was used to provide consistency and robustness estimates for survival parameters, which has proven to have vast computational advantage. The proposed generalized conditional score method is better in inference for the covariance structure of the random errors and the survival data simultaneously but may not be possible if the two-stage approach is implemented to cater for the covariance structure. However, in analysing the real data, the rule of significance was simply applied to select the polynomial degree of time for modelling the covariance structure using the modified Cholesky decomposition and to select a variable for the Cox model.

There is a need to apply caution when using the conditional score estimation approach and generalised conditional score estimation approach (proposed) for survival analysis under the joint modelling framework of longitudinal and survival data. This is because an adjusting term of variation of sufficient statistic is introduced to the exponential power of the hazard function to assist the inference. Sometimes, this can be a bit tricky as the exponential function is much more sensitive to the change of power. Without proper inference for the covariance of random errors, the generalised conditional score approach may also lead to biased inference for the survival parameters. It becomes expedient to utilise the classical likelihood-based approach to investigate further the impact on survival analysis if the assumption of independence of random errors is violated.

5.3 Joint Model with Cumulative Effects Association Structures

The results from this study are actuated by the biomedical inquiry about the association between biomarkers and the probability risk of renal disease. I proposed an extension to a Longitudinal-Survival joint model framework, where cumulative parameterization and statistical inference are used to compare parameterization of the current values in joint modelling and a truncated weighted-cumulative effect with parametric weight functions introduced to significant parameterization. Two similar but alternative weight functions (normal and skewed functions) were also researched for a cumulative effect, and used to directly estimate the scale and structure parameters from the data. The proposed weight function was further exemplified with a small simulation study. The inference indicates cumulative effect is adequate and the truncated weighted cumulative effect determines more precisely, the behaviour of the association that exists in time-varying covariate and the correlative risk of an event of interest. This allows the hazard function calculation at time t to rely on a

cumulative-effect for biomarker history and estimation in the most relevant duration of interest. Therefore, the current-value and cumulative-effect parameterization indicate remarkable illustrations for truncated weighted-cumulative effect parameterization.

However, an extension of a more general family of weight functions including more number of parameters to estimate, would be great development in the joint modelling. In a clinical perspective, it would be of a great interest to extend the methodology to allow functionality in the survival sub-model for recurrent events or progression, competing events, and interval-censored event. A left-truncated data and additional exogenous time-varying covariates can be included in the survival sub-model in the joint model.

This framework can be extended to other regression models for survival data such as AFT models.

Reference

Abrahamowicz, Michal, Gillian Bartlett, Robyn Tamblyn, and Roxane Du Berger. 2006. “Modeling Cumulative Dose and Exposure Duration Provided Insights Regarding the Associations between Benzodiazepines and Injuries.” *Journal of Clinical Epidemiology* 59(4):393–403.

Andrinopoulou, Eleni-Rosalina, Dimitris Rizopoulos, Marcel L. Geleijnse, Emmanuel Lesaffre, Ad J. J. C. Bogers, and Johanna J. M. Takkenberg. 2015. “Dynamic Prediction of Outcome for Patients with Severe Aortic Stenosis: Application of Joint Models for Longitudinal and Time-to-Event Data.” *Bmc Cardiovascular Disorders* 15:28.

Anon. 2012. “Joint Models for Longitudinal and Time-to- Event Data : With Applications in R.” 2012.

Argyropoulos, Christos and Mark L. Unruh. 2015. “Analysis of Time to Event Outcomes in Randomized Controlled Trials by Generalized Additive Models.” *PLoS ONE* 10(4):1–33.

Austin, Peter C., Douglas S. Lee, and Jason P. Fine. 2016. “Introduction to the Analysis of Survival Data in the Presence of Competing Risks.” *Circulation* 133(6):601–9.

Barrett, Jessica and Li Su. 2016. “Dynamic Predictions Using Flexible Joint Models of Longitudinal and Time-to-Event Data.” *Statistics in Medicine* 36(9):1447–60.

Bartlett, Jonathan W. and Ruth H. Keogh. 2018. “Bayesian Correction for Covariate Measurement Error: A Frequentist Evaluation and Comparison with Regression

Calibration.” *Statistical Methods in Medical Research* 27(6):1695–1708.

Berlin, Kristoffer S., Gilbert R. Parra, and Natalie A. Williams. 2014. “An Introduction to Latent Variable Mixture Modeling (Part 2): Longitudinal Latent Class Growth Analysis and Growth Mixture Models.” *Journal of Pediatric Psychology* 39(2):188–203.

Breslow, N. E., J. H. Lubin, P. Marek, and B. Langholz. 1983. “Multiplicative Models and Cohort Analysis.” *Journal of the American Statistical Association* 78(381):1–12.

Brown, Elizabeth R. 2009. “Assessing the Association between Trends in a Biomarker and Risk of Event with an Application in Pediatric HIV/AIDS.” *Annals of Applied Statistics* 3(3):1163–82.

Brown, Elizabeth R., Joseph G. Ibrahim, and Victor Degruttola. 2005. “A Flexible B-Spline Model for Multiple Longitudinal Biomarkers and Survival.” *Biometrics* 61(1):64–73.

Cox, David R. 1972. “Regression Models and Life-Tables.” *Journal of the Royal Statistical Society. Series B (Methodological)* 34(2):527–41.

Crowther, Michael J., Keith R. Abrams, and Paul C. Lambert. 2012. “Flexible Parametric Joint Modelling of Longitudinal and Survival Data.” *Statistics in Medicine* 31(30):4456–71.

Daniels, Michael J. and Yan D. Zhao. 2003. “Modelling the Random Effects Covariance Matrix in Longitudinal Data.” *Stat Med.* 22(10):1631–47.

David Collett. 2003. *Modelling Survival Data in Medical Research*.

Dupuy, Jean-Francois. 2014. “Accelerated Failure Time Models : A Review.”

International Journal of Performability Engineering 10(01):23–29.

E.L. Kaplan, Paul Meier. 1958. “Nonparametric Estimation from Incomplete Observations.” *Journal of the American Statistical Association* 53(282):457–81.

Elashoff, Robert M., Gang Li, and Ning Li. 2007. “An Approach to Joint Analysis of Longitudinal Measurements and Competing Risks Failure Time Data.” *Statistics in Medicine* 26:2813–35.

Faucett, Cheryl and Duncan Thomas. 1996. “Simultaneously Modelling Censored Survival Data and Repeatedly Measured Covariates: A Gibbs Sampling Approach.” *Statistics in Medicine* 15:1663–85.

Fisher, Lloyd D. and D. Y. Lin. 1999. “Time-Dependent Covariates in the Cox Proportional-Hazards Regression Model.” *Annual Review of Public Health* 20(6):145–57.

Fox, John and Sanford Weisberg. 2011. “Cox Proportional-Hazards Regression for Survival Data in R.” Pp. 1–20 in *An Appendix to An R Companion to Applied Regression, Second Edition*.

Henderson, R., P. Diggle, and A. Dobson. 2000. “Joint Modelling of Longitudinal Measurements and Event Time Data.” *Biostatistics* 1(4):465–80.

Hickey, Graeme L., Pete Philipson, Andrea Jorgensen, and Ruwanthi Kolamunnage-dona. 2016. “Joint Modelling of Time-to-Event and Multivariate Longitudinal Outcomes : Recent Developments and Issues.” *BMC Medical Research Methodology* 16(117):1–15.

Hougaard, P. 2000. *Analysis of Multivariate Survival Data*. Heidelber g: Springer.

Van Den Houten, Marijn M. L., Sandra C. P. Jansen, Anneroos Sinnige, Lijckle Van

Der Laan, Patrick W. H. E. Vriens, Edith M. Willigendael, Jan Willem H. P. Lardenoije, Jan Willem M. Elshof, Eline S. Van Hattum, Maarten A. Lijkwan, Ivan Nykliék, Ellen V. Rouwet, Mark J. W. Koelemay, Marc R. M. Scheltinga, and Joep A. W. Teijink. 2019. "Protocol for a Prospective, Longitudinal Cohort Study on the Effect of Arterial Disease Level on the Outcomes of Supervised Exercise in Intermittent Claudication: The ELECT Registry." *BMJ Open* 9(2):1–9.

Huang, Yangxin, X. Joan Hu, and Getachew A. Dagne. 2014. "Jointly Modeling Time-to-Event and Longitudinal Data: A Bayesian Approach." *Statistical Methods and Applications* 23(1):95–121.

Huong, Pham Thi Thu, Darfiana Nur, Hoa Pham, and Alan Branford. 2018. "A Modified Two-Stage Approach for Joint Modelling of Longitudinal and Time-to-Event Data." *Journal of Statistical Computation and Simulation* 88(17):3379–98.

Ibrahim, Joseph G., Ming-hui Chen, and Debajyoti Sinha. 2004. "Bayesian Methods for Joint Modeling of Longitudinal and Survival Data with Applications to Cancer Vaccine Trials." *Statistica Sinica* 14:863–83.

Ibrahim, Joseph G., Haitao Chu, and Liddy M. Chen. 2010. "Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data." *Journal of Clinical Oncology* 28(16):2796–2801.

Jin, Zhezhen. 2016. "Semiparametric Accelerated Failure Time Model for the Analysis of Right Censored Data." *Communications for Statistical Applications and Methods* 23(6):467–78.

Klein, By John P. and Melvin L. Moeschberger. 1997. "Survival Analysis : Techniques for Censored and Truncated Data (Statistics for Biology and Health)." *American Journal of Epidemiology* 149(2):142–84.

Kosmidis, Ioannis. 2014. “Bias in Parametric Estimation: Reduction and Useful Side-Effects.” *Wiley Interdisciplinary Reviews: Computational Statistics* 6(3):185–96.

Kosmidis, Ioannis and David Firth. 2010. “A Generic Algorithm for Reducing Bias in Parametric Estimation.” *Electronic Journal of Statistics* 4:1097–1112.

Laird, Nan M. and James H. Ware. 1982. “Random-Effects Models for Longitudinal Data.” *Biometrics* 38(4):963–74.

Lash, Timothy L., Matthew P. Fox, Richard F. MacLehose, George Maldonado, Lawrence C. McCandless, and Sander Greenland. 2014. “Good Practices for Quantitative Bias Analysis.” *International Journal of Epidemiology* 43(6):1969–85.

Lin, D. Y. and Zhiliang Ying. 1995. “Semiparametric Inference for the Accelerated Life Model with Time-Dependent Covariates.” *Journal of Statistical Planning and Inference* 44((1995)):47–63.

Liu, Qing and Donald A. Pierce. 1994. “A Note on Gauss-Hermite Quadrature.” *Biometrika* 81(3):624–29.

Moore, Dirk F. 2016. *Applied Survival Analysis Using R Series: Use R*. Springer.

Prentice, Ross. 1982.“Covariate Measurement Errors in the Analysis of Cohort and Case-Control Studies.”

Rizopoulos, Dimitris. 2010. “JM : An R Package for the Joint Modelling Of.” *Journal of Statistical Software* 35(9):1–33.

Rizopoulos, Dimitris. 2011a. “Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time- to-Event Data.” *Biometrics* 67(3):819–29.

Rizopoulos, Dimitris. 2011b. “Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data.” *Biometrics* 67(3):819–82.

Rizopoulos, Dimitris. 2012. *Joint Models for Longitudinal and Time-to-Event Data with Application in R*. CRC Press Taylor & Francis.

Rizopoulos, Dimitris, Laura A. Hatfield, Bradley P. Carlin, and Johanna J. M. Takkenberg. 2014. “Combining Dynamic Predictions From Joint Models for Longitudinal and Time-to-Event Data Using Bayesian Model Averaging.” *Journal of the American Statistical Association* 109(508):1385–97.

Robins, James and Anastosios A. Tsiatis. 1992. “Semiparametric Estimation of an Accelerated Failure Time Model with Time- Dependent Covariates.” *Biometrika* 79(2):311–19.

Rubin, Donald B. 2004. *Multiple Imputation for Nonresponse in Surveys*.

Song, Xiao and Ching-Yun Wang. 2014. “Proportional Hazards Model with Covariate Measurement Error and Instrumental Variables.” *J Am Stat Assoc* 109(504):1636–46.

Spiegelman, Donna, Bernard Rosner, and Roger Logan. 2000. “Estimation and Inference for Logistic Regression with Covariate Misclassification and Measurement Error in Main Study/Validation Study Designs.” *Journal of the American Statistical Association* 95(449):51–61.

Stefanski, Leonard A. and Raymond J. Carroll. 1987. “Conditional Scores and Optimal Scores for Generalized Linear Measurement- Error Models.” *Biometrika* 74(4):703–16.

Sweeting, Michael J. and Simon G. Thompson. 2011. “Joint Modelling of

Longitudinal and Time-to-Event Data with Application to Predicting Abdominal Aortic Aneurysm Growth and Rupture." *Biometrical Journal* 53(5):750–63.

Sylvestre, Marie Pierre and Michal Abrahamowicz. 2009. "Flexible Modeling of the Cumulative Effects of Time- Dependent Exposures on the Hazard." *Statistics in Medicine* 28:3437–3453.

Thomas, Duncan C. 1988. "Models for Exposure-Time-Response Relationships with Applications to Cancer Epidemiology." *Ann. Rev. Public Health* 9:451–82.

Tsiatis, Anastasios A. and Marie Davidian. 2001. "A Semiparametric Estimator for the Proportional Hazards Model with Longitudinal Covariates Measured with Error." *Biometrika* 88(2):447–58.

Tsiatis, Anastasios A. and Marie Davidian. 2004a. "Joint Modelling of Longitudinal and Time-to-Event Data: An Overview." *Statistica Sinica* 14(3):809–34.

Tsiatis, Anastasios A. and Marie Davidian. 2004b. "Joint Modelling of Longitudinal and Time to Event Data: An Overview." *Statistica Sinica* 14(3):809–34.

Vacek, Pamela M. 1997. "Assessing the Effect of Intensity When Exposure Varies over Time." *Statistics in Medicine* 16(5):505–13.

Vaupel, James W., Kenneth G. Manton, and Eric Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography* 16(3):439–54.

Wang, Chao, Naisyin Wang, and Suojin Wang. 2000. "Regression Analysis When Covariates Are Regression Parameters of a Random Effects Model for Observed Longitudinal Measurements." *Biometrics* 56(2):487–95.

Wang, Naisyin, R. J. Carroll, and Kung-yee Liang. 1996. "Quasilikelihood Estimation

in Measurement Error Models with Correlated Replicates." *Biometrics* 52(2):401–11.

Wang, Naisyin, Xihong Lin, Roberto G. Gutierrez, and Raymond J. Carroll. 1998. "Bias Analysis and SIMEX Approach in Generalized Linear Mixed Measurement Error Models." *Journal of the American Statistical Association* 93(441):249–61.

Wang, Naisyin, Xihong Lin, and Roberto G. Guttierrez. 1999. "A Bias Correction Regression Calibration Approach in Generalized Linear Mixed Measurement Error Models." *Communications in Statistics - Theory and Methods* 28(1):217–32.

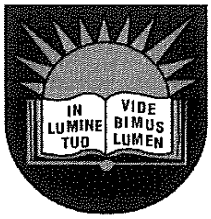
Wulfsohn, Michael S. and Anastasios A. Tsiatis. 2010. "A Joint Model for Survival and Longitudinal Data Measured with Error." *Biometrics* 53(1):330–39.

Wulfsohn, Michael S. and Anastasios A. Tsisatis. 1997. "A Joint Model for Survival and Longitudinal Data Measured with Error." *Biometrics* 53(1):330–39.

Wulfsohn, MS and A. .. Tsiatis. 1997. "A Joint Model for Survival and Longitudinal Data Measured with Error ." *Biometrics* 53(1):330–39.

Ye, Wen, Xihong Lin, and JMG Taylor. 2008. "A Penalized Likelihood Approach to Joint Modeling of Longitudinal Measurements and Time-to-Event Data." *Statistics and Its Interface* 1(December):33–45.

Appendix A: University Ethical Clearance Certificate



University of Fort Hare
Together in Excellence

ETHICAL CLEARANCE CERTIFICATE
REC-270710-028-RA Level 01

Certificate Reference Number:	QIN111SAZE01
Project title:	**A joint Modelling for Analysis of Longitudinal - Time to Event Data: Application to Medical Research and Clinical Trial Studies.**
Nature of Project	PhD in Statistics
Principal Researcher:	Adeboye Azeez
Supervisor:	Prof Y Qin
Co-supervisor:	Dr J Ndege

On behalf of the University of Fort Hare's Research Ethics Committee (UREC) I hereby give ethical approval in respect of the undertakings contained in the above-mentioned project and research instrument(s). Should any other instruments be used, these require separate authorization. The Researcher may therefore commence with the research as from the date of this certificate, using the reference number indicated above.

Please note that the UREC must be informed immediately of

- Any material change in the conditions or undertakings mentioned in the document;
- Any material breaches of ethical undertakings or events that impact upon the ethical conduct of the research.

The Principal Researcher must report to the UREC in the prescribed format, where applicable, annually, and at the end of the project, in respect of ethical compliance.

***Special conditions**:* *Research that includes children as per the official regulations of the act must take the following into account:*
Note: The UREC is aware of the provisions of s71 of the National Health Act 61 of 2003 and that matters pertaining to obtaining the Minister's consent are under discussion and remain unresolved. Nonetheless, as was decided at a meeting between the National Health Research Ethics Committee and stakeholders on 6 June 2013, university ethics committees may continue to grant ethical clearance for research involving children without the Minister's consent, provided that the prescripts of the previous rules have been met. This certificate is granted in terms of this agreement.

The UREC retains the right to

- Withdraw or amend this Ethical Clearance Certificate if
  - Any unethical principal or practices are revealed or suspected;
  - Relevant information has been withheld or misrepresented;
  - Regulatory changes of whatsoever nature so require;
  - The conditions contained in the Certificate have not been adhered to.

- Request access to any information or data at any time during the course or after completion of the project.

- In addition to the need to comply with the highest level of ethical conduct principle investigators must report back annually as an evaluation and monitoring mechanism on the progress being made by the research. Such a report must be sent to the Dean of Research's office.

The Ethics Committee wished you well in your research.

Yours sincerely

11/07/2018

Professor Pumla Dineo Gqola
Dean of Research

05 July 2018

Appendix B: Ethical Clearance from EC Department of Health

Province of the
EASTERN CAPE
HEALTH

Enquiries: Zonwabele Merile Tel no: 083 378 1202

Email: zonwabele.merile@echealth.gov.za Fax no: 043 642 1409

Date: 20 August 2018

RE: A JOINT MODELLING FOR LONGITUDINAL-TIME TO EVENT DATA WITH APPLICATION TO BIOMARKERS IN MEDICAL RESEARCH AND CLINICAL TRIAL STUDIES. (EC_201808_007).

Dear Mr A.N. Azeez

The department would like to inform you that your application for the abovementioned research topic has been approved based on the following conditions:

1. During your study, you will follow the submitted amended protocol with ethical approval and can only deviate from it after having a written approval from the Department of Health in writing.

2. You are advised to ensure, observe and respect the rights and culture of your research participants and maintain confidentiality of their identities and shall remove or not collect any information which can be used to link the participants.

3. The Department of Health expects you to provide a progress update on your study every 3 months (from date you received this letter) in writing.

4. At the end of your study, you will be expected to send a full written report with your findings and implementable recommendations to the Eastern Cape Health Research Committee secretariat. You may also be invited to the department to come and present your research findings with your implementable recommendations.

5. Your results on the Eastern Cape will not be presented anywhere unless you have shared them with the Department of Health as indicated above.

Your compliance in this regard will be highly appreciated.

SECRETARIAT: EASTERN CAPE HEALTH RESEARCH COMMITTEE

Appendix C: Simulation procedures used for the distributions in chapter 4

1. Mixture of normal distribution

Example of this is described in Davidian and Gallant (1993).

Mixture of bimodal random effects are generated from two normal distributions of $N(\mu, RR^T)$ and $N(-\mu, RR^T)$ with proportion of mixing of *p* and $\mu = \{(sep/2)(\sqrt{r_{11}^2 + r_{11}^2}, 0)^T\}$, where R represents the upper triangular matrix. We denoted $R = \begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix}$ and $RR^T = \begin{pmatrix} r_{11}^2 + r_{12}^2 & r_{12}r_{22} \\ r_{12}r_{22} & r_{22}^2 \end{pmatrix}$.

We let $Y \sim N(\mu, RR^T)$ and $Z = \begin{cases} Y & \text{if } \mu \le p \text{ and } \mu \sim U(0,1) \\ -Y & \text{if } \mu > p \text{ and } \mu \sim U(0,1) \end{cases}$

Then $E(\mathrm{Z}) = p\mu + (1-p)(-\mu) = 2p\mu - \mu = (2p-1)\mu$ and $E(\mathrm{Z}^2) = E(Y^2)$

Therefore, $Var(Z) = RR^T - \left((2p-1)\mu\right)^2$ and if $p = 0.5$, $E(Z) = 0$ as a special case.

We generated the mixture normal distributions by $Z_0 = Z - (2p-1)\mu + \mu_0$ with μ_0 and variance Σ_0 for $p = 0.5$ and $sep = 4$

2. Skew-normal distributions

We followed the skew-normal distribution discussed by Azzalini and Dalla (1996). The density of a random variable Z said to be a skew-normal with parameter λ is written as $Z \sim SN(\lambda)$ and expressed as

$$f(z;\lambda) = 2\varphi(z)\phi(\lambda z) \quad z \in \mathbb{R}$$

where $\varphi(z)$ and $\phi(\lambda z)$ represent the mean = 0 and variance = 1density and distributional function, respectively. The parameter λ varies in $(-\infty, \infty)$ in accordance

with the procedure of the skewness and $\lambda = 0$ corresponds to $N(0,1)$ density. For a random number generation, we used a more efficient variant of

$$Z = \begin{cases} Y & \text{if } \lambda Y > W \\ -Y & \text{if } \lambda Y \le W \end{cases}$$

to avoid rejection of samples. If Y_0 and Y_1 are independent variables and $\delta \in (-1,1)$, then

$$Z = \delta |Y_0| + (1-\delta^2)^{\frac{1}{2}} Y_1 \text{ is } SN(\delta(\lambda))$$

where $\delta(\lambda) = \delta/(1-\delta^2)^{\frac{1}{2}}$ and $\delta(\lambda) = \lambda/(1-\lambda^2)^{\frac{1}{2}}$. Also, it can be expressed as

$$Z_j = \delta_j |Y_0| + (1-\delta_j{}^2)^{\frac{1}{2}} Y_j \text{ where } j = 1,...,k$$

However, random variable $Z = (Z_1,..,Z_k)^T$ is a k-dimensional skew-normal variable with vector of shape parameter λ and dependent parameter ψ. For briefness, we denoted $Z \sim SN_k(\lambda,\psi)$ used for the expected mean and variance as

$$E(|Y_0|) = \int_{-\infty}^{0} (-y) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + \int_{0}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

$$= 2\int_{0}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 2\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u} du = \sqrt{\frac{2}{\pi}}$$

and $E(|Y_0|^2) = E(|Y_0^2|) = 1$, where $\text{var}(|Y_0|) = 1 - \frac{2}{\pi}$, $E(Z) = \sqrt{\frac{2}{\pi}} \omega_1(\delta)$ and

$Var(Z) = \omega_1(\delta)\,\text{var}(|Y_0|)\omega_1^T(\delta) + \omega_2(\delta)\psi\omega_2^T(\delta) = \Sigma$

we defined the following as

$$\omega_1(\delta) = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_k \end{pmatrix} \text{ and } \omega_2(\delta) = \begin{pmatrix} (1-\delta_1{}^2)^{\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & (1-\delta_2^2)^{\frac{1}{2}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & (1-\delta_k{}^2)^{\frac{1}{2}} \end{pmatrix}$$

where $diag(\Sigma) = (1 - \frac{2}{\pi}\delta_1{}^2, .., 1 - \frac{2}{\pi}\delta_k{}^2)^T$.

also, denoted $Z_0 = CZ$, $Var(Z_0) = \Sigma_0$ and $diag(\Sigma_0) = (\sigma_1^2, .., \sigma_k^2)$, where

$$C = diag\left(\frac{(diag(\Sigma_0))^{1/2}}{(diag(\Sigma))^{1/2}}\right) = \begin{pmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ 0 & 0 & \cdots & c_k \end{pmatrix}$$

$\Sigma = C^{-1}\Sigma_0 C^{-1}$, $\psi = \omega_1^{-1}(\delta)\left(\Sigma - \left(1 - \frac{2}{\pi}\right)\omega_1(\delta)\omega_1^T(\delta)\right)\omega_2^T(\delta)$ and $E(\mathrm{Z}_0) = CE(Z) = \sqrt{\frac{2}{\pi}}C\omega_1(\delta).$

Therefore, the skew normal distribution was generated using $Z_0 = \sqrt{\frac{2}{\pi}}C\omega_1(\delta) + \mu$.

Appendix D: Chapter four analysis codes

Joint model for Correlated Measurement Error for Repeated Covariates

WinBUGS and R codes examples

```
Model
{
  For (I in 1:N) {
  For (j in 1:T) {
        Y[I,j]~dnorm(mu[I,j], tau.c) →Yij ~N(uij+rc)
        Mu[I,j]<-alpha[i]+beta[i] * (x[i]-xbar) → uij =αi + βi(xi-ẍ)
  }
  Alpha[i]~dnorm(alpha.c, alpha.tau)
  Beta[i]~dnorm(beta.c, beta.tau)
}
  Tau.c~dgamma(0.001, 0.001)
  Sigma←1/sqrt(tau.c)
  Alpha.c~dnorm(0.0, 1, 0E-6)
  Alpha.tau~dgamma(0.001, 0.001)
  Beta.c~dnorm(0.0, 1, 0E-6)

Model {
 For (I in 1:N) {
       For (j in 1:M) {
            Y[I,j]~dnorm(muy[I , j], tauz)

muy[i]←beta1[1]+beta1[2]*t[j]+beta1[3]*t[j]*drug[i]beta1[4]*gender[i]+beta1[5]*previou
s[i]+beta1[6]*status[i]+u[I,1]+u[I,2]*t[j]
     }
     surt[i]~dweib(1, mut[i]) I(surt.cen[i], )
     log(mut[i])←beta2[1]+beta2[2]*drugs[i]+
beta2[3]*gender[i]+beta2[4]*previous[i]+beta2[5]*status[i]+r1*u[I,1]+r2*u[I,2]+r3*(u[I,1
]+u[I,2]*tee[i]
     u[I,1:2]~dmnorm(u0[ ], tau [ , ]
}
  tau[1:2, 1:2]~dwish (R[ , ], 23)
  beta1 [1:6]~dmnorm(betamu1[ ], Sigma1[ , ]
  tauz~dgamma(0.1, 0.1)
  beta2 [1:5]~dmnorm(betamu2[ ], Sigma2[ , ]
  r1~dnorm (0, 0.01)
```

```
  r2~dnorm (0, 0.01)
  r3~dnorm (0, 0.01)
}

Load readxl, mcmcplots
Load data
plot(tb$bmi ~ tb$time, xlab = "time length", ylab="bmi length")
sink("mod1.txt")
> cat("
+ MODEL LR1 {
+ for(i in 1:N) {
+ age[i] ~ dnorm(mu[i], tau)
+ mu[i] <- alpha + beta*bmi[i]
+ }
+ alpha ~ dnorm(0,0.001)
+ beta ~ dnorm(0,0.001)
+ tau <- pow(sigma, -2)
+ sigma ~ dunif(0,10)
+ }
+ ", fill = TRUE)
> sink()
> N = length(age)
> data = list("N","age","bmi")
> params = c("alpha", "beta")
> inits <- function () {list(alpha = rnorm(1), beta = rnorm(1), sigma = 1)}
> nc <- 3
> ni <- 5000
> nb <- 1000
> nt <- 1
> bugs.out <- bugs(data=data, inits=inits, parameters.to.save=params,
model.file="mod1.txt", n.chains=nc, n.iter=ni, n.burnin=nb, n.thin=nt, debug=TRUE,
DIC=TRUE, bugs.directory = "C:\\Program Files\\WinBUGS14",
working.directory=getwd())
> print(bugs.out, digits = 3)
> bugs.summary <- bugs.out$summary
> bugs.DIC <- bugs.out$DIC
> bugs.summary
> bugs.DIC
> plot(bugs.out)
> mcmcplot(bugs.out)


/*Specify the initial value for fixed effects parameters and covariances;*/
parms
```

```
beta0=5.7510
beta1=-0.0013
alpha1=-1.9965
alpha2=-3.3536
alpha3=-3.9501
alpha4=-3.5721
gamma=0.0411
eta1=1.0508
PsiSq=1.4594
SigmaSq=1.8513;
/*Non-negative constraints for variances;*/
bounds PsiSq SigmaSq >0;
/*Specify log-likelihood function;*/
MeanYij = beta0+beta1*tij+ui0;
VarYij = SigmaSq;
ll_long = (1-IND)*(-0.5*log(VarYij)-0.5*(Yij-meanYij)**2/VarYij);
p=alpha1*t1+alpha2*t2+alpha3*t3+alpha4*t4+gamma*age+eta1* ui0;
hij=1/(1+exp(-p));
ll_surv=ind*(yij*log(hij) +(1-yij)*log(1-hij));
ll=ll_surv +ll_long;
/*Specify model;*/
model yij ~ general(ll);
random ui ~ normal(0, PsiSq) sub=id;
run;


#################################################################
#
#  Simulation to compare sampling properties of three
#  different estimators for the mean of a distribution
#  based on an iid sample of size n:
#
#################################################################

#  function to view the first k lines of a data frame

view <- function(dat,k){

   message <- paste("First",k,"rows")
   krows <- dat[1:k,]
   cat(message,"\n","\n")
   print(krows)

}

#  function to calculate summary statistics across the 1000
#  data sets

simsum <- function(dat,trueval){
```

```
  S <- nrow(dat)

  MCmean <- apply(dat,2,mean)
  MCbias <- MCmean-trueval
  MCrelbias <- MCbias/trueval
  MCstddev <- sqrt(apply(dat,2,var))
  MCMSE <- apply((dat-trueval)^2,2,mean)
#  MCMSE <- MCbias^2 + MCstddev^2   # alternative lazy calculation
  MCRE <- MCMSE[1]/MCMSE

  sumdat <- rbind(rep(trueval,3),S,MCmean,MCbias,MCrelbias,MCstddev,MCMSE,
        MCRE)
  names <- c("true value","# sims","MC mean","MC bias","MC relative bias",
          "MC standard deviation","MC MSE","MC relative efficiency")
  ests <- c("Sample mean","Trimmed mean","Median")

  dimnames(sumdat) <- list(names,ests)
  round(sumdat,5)
}

#  function to generate S data sets of size n from normal
#  distribution with mean mu and variance sigma^2

generate.normal <- function(S,n,mu,sigma){

 dat <- matrix(rnorm(n*S,mu,sigma),ncol=n,byrow=T)

#  Note: for this very simple data generation, we can get the data
#  in one step like this, which requires no looping.  In more complex
#  statistical models, looping is often required to set up each
#  data set, because the scenario is much more complicated.  Here is
#  a loop to get the same data as above; try running the program and see
#  how much longer it takes!

#  dat <- NULL
#
#   for (i in 1:S){
#
#     Y <- rnorm(n,mu,sigma)
#     dat <- rbind(dat,Y)
#
#   }

  out <- list(dat=dat)
  return(out)
}

#  function to generate S data sets of size n from gamma
#  distribution with mean mu, variance sigma^2 mu^2
```

```
generate.gamma <- function(S,n,mu,sigma){

   a <- 1/(sigma^2)
   s <- mu/a

   dat <- matrix(rgamma(n*S,shape=a,scale=s),ncol=n,byrow=T)

#  Alternative loop

#   dat <- NULL
#
#   for (i in 1:S){
#
#      Y <- rgamma(n,shape=a,scale=s)
#      dat <- rbind(dat,Y)
#
#   }

   out <- list(dat=dat)
   return(out)
}

#  function to generate S data sets of size n from a t distribution
#  with df degrees of freedom centered at the value mu (a t distribution
#  has mean 0 and variance df/(df-2) for df>2)

generate.t <- function(S,n,mu,df){

    dat <- matrix(mu + rt(n*S,df),ncol=n,byrow=T)

#   Alternative loop

#   dat <- NULL
#
#   for (i in 1:S){
#
#      Y <- mu + rt(n,df)
#      dat <- rbind(dat,Y)
#
#   }

   out <- list(dat=dat)
   return(out)
}


#  function to compute the 20% trimmed mean

trimmean <- function(Y){mean(Y,0.2)}
```

```
#  set the seed for the simulation

set.seed(3)

#  set number of simulated data sets and sample size

S <- 1000

n <- 15

#  generate data  --Distribution choices are normal with mu,sigma
#  (rnorm), gamma (rgamma) and student t (rt)

#  a possible "fair" comparison would be to generate data from each
#  of these distributions with the same mean and variance and see how
#  the three methods perform on a relative basis under each condition

mu <- 1
sigma <- sqrt(5/3)

# out <- generate.normal(S,n,mu,sigma)  # generate normal samples
# out <- generate.gamma(S,n,mu,sigma)  # generate gamma samples
  out <- generate.t(S,n,mu,5)   # generate t_5 samples

outsampmean <- apply(out$dat,1,mean)

outtrimmean <- apply(out$dat,1,trimmean)

outmedian <- apply(out$dat,1,median)

summary.sim <- data.frame(mean=outsampmean,trim=outtrimmean,
          median=outmedian)

#view(round(summary.sim,4),5)

#  get summary statistics for each estimator

results <- simsum(summary.sim,mu)

#####################################################################

sampmean.ses <- sqrt(apply(out$dat,1,var)/n)

#  take the average

ave.sampmeanses <- mean(sampmean.ses)

#  coverage of usual confidence interval based on sample mean

t05 <- qt(0.975,n-1)
```

```
coverage <- sum((outsampmean-t05*sampmean.ses <= mu) &
        (outsampmean+t05*sampmean.ses >= mu))/S


pop <- 2
samp <- rnorm(100, 2, sd = 0.5)
bias(samp, pop)
bias(samp, pop, type = 'relative')
bias(samp, pop, type = 'standardized')

dev <- samp - pop
bias(dev)

# equivalent here
bias(mean(samp), pop)



> n.sim1 <- 500; set.seed(123)
> x1 <- rnorm(n.sim1, mean = 10, sd = 5)
> x2 <- rbinom(n.sim1, size = 100, prob = 0.5)
> e <- rnorm(n.sim1, mean = 0, sd = 1)
> b1 <- 2.5
> b2 <- -5
> a <- 2
> y <- a + b1 * x1 + b2 * x2 + e
sim.dat <- data.frame(y, x1, x2)
> freq.mod <- lm(y ~ x1 + x2, data = sim.dat)

model {
for(i in 1:N){
y[i] ~ dnorm(mu[i], tau)
mu[i] <- alpha + beta1 * x1[i] + beta2 * x2[i]
}
alpha ~ dnorm(0, .01)
beta1 ~ dunif(-100, 100)
beta2 ~ dunif(-100, 100)
tau ~ dgamma(.01, .01)
}

> sink("bayesmod.txt")
> cat("
modelf
for(i in 1:N)f
y[i] ~ dnorm(mu[i], tau)
mu[i] <- alpha + beta1 * x1[i] + beta2 * x2[i]
g
alpha ~ dnorm(0, .01)
```

```
beta1 ~ dunif(-100, 100)
beta2 ~ dunif(-100, 100)
tau ~ dgamma(.01, .01)
g
", fill=TRUE)
> sink()

> y <- sim.dat$y
> x1 <- sim.dat$x1
> x2 <- sim.dat$x2
> N <- nrow(sim.dat)

> bayes.mod.fit.R2WinBUGS <- bugs(model.file = "bayes.mod",
+ data = sim.dat.bugs,
+ parameters.to.save = bayes.mod.params,
+ inits = bayes.mod.inits,
+ n.chains = 3,
+ n.iter = 5000,
+ n.burnin = 1000,
+ n.thin = 1,
+ bugs.directory = "C:/Users/Azeez/WinBUGS14/")
```

Joint Model with Cumulative Effects Association Structures

```
library sn, JMbayes, splines2, splines: Loading required package: stats4
load data into R
> longk$status2 <- as.numeric(longk$status != "alive")
> survk$status2 <- as.numeric(survk$status != "alive")
> sfit <- survfit(Surv(years, status2) ~ drug, data = survk)
> lme.Fit1<-lme(log(serBilir)~ns(year,2) + age + spiders + albumin + sex, data =
longk, random =~ns(year,2)|id, method = "REML")
coxFit<-coxph(Surv(years, status2)~1, data = survk, x=TRUE)
> coxFit1 <- coxph(Surv(years, status2) ~ drug * age, data = survk, x = TRUE)
> lmeFit <- lme(log(serBilir) ~ ns(year, 2), data = longk, random = ~ ns(year, 2) | id)
> lmeFitt <- lme(log(serBilir) ~ ns(year, 2), data = longk, random = ~ ns(year, 2) | id,
method = "REML")
> jointFitt <- jointModelBayes(lmeFitt, coxFit1, timeVar = "year", n.iter = 30000) for
MCMC iterations:
> plot(jointFitt) Hit <Return> to see next plot:
> iForm<-list(fixed = ~0 + year + ins(year, 2), random = ~0 )
> iForm<-list(fixed = ~0 + year + ins(year, 2), random = ~0 + year + ins(year, 2),
indFixed = 1:3, indRandom = 1:3)
> jointFit.s <- update(jointFitt, param = "td-extra", extraForm = iForm) for MCMC
iterations:

> wf <- function(u, parms, t.max)
+ num <- dnorm(x = u, sd = parms)
> wf
```

```
> den <- pnorm(q = c(0, t.max), sd = parms)
> num / (den[2L] - den[1L])
> wf <- function(u, parms, t.max)
   num <- dnorm(x = u, sd = parms)
> summary(jointFitt, include.baseHazCoefs = TRUE)
> jointFit2 <- update(jointFitt, estimateWeightFun = TRUE, weightFun = wf,
priorShapes = list(shape1 = dunif), priors = list(priorshape1 = c(0, 5))) for MCMC
iterations:
> plot(jointFit2, which = "weightFun", max.t = max(survk$year))
> plot(jointFit2) Hit <Return> to see next plot:
> wfsn <- function(u, parms, t.max)
+     num <- dst(x = u, omega = parms[1], alpha = parms[2])
>     den <- pst(x = c(0, t.max), omega = parms[1], alpha = parms[2])
> wfsn <- function(u, parms, t.max) num <- dst(x = u, omega = parms[1], alpha =
parms[2])
> jointFit3 <- update(jointFitt, estimateWeightFun = TRUE, weightFun = wfsn,
priorShapes = list(shape1 = dunif, shape2 = dunif), priors = list(priorshape1 = c(0, 5),
priorshape2 = c(0, 10)))
MCMC iterations:
> plot(jointFit3, which = "weightFun", max.t = 0.5)
> plot(jointFit2, which = "weightFun", max.t = 0.5)
> jointFit4 <- update(jointFitt, param = "shared-RE", n.iter = 50000)
MCMC iterations:
> plot(jointFit4, which = "weightFun", max.t = 0.5)
> plot(jointFit4)
> fixef.JMbayes(jointFitt)
> anova.JMbayes(jointFitt, jointFit2, jointFit3, jointFit4)
> logLik.JMbayes(jointFitt)
> logLik.JMbayes(jointFitt, jointFit2, jointFit3, jointFit4)
> logLik.JMbayes(jointFit2)
> logLik.JMbayes(jointFit3)
> logLik.JMbayes(jointFit4)
```

Sensitivity analysis

```
> jointFitt10knots <- update(jointFitt, lng.in.kn = 10L)
MCMC iterations:
> plot(jointFitt10knots) Hit <Return> to see next plot:
> jointFitt20knots <- update(jointFitt, lng.in.kn = 20L)
MCMC iterations:
> cbind("10 knots" = fixef(jointFitt10knots), "15 knots" = fixef(jointFitt), "20 knots" =
fixef(jointFitt20knots))
> cbind("10 knots" = fixef(jointFitt10knots, process = "Event"), "15 knots" =
fixef(jointFitt, process = "Event"), "20 knots" = fixef(jointFitt20knots, process =
"Event"))
```

Appendix E: Motivating data subject-specific Longitudinal profiles

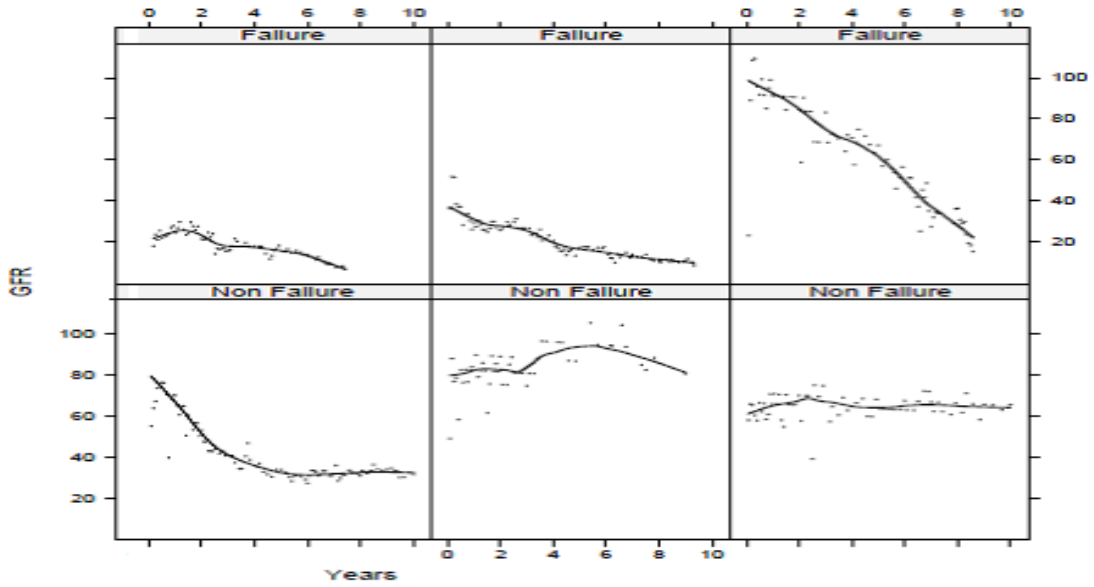
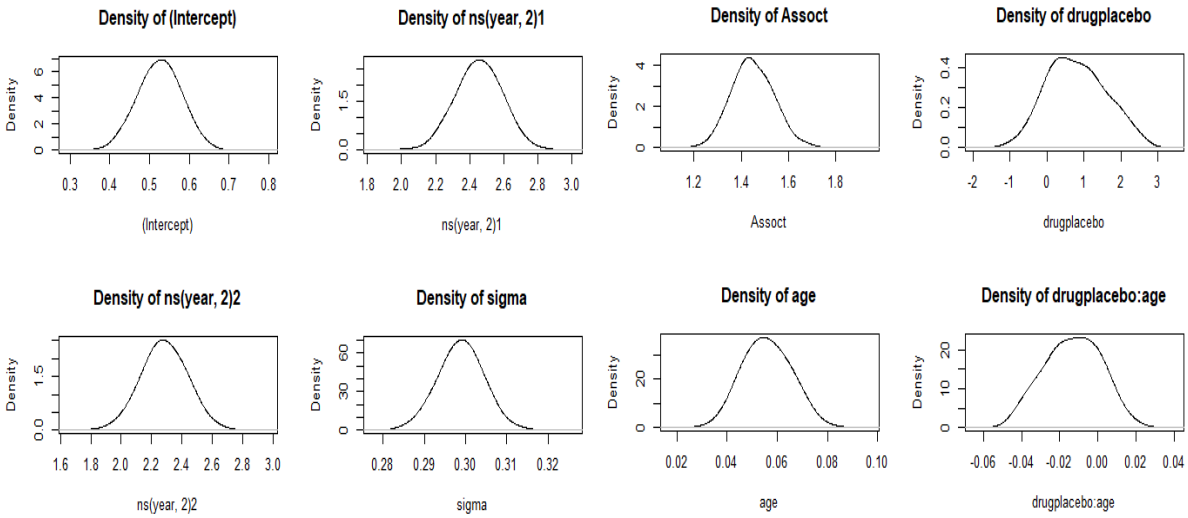


Figure-1s: Sample within subject-specific longitudinal profiles for six selected patients

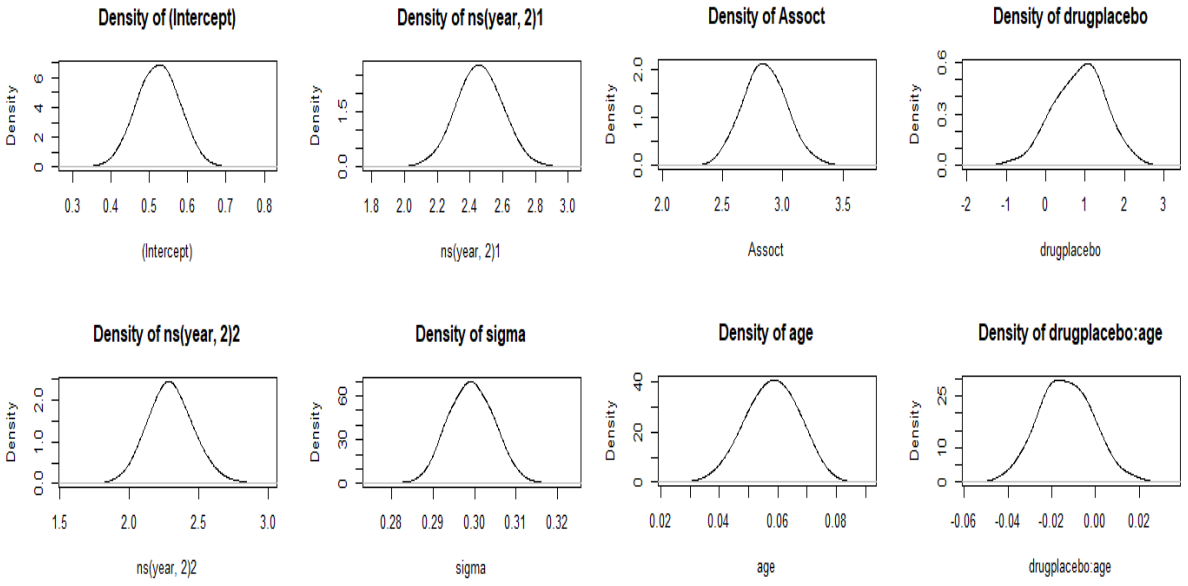
Appendix F: MCMC diagnostic plots

A. Kernel density estimation plots for the parameters of the longitudinal and survival submodels from all the joint model compared.

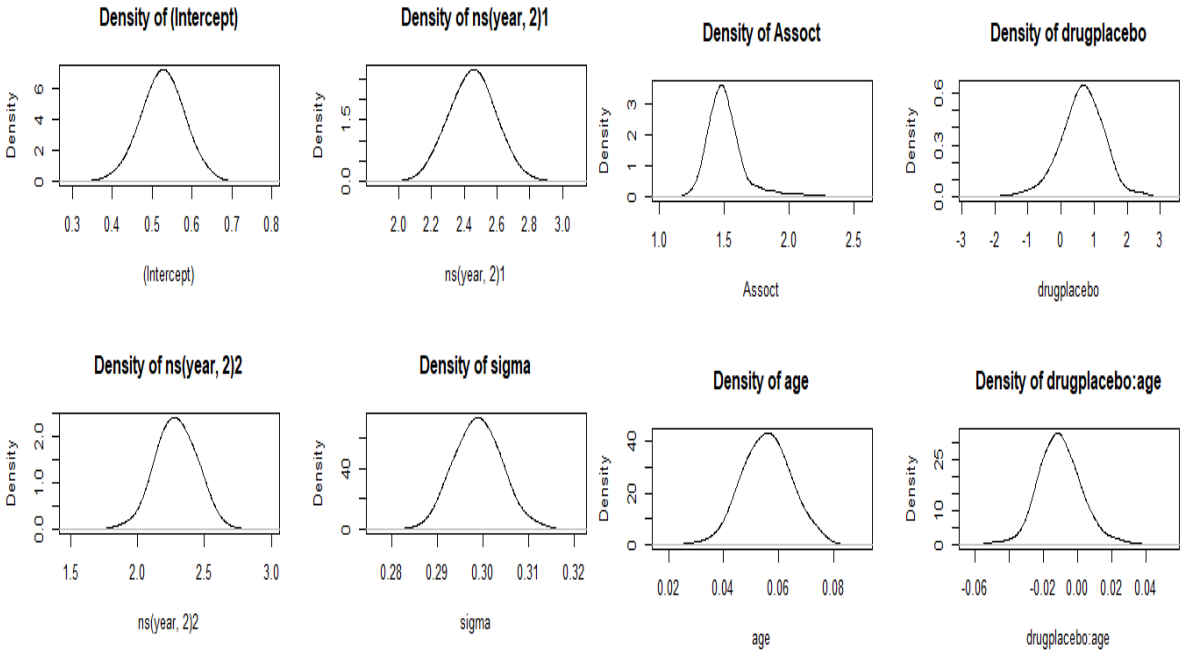
Model 1: Current value parameterization



Model 2: Cumulative effect

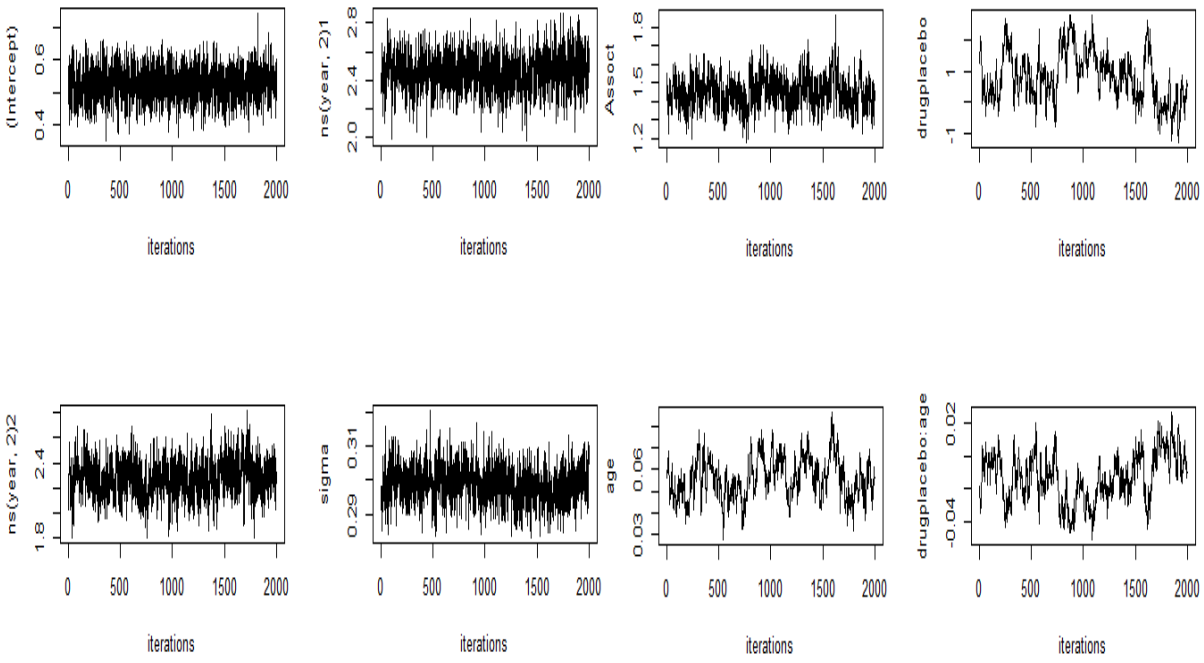


Model 3: Weighted Normal effect

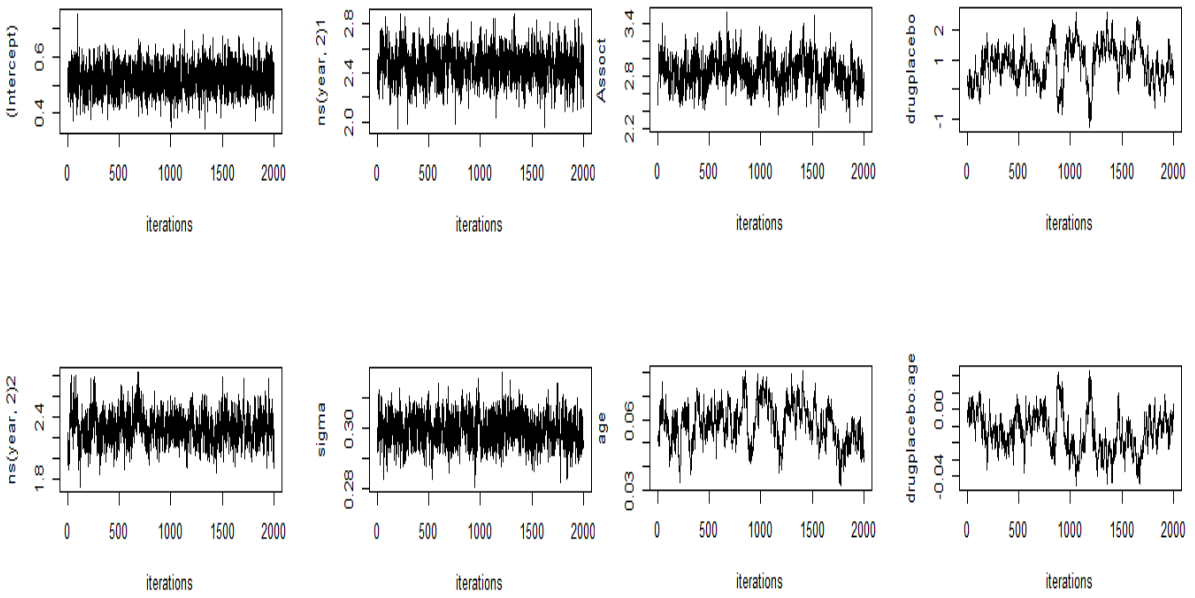


B. Trace plots for the parameters of the longitudinal and survival submodels from joint models.

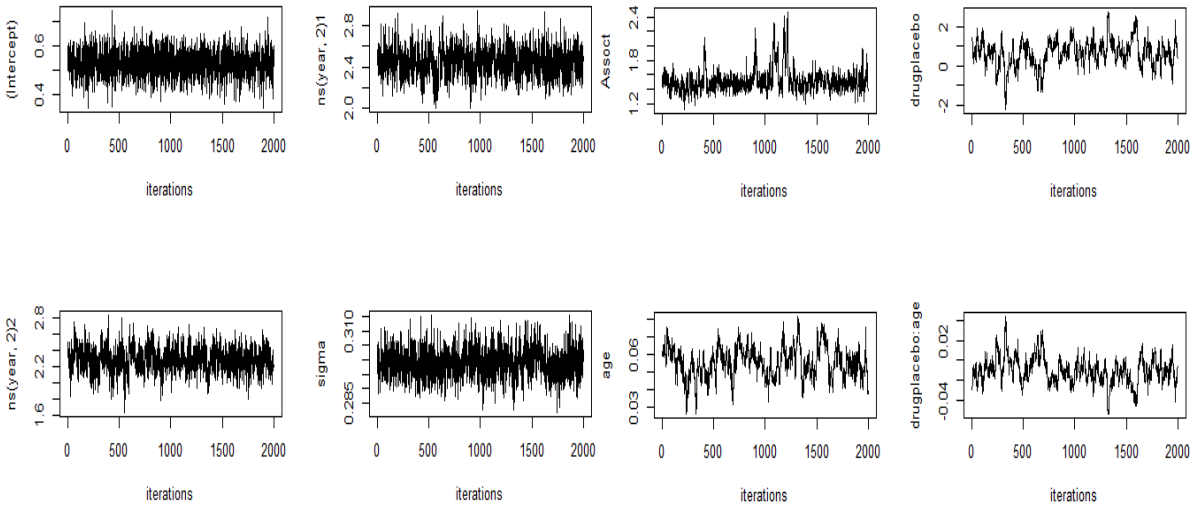
Model 1: Current value effect



Model 2: Cumulative effect



Model 3: Weighted normal effect



Model 4: Weighted skewed normal effect