Joint analysis of longitudinal and time to event data using accelerated failure time models: A Bayesian approach

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Collaborative Biostatistics Program of School of Public Health
University of Saskatchewan
Saskatoon

By
Nyla Basharat

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Director
Collaborative Biostatistics Program
School of Public Health
104 Clinic Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C9
Canada
OR
Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Joint modeling is a collection of statistical methods to properly handle a longitudinal response while investigating its effects on time to the occurrence of an event. Joint modeling also allows an investigation of the effects of baseline covariates on both the longitudinal response and the event process. In practice, the inspiration of biostatistical research arises from clinical and biomedical studies. The data collected from these studies have always been getting attention due to their particular features that need special consideration when doing an analysis. New statistical methods have developed over time to handle an analysis of such data coming from these sources. A typical clinical study often involves collecting repeated measurements on a biomarker (e.g., lvmi measurements) along with an observation of the time to the occurrence of an event (e.g., death), resulting in a joint modeling setup, a model becomes increasingly popular in clinical studies. Joint models can be formulated with a probability distribution (parametric models) or without assuming a probability distribution (Cox model or semi-parametric Cox PH model) for time-to-event process.

In general, parametric models are pivotal in the joint modeling of longitudinal and time-to-event data. A non-parametric or semi-parametric model usually leads to an underestimation of standard errors of the parameter estimates in the joint analysis. However, selection for the joint model framework is quite limited in the literature. The best choice for the selection of longitudinal model can be made based on the observed longitudinal data, and the best survival model can be selected based on the survival data, using standard model selection procedures for these models.

In this thesis, we develop and implement a Bayesian joint model framework, consisting of longitudinal process involving continuous longitudinal outcome and two parametric accelerated failure time (AFT) models (Log-logistic (model 1) and Weibull (model 2)) for survival process. We introduce a link between the parametric AFT survival processes and the longitudinal process via one parameter of association (ϕ) corresponding to shared random effects. A linear mixed-effect model approach is used for the analysis of longitudinal process with the normality assumption of longitudinal response along with normal and independent distribution assumption for both random effects and the error term of the longitudinal process. Finally, Bayesian approach using the Markov chain Monte Carlo method with the Gibbs Sampler technique is adopted for the statistical inference.

The motivating ideas behind our work on Bayesian joint models using parametric AFT event processes are: (a) although there are well-known techniques to test the proportionality assumption for the Cox PH model, checking this assumption for joint modeling has received less attention in the literature. To our knowledge, no statistical package is available to check the PH assumption under the joint modeling setup. AFT models are particularly useful when the PH assumption is in question, (b) there are two integrals

involved in the specification of joint models: (1) a unidimensional integral with respect to time which is relatively straightforward to approximate using numerical techniques, and (2) a multidimensional integral with respect to random effects which is the main computational burden to fit a joint model. It is relatively straightforward to handle (2) under the Bayesian framework, implemented using Markov Chain Monte Carlo (MCMC) techniques, (c) Bayesian approach does not depend on asymptotic approximation for statistical inference and (d) availability of software makes Bayesian implementation for complicated models relatively more straightforward and simple than frequentist methods.

We also develop computational algorithms to fit the proposed Bayesian joint model approach and implemented it in WinBUGS (a Bayesian software) and R software. Analysis are performed with an application to aortic heart valve replacement surgery data (available in joineR package in R software) to illustrate the performance of our two proposed models with the aim of comparing the efficiency of two types of valves based on tissue type (Stentless porcine tissue or Homograft) implanted during surgery and the association between internal covariate (longitudinal response: log.lvmi) and the occurrence of an event (death) after the surgery. Model selection is performed using the deviance information criterion (DIC).

Study analysis results for both joint models indicate the statistically significant and strong association between internal covariate (longitudinal response: log.lvmi) and the relative risk of death after aortic valve replacement surgery. Results show that one gm/m^2 increase in the value of log.lvmi after the surgery reduces the relative risk of death by about 62 % (model 1) and 60 % (model 2), respectively, after controlling for other factors. Moreover, age of the patient (age) and preoperative left ventricular ejection fraction (lv) are found statistically significant for the risk of death after surgery. However, we found no significant difference between the efficiency of two types of valves implanted during surgery based on tissue type (Stentless porcine tissue or Homograft) associated with reducing the risk of death in the patients after surgery. Finally, based on DIC, we recommend, Bayesian joint AFT model with Weibull distribution fits the motivated data set more efficiently than Bayesian joint AFT model with Log-Logistic distribution.

Developing joint models using AFT event processes, writing the model in a hierarchical framework for Bayesian implementation and developing computational algorithms to fit proposed joint models is the novelty of this thesis.

Acknowledgements

First of all, I am, grateful to Almighty Allah, the Most Gracious and The Most Merciful, for the protection and countless blessings that provides energy and passion for continuing my work and empowering me for what I certainly not be able to do.

I would like to express my sincere, deepest gratitude and appreciation to my supervisor Dr. Shahedul A. Khan for his motivation, patience, immense knowledge, consistent support, excellent guidance and providing me with an excellent atmosphere for doing this research. I spend countless hours with Dr. Shahedul A. Khan to discuss my research work. Dr. Khan brought his expertise in Joint modeling approach to open my eyes for this area of research and R and WinBUGS software implementation to help me for the completion of the study results.

I would also like to express my deepest gratitude to my committee member, Dr. Punam Pahwa and committee chair, Dr. Cindy Feng for reading my thesis and valuable suggestions. I would never have been able to finish my dissertation without their guidance. I would also like to acknowledge my class fellows, They were always willing to help and give their best suggestions and support. I am also grateful to the Department of Mathematics and Statistics, and Department of Biostatistics, University of Saskatchewan, for their financial support and all the facilities including software availability.

I express my sincere and deepest gratitude to my all family members for the support – emotional, moral and financial they provided me throughout my life. In particular, I would like to thank my mother, my father (late), my sisters and my brothers. They were always prayed for me, supporting and encouraging me with their best wishes throughout the time of my research and this thesis would certainly not exist without them.

Last but not least, I would like to thank my husband who was important to the successful completion of this thesis. He was always there cheering me up and stood by me through the good and bad time.

I would like to dedicate this thesis to my dearest, strong & lovely mother (Parveen Akhter), to memory of my dearest & kind father (Basharat Ali), who have offered me unconditional love and support throughout my academic career. There is no doubt that without their continuous pray, support, encouragement, motivation and guidance, I could not be able to complete this thesis and all the achievements in my life.

Contents

List of Tables

List of Figures

List of Abbreviations

JM	Joint Model
AFT	Accelerated Failure Time
PH	Proportional Hazard
LL	Log-Logistic
WHO	World Health Organization
LVMI	Left Ventricular Mass Index
LVM	Left Ventricular Mass
AVR	Aortic Valve Replacement
LME	Linear Mixed Effect
nlme	Non Linear Mixed Effect
EM	Expectation Maximization
MLE	Maximun Likelihood Estimator
MCMC	Markov Chain Monte Carlo
BIC	Bayesian Information Criterion
AIC	Akaike Information Criterion
WAIC	Watanabe-Akaike Information Criterion
DIC	Deviance Information Criteria

1 Introduction

In many areas of study including biological, environmental, medical and health applications, a common problem is to investigate the effects of a longitudinal response on time to the occurrence of an event. That is, the longitudinal response is considered to be a time-dependent covariate of the event process. Since data are collected only intermittently in a follow-up study, missing observations and measurement errors are commonly seen in longitudinal data (Rizopoulos 2012). Therefore, analysis of longitudinal data requires special attention to take into account measurement error and missing data. The problem becomes even more challenging when the objective is to model an event process along with a longitudinal response in order to explore the effects of the longitudinal response on time to the occurrence of the event. For example, CD4 cell levels may be recorded longitudinally along with an observation of the onset of AIDS (Abrams et al. 1994), and we might be interested to investigate the effects of CD4 cell counts on time to the onset of AIDS.

The modern approach to handle an analysis when both longitudinal and survival responses are collected is to jointly model the longitudinal response and the time-to-event outcome through shared random effects (Faucett and Thomas 1996, Wulfsohn and Tsiatis 1997, Hogan and Liaired 1997, Henderson et al. 2000, and Yu et al. 2004). Joint modeling of longitudinal and time-to-event processes is the central theme of this thesis.
To summarize, joint modeling is a collection of statistical methods to analyze longitudinal and survival data simultaneously when (Murawska 2014)

1. there is an event process, where the event time distribution depends on a longitudinally measured internal covariate or longitudinal response;
2. the longitudinal data are subject to measurement error and missing observations; and
3. the objective is to understand the effect of the longitudinal response on time to the occurrence of the event.

In general, joint modeling is a collection of statistical methods to properly handle a longitudinal response while investigating its effects on time to the occurrence of an event. (Wulfsohn and Tsiatis 1997, Tsiatis et al. 1995). Joint modeling also allows an investigation of the effects of baseline covariates on both the longitudinal response and the event process. An excellent review of the joint modeling approach is given in Tsiatis and Davidian (2004).

Two general classes of models are in common use for regression analysis of time-to-event data: accelerated failure time (AFT) models where the covariates act multiplicatively on time (i.e., the effect of covariates is to decelerate or accelerate the time to the occurrence of the event), and proportional hazards (PH) models where the covariates act multiplicatively on the hazard function (Cox 1972). Parametric PH models are commonly considered to describe the time-to-event process of joint modeling. Note that the use of the semi-parametric Cox PH model (Cox 1972) in joint analysis leads to an underestimation of the standard errors of the parameter estimates (Hsieh et al. 2006, Rizopoulos 2012), and therefore most methods used for joint modeling are based on parametric response distributions (Hwang and Pennell 2014). The key assumption in the PH model is the proportionality assumption concerning the effects of covariates. Although there are well-known techniques to test the proportionality assumption for the Cox PH model (Kleinbaum and Klein 2012), checking this assumption for joint modeling has received less attention in the literature.

An alternative framework involves considering AFT models, which is particularly useful when the proportionality assumption fails. For joint modeling, Tseng et al. (2005) described the maximum likelihood approach under the AFT assumption, and highlighted some challenges of modeling the baseline hazard function in a likelihood setting. To circumvent this problem, they only considered a piecewise constant baseline hazard function. As far as computational resources are concerned, the JM package (Rizopoulos 2010) of the statistical software R (R Core Team 2018) has been developed to fit joint models based on the PH assumption of the event process. In this thesis, we propose joint model methodology for AFT models, and develop a Bayesian framework for statistical inference.

1.1 Motivating Example: Aortic Valve Replacement and Heart Functioning

Prevalence of heart diseases is increasing worldwide and is the leading cause of death. In 2015, the deaths caused by heart diseases were 45 % of all non-transmissible disease deaths, suggesting approximately 8.9 million people died from heart diseases worldwide (Wang et al. 2016). Heart disease is the second leading cause of death. In Canada after cancer, heart diseases accounted for almost 20 % of all other causes of deaths (Lix et al. 2018).

Heart valve disease is a common heart condition which occurs if one or more of the heart valves do not work well. Globally, the burden of heart valve disease is increasing. The main reason for this increase in heart valve disease is linked with the ageing of the world population and the failure to eradicate rheumatic heart disease in the developing world (Carapetis et al. 2005). The burden of heart valve disease increased by 13 % in the population with age $\geq$ 75 years old (Nkomo et al. 2006). Untreated valve disease leads to premature death, whereas valve surgery may prolong life (Nishimura et al. 2014).

Tricuspid, pulmonary, mitral, and aortic valves are the four types of valves of the heart. Each valve has tissue flaps, and with the heartbeat, these tissue flaps open and close. The function of flaps is to make sure blood flows in the right direction through the four chambers and moves to the rest of the body, and these flaps are called valve leaflets ($https://my.clevelandclinic.org/health/articles/17067-heart-valves$).

The aortic valve is a one-way valve, and the function of the aortic valve is to prevent the flow of blood back into the heart. When the aortic valve narrows, valve leaflets lose controlling its ability to open normally; pumping blood into the aortic becomes harder. Aorta is the largest and main artery in the body. The aorta transmits blood away from the heart to the rest of the body. Stiffening of the valve leaflets causes aortic stenosis. Aortic stenosis usually develops among individuals aged $\geq$ 65 years old, but can also develop in younger people born with an abnormal valve or develop rheumatic heart disease ($https://www.nhlbi.nih.gov/health-topics/heart-valve-disease$).

Currently, there are no medicines, which can cure heart valve disease. However, medicines and change in lifestyle can treat symptoms and delay problems for many years. Ultimately, surgery is done to repair or replace a damaged heart valve. If the heart valve cannot be repaired, the faulty valve is removed and replaced during surgery ($https://www.nhlbi.nih.gov/health-topics/heart-valve-disease$). The most common heart valve surgery performed is the aortic valve replacement (AVR) surgery (Bridgewater et al. 2011). In western countries, the most common form of valvular heart disease is the aortic stenosis, and the standard treatment is aortic valve replacement (AVR). Annually in the United States alone, up to 85,000 AVR surgeries are performed (Nishimura et al. 2014).

Usually, two types of valves are used for aortic valve replacement (AVR) surgery: biological and mechanical. Mechanical valves come in three main types caged ball, tilting-disc and bileaflet with various modifications on these designs (Gott and Alejo et al. 2003). Now a days caged ball valves are no longer implanted, but many patients are still living with this type of valve (Pibarot and Dumesnil 2009). Bileaflet valves are the most frequently implanted mechanical valves in the patients today (Bloomfield 2002). On the other hand, biological valves are made from pig, cow, or human heart tissues. Biological valves can be of numerous types such as bovine, pericardial or porcine valves, stented or stentless, homografts, and pulmonary autografts (Chan et al. 2011). Most biological valves are attached on a stent for ease of implantation. However, stentless valves may have superior hemodynamic with lower gradients and larger effective opening areas. Conflicting evidence exists in the literature and it is uncertain about the preference of stentless valves over stented valves (Gulbins and Reichenspurner 2009).

Another type of biological valve is homograft valve, made from tissues of a human donor. The age of

a patient is of significant attention in the choice of the type of valve: older patients generally being offered biological valves and younger patients being offered mechanical valves. More specifically, biological valves are commonly used when the age of the patient is between 60 and 65 years. Biological valves may also be offered to younger patients who have multiple co-morbidities and if the life expectancy is lower than the assumed durability of the biological valve (Vahanian et al. 2007).

Nowadays, homografts valves are not used frequently. There is conflicting evidence in the literature about its functionality. A recent study (El-Hamamsy et al. 2010) reported a significantly higher rate of valve dysfunction at eight years in homografts compared to stentless valves (63 % versus 15 %, $p < 0.001$), and also indicate the higher rate of re-operation (10 % versus 0 % , $p = 0.024$). Evidence from another study suggested 16 % and 12 % mortality with stentless and homograft valves, respectively (Henryk et al. 2003).

Lim et al. (2008) described a longitudinal study on detecting effects of different heart valves, differing on type of tissue, implanted in the aortic position. The data consisted of longitudinal measurements on left ventricular mass index (lvmi) for each patient after surgery. The patients were followed prospectively after surgery and time to death was recorded, resulting in survival analysis data. Several baseline covariates were also recorded, including age, sex, preoperative left ventricular ejection fraction and implanted aortic prosthesis (homograft or stentless porcine tissue). Here we are interested in the association between lvmi and the risk for death. In addition, the effects of the baseline covariates on the survival outcome as well as on lvmi are also of interest. This is a setup of the joint modeling problem, with death is the event of interest and lvmi is the longitudinal response (time-dependent covariate for the event process). The dataset is publicly available in the joineR package (Philipson et al. 2018) of R.

1.2 Statistical Background

The inspiration of biostatistical research arises from clinical and biomedical studies. The data collected from these studies have always been getting attention due to their particular features that need special consideration when doing an analysis. New statistical methods have developed over time to handle an analysis of such data coming from these sources. A typical clinical study often involves collecting repeated measurements on a biomarker (e.g., lvmi measurements) along with an observation of the time to the occurrence of an event (e.g., death), resulting in a joint modeling setup.

Follow-up studies are common in clinical trails, where measurements are taken repeatedly on the same subject (e.g., human , animal, electronic component) at different points over time. Data resulting from such follow-up studies are called longitudinal data. Analysis of longitudinal data is an active research area,

particularly due to various details not part of standard data analysis, including participant dropout from longitudinal studies and inherent correlation between repeated measures from the same participant (Rizopoulos 2012). Longitudinal measurements can be either continuous or discrete (Salkind and Rasmussen 2008). In this thesis, the focus is on continuous repeated measurements. Many well-established methods are available in statistical literature to analyze the longitudinal data. For example, linear mixed-effects (LME) models (Laird and Ware 1982) and marginal and transitional models (Liang and Zeger 1986) are highly efficient in describing data with ranging characteristics.

Another common feature of a follow-up study involves understanding the distribution of time to the occurrence of an event. Data collected on times to the occurrence of an event are commonly known as survival or time-to-event data (Lawless 2011). Over the last few decades, survival analysis continues to be an active research area in statistical field. A typical survival analysis involves regression methodology, exploring the effects of certain covariates on times to the occurrence of the event (Collett 2015). The event of interest could be death, disease, start time of smoking etc. Analysis of time-to-event data also provides valuable information on, e.g., the effectiveness of a treatment for a specific type of disease and the covariates that are significantly associated with the recovery of a patient.

Time-to-event data have two key characteristics: censoring and truncation. These two characteristics make survival analysis different from standard statistical analysis. In practice survival data may contain either complete or incomplete information on times to the occurrence of the event. If the exact time to the occurrence of the event is observed over the follow-up period, then we have complete information. On the other hand, if only partial information is available in that we only know an individual has survived to a certain time point (exact time to failure is not known), then this leads to incomplete or censored observation. In this thesis, we only consider right censoring, which occurs when (Kleinbaum and Klein 2012):

1. the event is not experienced by the end of the study,
2. lost to follow-up during the study, and
3. withdraw from the study.

Moreover, there are two types of censoring mechanism: informative and non-informative censoring (Rizopoulos 2012). Informative censoring occurs when the probability of censoring is related to the expected failure time, whereas non-informative censoring occurs when the probability of censoring depends on covariates (e.g. gender, age, etc.) unrelated to the event process (Rizopoulos 2012).

Another common feature of time-to-event data is truncation. Truncation occurs when there is a late entry of a participant into the study (Lee and Wang 2003). Due to censoring and truncation, standard statistical methods cannot handle an analysis of survival data. There are many well-established parametric

and non-parametric methods available for survival analysis. Among all these, the most widely used method in epidemiological and medical studies is the Cox proportional hazards (PH) model (Cox 1972).

When dealing, longitudinal and time-to-event data simultaneously, statistical analysis becomes complicated due to the presence of missing data and within-individual correlations in the longitudinal response, and building a model to explore the association between the longitudinal response and the event process. To address these issues, joint analysis of survival with repeated measures is growing rapidly in recent years (Faucett and Thomas 1996, Wulfsohn and Tsiatis 1997, Hogan and Liaired 1997, Henderson et al. 2000, and Yu et al. 2004), and many methods have been proposed for valid and efficient statistical inference (Hatfield et al. 2012).

To analyze longitudinal and time-to-event data separately and in joint modeling framework, standard methods and computer packages are available in the literature (Guo and Carlin, 2004). Early work on joint modeling includes Wulfsohn and Tsiatis (1997), Henderson et al. (2000), Wang and Taylor (2001), and Tsiatis and Davidian (2004). More recent work includes Ye et al. (2008), Rizopoulos (2009), Wu et al. (2010), Alber and Shih (2010), Huang et al. (2011), and Hatfield et al. (2012). Chapter 2 reviews the fundamental framework of joint models.

Some authors proposed extensions of the classical joint modeling framework. A joint model developed by Chi and Ibrahim (2006) considered multivariate longitudinal and survival data. Liu and Huang (2009) and Kim et al. (2012) discussed methods to simultaneously take into account longitudinal and recurrent event data, times of event (death) and time of recurrent events simultaneously. Elashoff et al. (2008) and Rizopoulos (2012) proposed joint models for competing risk problems. Dantan et al. (2011) proposed joint models to explore association between an illness-death model and longitudinal response. He and Luo (2013) considered the longitudinal data with multiple outcomes of different types and outcome-dependent terminal events.

Fitting a joint model is a computationally intensive task as it requires to approximate multiple integrals that do not have an analytic solution except in very special cases. Currently there are several software implementations available to fit joint models. For example, the JM package (Rizopoulos 2010) in R fits joint models for a continuous longitudinal outcome and an event time process using the maximum likelihood method, whereas the joineR package (Philipson et al. 2012) fits joint models following the formulation of Henderson et al. (2000).

Most of the joint models and software packages have been developed based on parametric and semi-parametric PH models (Guo and Carlin 2004, Terrera et al. 2011, and Wulfsohn and Tsiatis 1997). However, if there is a violation of PH assumption, the attractive alternative model choice is the parametric accelerated

failure time (AFT) models (Tseng et al. 2005). In this thesis, we consider the Weibull and log-logistic AFT models, and develop Bayesian methods for statistical analysis and computation.

To summarize, this work on Bayesian joint models using parametric AFT event processes is motivated by the following facts.

1. Although there are well-known techniques to test the proportionality assumption for the Cox PH model, checking this assumption for joint modeling has received less attention in the literature. To our knowledge, no statistical package is available to check the PH assumption under the joint modeling setup. AFT models are particularly useful when the PH assumption is in question.

2. There are two integrals involved in the specification of joint models: (a) a unidimensional integral with respect to time which is relatively straightforward to approximate using numerical techniques, and (b) a multidimensional integral with respect to random effects which is the main computational burden to fit a joint model. It is relatively straightforward to handle (b) under the Bayesian framework, implemented using Markov Chain Monte Carlo (MCMC) techniques.

3. Bayesian approach does not depend on asymptotic approximation for statistical inference.

4. Availability of software makes Bayesian implementation for complicated models relatively more straightforward and simple than frequentist methods.

1.3 Objectives of the Thesis

The background and objectives of this study are described in various places in the above sections. Here we summarize our objectives below.

1. Developing joint models using AFT event processes.

2. Writing the model in a hierarchical framework for Bayesian implementation.

3. Developing computational algorithms to fit joint models, implemented in WinBUGS (a Bayesian software) and R.

4. Analyzing the aortic valve data using our approach.

1.4 Organization of the Thesis

Chapter 2 reviews the fundamental statistical models and methods for longitudinal and survival data, leading to the the joint model framework. Basic concept of Bayesian inference and MCMC are also presented in chapter 2. In chapter 3, we present the formulation of the proposed joint models in a Bayesian framework.

There, we also discuss computational algorithms and software implementation. The analysis of aortic valve data is presented in chapter 4. In chapter 5, we summarize our findings, highlighting some limitations of our work and presenting scope for future research.

2 Joint model: General Overview

Joint models consist of two basic sub-models: longitudinal and time to event sub-models. In this chapter, we will discuss the fundamental concepts and standard formulation of the joint model and Bayesian inference, which leads to the foundation of this thesis. We start with the basic concepts and methods of longitudinal data analysis for a continuous response variable (section 2.1). Then, we review the essential elements of time-to-event (section 2.2) data analysis, discuss non-parametric methods (section 2.2.1), parametric methods (section 2.2.2) and PH and AFT regression models in section 2.2.3. Finally, we review existing standard joint longitudinal-survival model, methods, and estimation processes (section 2.3), and also discuss Bayesian Inference (section 2.4), and Markov chain Monte Carlo (MCMC) (section 2.5).

2.1 Longitudinal Sub-model

Longitudinal data consist of the measurements on response variable taken from same individual over several observation times and these measurements can be continuous or discrete (e.g., binary) (Salkind and Rasmussen 2008). In this thesis, we only consider continuous outcome variable. Longitudinal data can be balanced or unbalanced. For balanced data, all individuals are measured at a common set of occasions, whereas for unbalance data, numbers of measurements and measurement times vary across individuals (Colosimo et al. 2012 and Fitzmaurice et al. 2008). Our proposed joint modeling framework can handle both balanced and unbalanced longitudinal data.

Longitudinal data analysis involves various details not part of standard data analysis, including participant dropout from longitudinal studies and the inherent correlation (Wu 2009 and Viviani 2012) between repeated measures from the same participant (commonly known as a within-subject variation). In this thesis, we consider only linear models to describe a longitudinal process. The following section gives an overview of the linear mixed-effects model (LME) for longitudinal data.

2.1.1 Linear Mixed-Effects Model

In statistical literature, many well-established methods are available for longitudinal data analysis, including linear mixed-effects (LME) models (Laird and Ware 1982) and marginal and transitional models based on generalized estimating equations (GEE) approach (Liang and Zeger 1986). A linear mixed-effects model is

different from a standard regression model in that it takes into account within-subject correlation, between-subject correlation, random effects and provides a valid statistical inference for longitudinal data. Linear mixed-effects models can describe longitudinal data with ranging characteristics. For example, LME models allows characterization and comparison of changes in the response of interest over time, accommodation of incomplete data, handle both balanced and unbalanced data and model the covariance in a parsimonious way (Fitzmaurice et al. 2008). The formulation of the model is simple, and the application of the maximum likelihood method for statistical inference is straightforward.

Let n randomly selected and independent subjects participated in the study, and m_i denote the number of observations for individual i, $i = 1, 2, \ldots, n$. Let y_i be the $m_i \times 1$ vector of responses for individual i. The linear mixed-effects model can be written as (Laird and Ware 1982)

$$y_i = \mathbf{x_i}\boldsymbol{\alpha} + \mathbf{z_i}\mathbf{b_i} + \epsilon_i, \tag{2.1}$$

where $\mathbf{x_i}$ and $\mathbf{z_i}$ are known design matrices, $\boldsymbol{\alpha}$ is a vector of fixed-effects parameters, $\mathbf{b_i}$ is a vector of random effects, and ϵ_i is a vector of measurement errors. We assume that $\mathbf{b_i} \sim N(0, \Sigma)$ (Verbeke 1997), $\epsilon_i \sim N(0, \sigma^2 I_i)$ (McCrink et al. 2013), and $\mathbf{b_i}$ and ϵ_i are independent, where Σ is the covariance matrix for $\mathbf{b_i}$ (Verbeke 1997), σ^2 is the variance for each of the components of ϵ_i, and I_i is an identity matrix of order $m_i \times m_i$. The marginal density function of y_i is (Verbeke 1997)

$$f(y_i) = \int f(y_i|\mathbf{b_i}) f(\mathbf{b_i}) d\mathbf{b_i}. \tag{2.2}$$

where $y_i|\mathbf{b_i} \sim N(\mathbf{x_i}\boldsymbol{\alpha} + \mathbf{z_i}\mathbf{b_i}, \sigma^2 I_i)$ and $f(\mathbf{b_i})$ is the density functions of $\mathbf{b_i}$. Estimation of the parameters can be obtained using the expectation-maximization (EM) algorithm (Dempster and Laird 1977, Laird and Ware 1982) or Newton-Raphson method (Jennrich and Schluchter, 1986).

2.2 Survival Analysis

In this section, we discuss the basic concepts of survival analysis.

Important Quantities of Survival Analysis

Let T be a non-negative random variable representing survival times. Lifetimes distributions can be characterized by any of the following five quantities (for details, see Lawless 2011).

1. Cumulative Distribution Function (CDF)

The cumulative distribution function is the probability that the event will occur before time t.

$$F(t) = Pr(T \leq t). \tag{2.3}$$

F(t) can take the value from 0 to 1 and is a monotone increasing function of t.

2. Survival Function

Survival function, denoted by S(t), is the probability that a subject survives longer than time t. The survival function can be written as

$$S(t) = Pr(T > t). \tag{2.4}$$

S(t) is a monotone decreasing function, with S(0) = 1 and S(∞) = 0. Note that

$$S(t) = 1 - F(t). \tag{2.5}$$

3. Probability Density Function (pdf)

The probability density function is the derivative of the cumulative distribution function (2.3):

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}. \tag{2.6}$$

4. Hazard rate or Hazard Function

Hazard rate, denoted by h(t), is the chance that a subject who has survived up to time t experiences the event of interest at the next instant in time. In other words, hazard function is the instantaneous probability of an event to occur within a small time change $t + \Delta t$ given that event has not occurred until time t:

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr[t \leq T \leq t + \Delta t | T \geq t]}{\Delta t}. \tag{2.7}$$

The hazard function can also be expressed in terms of $f(t)$ and $S(t)$:

$$h(t) = \frac{f(t)}{S(t)}. \tag{2.8}$$

5. Cumulative Hazard Function

Cumulative hazard function H(t) is simply an integration over hazard function from zero to time t and written as

$$H(t) = \int_0^t h(u)du. \tag{2.9}$$

The functions f(t), S(t) and h(t) are mathematically equivalent specifications of the distribution of T. It is easy to see that (Lawless, 2011)

$$h(t) = -\frac{d}{dt}\log[S(t)], \tag{2.10}$$

$$S(t) = \exp(-\int_0^t h(s)ds), \tag{2.11}$$

$$f(t) = h(t)S(t) = h(t)\exp(-\int_0^t h(s)ds), \tag{2.12}$$

and

$$H(t) = -logS(t). \tag{2.13}$$

2.2.1 Non-Parametric Methods of Survival Analysis

In this section, we present two non-parametric methods to estimate survival function and cumulative hazard function.

Kaplan-Meier Estimate of Survival Function

Kaplan and Meier (1958) proposed a non-parametric method to estimate survival probabilities. Let there be p distinct ordered lifetimes $t_{(1)} < t_{(2)} < \ldots < t_{(p)}$. Also, let n_{t_i} be the number of individuals at risk at time $t_{(i)}$, and let d_{t_i} be the number of individuals died at $t_{(i)}$. The Kaplan-Meier (KM) estimate of the survival function is given by (for details, see Lawless (2011))

$$\widehat{S(t)} = \prod_{t_i \leq t}\left(1 - \frac{d_{t_i}}{n_{t_i}}\right), \;\; t \geq 0. \tag{2.14}$$

Nelson-Aalen Estimate Cumulative Hazard Function

The cumulative hazard function for the Nelson-Aalen estimate is given by

$$\widehat{H(t)} = \Sigma_{i|t_i \leq t}\left(\frac{d_{t_i}}{n_{t_i}}\right), \;\; t \geq 0 \tag{2.15}$$

Note that cumulative hazard plots are the diagnostic tool and are used to check the assumption of a parametric model. For further details about the estimation of H(t) and cumulative hazard plots, see Nelson (1972) and Aalen (1978).

2.2.2 Parametric Distributions for Survival Analysis

For a positive-valued random variable T, Table 2.1 displays the probability density functions and survival functions for some of the widely used distributions in survival analysis (Lawless 2011). See Appendix A for details about the characteristics of these distributions. As we will see in the subsequent sections, these distributions play a very important role in formulating regression models for survival data analysis.

Table 2.1: Most widely used Parametric distributions for survival analysis

Distribution	Parameters	Probability Density Function: f(t)	Survival Function S(t)
Exponential	$\rho > 0$	$\rho\ \exp(-\rho t)$	$\exp(-\rho t)$
Weibull	$\rho, \kappa > 0$	$\kappa\ \rho\ (\rho t)^{(\kappa-1)}\ \exp\ [-(\rho t)^{\kappa}]$	$\exp\ [-(\rho t)^{\kappa}]$
Log-Logistic	$\rho, \kappa > 0$	$\kappa\ \rho\ (\rho t)^{(\kappa-1)}[1 +\ (\rho t)^{\kappa}]^{-2}$	$[1 +\ (\rho t)^{\kappa}]^{-1}$
Logistic	$\mu \varepsilon \mathbb{R}, \sigma > 0$	$\sigma^{-1} \exp\ [(t-\mu)\sigma^{-1}]$ $[1 + \exp(\ (t-\mu)\sigma^{-1})]^{-2}$	$[1 + \exp(\ (t-\mu)\sigma^{-1})]^{-1}$
Log-normal	$\mu \varepsilon \mathbb{R}, \sigma > 0$	$[t\ \sigma\sqrt{2\pi}]^{-1} \exp[-(\log(t)-\mu)^2(\sqrt{2}\sigma)^{-2}]$	$1 - \Phi([\log(t)][\sigma]^{-1})$
Gamma	$\rho, \kappa > 0$	$[\rho^{\kappa} t^{\kappa-1} \exp(-\rho t)][\Gamma(\kappa)]^{-1}$	$1 - I_k(\rho t)$
Gompertz	$a > 0$	$t \exp[ta + ta^{-1}(1 - \exp(ta))]$	$\exp[ta^{-1}(1 - \exp(ta))]$

2.2.3 Regression Models For Survival Analysis

Regression models are used to quantify the effects of covariates (time-independent or time-dependent) on the response variable. In the analysis of survival data, accelerated failure time (AFT) and proportional hazards (PH) are the two most popular families for regression analysis (Lawless 2011). The Cox proportional hazards model (Cox 1972) is a semi-parametric PH model and is appealing due to its robustness property against the distributional assumption and relative risk interpretation of the parameters. An alternative modeling framework when the PH assumption fails is the AFT family (Collett 2015). In the following sections, we briefly present the formulations of these two families for regression analysis (for more details, see Lawless 2011).

Proportional Hazards (PH) Model

The key feature of the PH model is the proportionality assumption. According to proportional hazards assumption: the effect of the covariate is to increase or decrease the hazard by a proportionate amount which does not depend on t ((Lawless 2011). In other words, a covariate has multiplicative effects on the hazard in a PH model. The PH model can be conveniently expressed using the hazard function as follows

(Cox 1972):

$$h(t|\mathbf{x}) = h_0(t)\exp[\mathbf{x}'\boldsymbol{\beta}]. \tag{2.16}$$

where $h_0(t)$ is the baseline hazard function , $\mathbf{x}$ is a vector of covariates, and $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients. $h_0(t)$ may have a specified parametric form or it may be left as an arbitrary nonnegative function. The semi-parametric Cox PH model assumes an arbitrary (unspecified) nonnegative function for $h_0(t)$, whereas parametric PH models ((Lawless 2011) assume a parametric form for $h_0(t)$ such as Weibull or exponential.

Note that the assumption of PH is strong and it is important that it be checked. The AFT family is an attractive alternative to the PH models when this assumption fails (Wei, 1992).

Accelerated Failure Time Model

Accelerated failure time (AFT) models can be used to predict times to failure. Proportional hazards models the effect of predictors on the hazard function, whereas the AFT models assume a direct relationship between survival times and covariates (George 2014). To clarify, consider a predictor signifying the presence or absence of a disease. The AFT model assumes that the disease either accelerates or decelerates the rate of survival of a patient. In other words, if $S_1(t)$ and $S_2(t)$ denote the survival functions for the presence and absence of the disease, respectively, then the AFT model assumes the relationship $S_1(t) = S_2(\eta t)$, where η is the acceleration factor. If $\eta > 1$, then a diseased individual has longer median survival time compared to a non-diseased individual, whereas if $\eta < 1$, then a diseased individual has shorter median survival time compared to a non-diseased individual (Collett 2015).

The AFT model can be expressed in terms of log survival time as (George 2014)

$$\log T = -\mathbf{x}'\boldsymbol{\beta} + \epsilon. \tag{2.17}$$

If T has a Weibull, log-logistic or log-normal distribution, then $\log T$ can be expressed in the form equation (2.17), with ϵ having a standard extreme value distribution, a standard logistic distribution and a standard normal distribution, respectively. Thus, Weibull, log-logistic and log-normal distributions belong to the AFT family.

For time-independent covariates, the probability distribution function, survival function, and hazard function of T can be expressed as (Kalbfleisch and Ross 2002)

$$f(t|\mathbf{x}) = f_0(t\ \exp(-\mathbf{x}'\boldsymbol{\beta}))\exp(-\mathbf{x}'\boldsymbol{\beta}), \tag{2.18}$$

$$S(t||\mathbf{x}) = S_0(t\ \exp[\mathbf{x}'\boldsymbol{\beta}]), \tag{2.19}$$

$$h(t|\mathbf{x}) = h_0(t\ \exp(-\mathbf{x}'\boldsymbol{\beta}))\exp(-\mathbf{x}'\boldsymbol{\beta}), \tag{2.20}$$

where $f_0(\cdot)$, $S_0(\cdot)$ and $h_0(\cdot)$ are the baseline probability density function, baseline survival function and baseline hazard function, respectively. For time-dependent covariates $\mathbf{x}$. we define:

$$\psi(t) = \int_0^t \exp(-\mathbf{x}'(s)\boldsymbol{\beta})ds, \tag{2.21}$$

which can be considered as a transformed time scale defined by the covariate process. With this transformation, the probability distribution function, survival function, and hazard function for AFT models can be expressed as (Cox and Oakes, 1984)

$$f(t|\mathbf{x}(t)) = f_0(\psi(t))\ \exp(-\mathbf{x}'(t)\boldsymbol{\beta}), \tag{2.22}$$

$$S(t|\mathbf{x}(t)) = S_0(\psi(t)), \tag{2.23}$$

$$h(t|\mathbf{x}(t)) = h_0(\psi(t))\ \exp(-\mathbf{x}'(t)\boldsymbol{\beta}). \tag{2.24}$$

Note that the Weibull and exponential distributions are the only distributions which are closed under both PH and AFT family. Figure 2.1 shows some widely used regression models for time-to-event data. In this thesis, we will use Weibull and log-logistic AFT models with time-dependent covariates to formulate joint models.

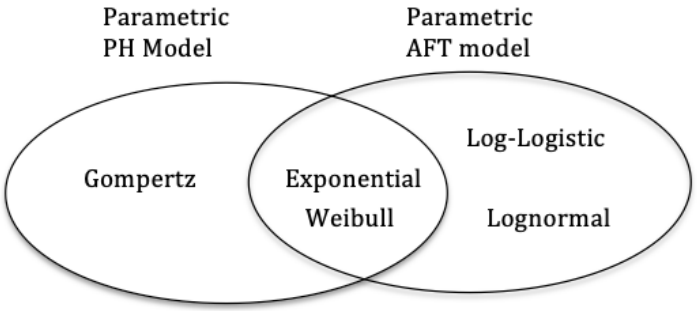


Figure 2.1: Commonly used AFT and PH models for survival data analysis.

2.3 Joint Model (JM)

Recall that the objective of joint modeling is to estimate the effect of a longitudinal response on time to the occurrence of an event. To achieve this goal, the longitudinal response is considered as an internal

time-dependent covariate for the event process under a regression setting. Note that valid inference requires a framework in which the underlying relationship between the event process and the longitudinal response is explicitly acknowledged. Although developing such a framework is conceptually straightforward, the implementation becomes complicated due to the nature of the data observed. Therefore, in order to properly incorporate the internal covariate into the event process for a valid inference, it is desirable to use a comprehensive framework that takes into account measurement errors and missing data (Rizopoulos 2012). In this section, we present the standard approach of joint modeling, where the longitudinal response is linked to the time-to-event process through shared random effects (Henderson et al. 2000).

As mentioned above, joint models consist of two sub-models: a model that takes into account the measurement error and missing data in the time-dependent covariate to estimate its true values (longitudinal model), and another model that uses these estimated values to quantify the association between this covariate and the time to the occurrence of the event (time-to-event model). The motivating idea behind the joint modeling techniques is to couple the time-to-event model with the longitudinal model (Rizopoulos 2012). The standard approach is to consider a linear mixed-effects model for the time-dependent covariate (i.e., the longitudinal response) and a PH model for the association analysis (Guo and Carlin 2004).

The maximum likelihood method is commonly used for statistical inference (McCrink et al. 2013, Schluchter 1992), implemented via Newton-Raphson method or EM algorithm ((Dempster and Laird 1997). Let there be n subjects with lifetimes denoted by $T_1, T_2, \ldots, T_n$. Assuming that the data are subject to right censoring, we observe $t_i = \min(T_i, C_i)$, where $C_i > 0$ corresponds to a potential censoring time for subject i. Letting $\delta_i = \mathrm{I}(T_i \leq C_i)$ that equals 1 if $T_i \leq C_i$ and 0 otherwise, the observed data for individual i consist of $\{t_i, \delta_i\}$, $i = 1, 2, \ldots, n$, where t_i is a lifetime or censoring time according to whether $\delta_i = 1$ or 0, respectively. Also, assume that the i^{th} subject provides a set of longitudinal quantitative measurements $\{y_{ij}, j = 1, 2, \ldots, n_i\}$ at times $\{s_{ij}, j = 1, 2, \ldots, n_i\}$. Then the likelihood contribution for individual i can be written as (Rizopoulos 2012)

$$L_i = \int \prod_{j=1}^{n_i} p(y_{ij}|\mathbf{b_i}, \boldsymbol{\theta}) p(t_i, \delta_i|\mathbf{b_i}, \boldsymbol{\theta}) p(\mathbf{b_i}|\boldsymbol{\theta}) d\mathbf{b_i}, \tag{2.25}$$

where $\boldsymbol{\theta}$ denotes the full parameter vector and $\mathbf{b_i}$ denotes the vector of shared random effects for individual i.

The main difficulty with the maximum likelihood method is that it requires evaluation of multiple integrals: a one-dimensional integral with respect to time for $p(t_i, \delta_i|\mathbf{b_i}, \boldsymbol{\theta})$, and a multi-dimensional integral with respect to the random effects b_i. These integrals do not have an analytic solution in general (Rizopoulos 2012). The integral with respect to time can be approximated using the Gauss-Kronrod rule (Press et al. 2007), but the integrals with respect to the random effects are computationally expensive and complicated, especially when the dimension of the random effect increases (equation 2.25). The numerical approximation

methods such as Gauss Hermit rule (Wulfsohn and Tsiatis 1997, Henderson et al. 2000, Song et al. 2002 and Rizopoulos et al.2008), Laplace approximation (Rizopoulos et al. 2009), and Bayesian estimation approach (Xu and Zeger 2001, Guo and Carlin 2004) are used to solve this issue.

Bayesian approach is also considered to fit joint models (e.g., Faucett and Thomas, 1996; Faucett et al., 2002; Xu and Zeger, 2001; Brown et al., 2005; Guo and Carlin, 2004). Although expensive in terms of computational time, Bayesian approach has some advantages as described below.

- The approximation of the integrals with respect to the random effects is straightforward via MCMC.
- As opposed to the maximum likelihood method, asymptotic approximation is not required for statistical inference.
- With the availability of Bayesian software (e.g., WinBUGS, JAGS), it is relatively easier to implement MCMC for Bayesian inference.

2.4 Bayesian Inference

In statistical inference, for the interpretations of probability, there are two schools of thought: Bayesian inference and frequentist inference (Boldgiv 2004). The difference between these two approaches is based on the nature of probabilities. The most widely used statistical methods are known as frequentist (or classical) methods. These methods treat the unknown parameters as a fixed constant and describe the probability as the limit of relative frequencies of an event. On the other hand, in Bayesian methods, the parameters are treated as random variables (Van et al. 2014).

The term "Bayesian" comes from the well-known Bayes' theorem, named after the Reverend Thomas Bayes (Bayes and Price 1763). He was an English statistician and Presbyterian minister in the eighteenth-century. Bayes was captivated in solving the question, what is the probability of one event? He first used conditional probability to provide an algorithm that uses evidence to calculate the limits on an unknown random parameter.

Generally, Bayesian inference is used as a complement to frequentist inference. The line drawn between the two approaches is ambiguous. The Bayesian approach is different from traditional frequentist approaches. The actual distinction between Bayesian and frequentist is that how someone explains the probability (Bolstad and Curran 2016). In this thesis, we describe the Bayesian concepts based on prior and posterior distributions, authentically used in applied work.

Bayesian inference is a tremendously powerful tool for modelling the random variable, such as the value

of a regression parameter, a demographic statistic, or the part of speech of a word. Bayesian approach is used to handle the following uncertainty situations in the model (Boldgiv 2004)

- Data is limited.
- Model overfitting.
- If the information is not contained in the data, but we want to model the data and we have to believe that some facts are more likely than others.

Bayesian approach is a useful tool to model the data when the parameters of interest are considered as random variables (Van et al. 2014). For example, we are interested in estimating the parameter or vector of parameter $\boldsymbol{\theta}$ from the data y with n number of observations and we want to use the statistical model described by the marginal density $p(y|\boldsymbol{\theta})$ (conditional probability of data y given parameters $\boldsymbol{\theta}$).

The fundamental elements of the Bayesian approach are (Bolstad and Curran 2016):

- Prior distribution; the probability distribution for $\boldsymbol{\theta}$ formulated as π ($\boldsymbol{\theta}$). The prior distribution describes the mean, the spread and the skewness about the parameter before the data is assessed.
- A statistical model p($y|\boldsymbol{\theta}$) called likelihood and it describes the distribution of observed data y given $\boldsymbol{\theta}$.
- Update the information by combining the information from the prior distribution and the posterior distribution (data through the calculation of the p ($\boldsymbol{\theta}|y$) of the parameter or the vector of parameters.

Thus, in the Bayesian paradigm, the main idea is to combine data and prior knowledge on a parameter (or a vector of parameters) and likelihood (p($y|\boldsymbol{\theta}$)) to determine its posterior distribution (the conditional density of the parameter given the data).

2.4.1 Prior Distribution

A prior probability distribution, often known as prior, is the probability distribution of an unknown quantity that would express one's beliefs about this quantity before some evidence is taken into account (Gelman et al. 2013). The uncertain quantity may be a parameter or vector of parameters of the model or a latent variable rather than an observable variable. Parameters of prior distributions are known as hyperparameters (Bolstad and Curran 2016). Many methods can be used to create or choose the priors (Carlin and Louis 2008). For example, a prior can be determined from past information, such as previous experiments or a prior can be predicted from the subjective conclusion/ assessment of an experienced expert. Here we will discuss some widely used priors.

Non-informative Priors

Non-informative or uninformative priors are created to reflect a balance among outcomes when no information is available. A non-informative prior can also be used if the prior distribution is "flat" relative to the likelihood function. These priors have minimal impact on the posterior distribution of $\boldsymbol{\theta}$. For more formal development of non-informative priors, see Box and Tiao (2011).

Improper Priors

A prior $\pi(\boldsymbol{\theta})$ is said to be improper if

$$\int \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = \infty.$$

For example, a uniform prior distribution defined on real line as $\pi(\boldsymbol{\theta}) \propto 1$ for $-\infty < \boldsymbol{\theta} < \infty$ is an improper prior. Improper priors usually generate non-informative priors and proper posterior distributions and are frequently used in Bayesian inference.

Informative Priors

A prior distribution, which is not dominated by the likelihood, is called an informative prior and has an impact on the posterior distribution.

Conjugate Priors

If the prior distribution and posterior distributions belong to the same family of distributions, then this kind of prior is called a conjugate prior. In this case, both posterior and the prior distribution has the same distributional form. A desire for computational convenience partially drove the development of conjugate priors and the conjugacy provides a practical way to obtain posterior distributions (Bolstad and Curran 2016).
In this thesis, we consider minimal informative and conjugate priors.

2.4.2 Posterior Distribution

Using Bayes' theorem, we can compute the posterior probability distribution (conditional distribution of the uncertain quantity given the data) by taking the product of the prior and the likelihood function. Thus, applying Bayes' theorem the posterior distribution of unknown quantity ($\boldsymbol{\theta}$) given data (y) can be defined as (Van et al. 2014)

$$p(\boldsymbol{\theta}|y) = \frac{p(\boldsymbol{\theta}, y)}{p(y)} = \frac{p(y|\boldsymbol{\theta})\pi(\theta)}{p(y)} = \frac{p(y|\theta)\pi(\boldsymbol{\theta})}{\int p(y|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}.$$

where

$$p(y) = \int p(y|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

is known as the normalizing constant of the posterior distribution. Moreover, the likelihood function $L(\boldsymbol{\theta})$ is a function proportional to $p(y|\boldsymbol{\theta})$, i.e., $L(\boldsymbol{\theta}) \propto p(y|\boldsymbol{\theta}a)$. Posterior mean from the sample X_t can be defined as (Zhao et al. 2008)

$$E[f(X)] \approx \frac{1}{n}\sum_{t=1}^{n} f(X_t).$$

Bayesian inference is based on the posterior distribution of the parameters given the data (Van et al. 2014). Because the posterior density is an actual probability density, one way to do this is to sample values from the distribution and then to compute the sample statistics. The posterior density of a parameter describes its behavior over a range of values (the support of the parameter space). We generate a sample of values from the posterior distribution and then use these values to approximate posterior means or median, standard deviations, and 95 % credible intervals, or any other quantity of our interest.

Bayesian Point Estimates

The approximated posterior mean or median is considered as the point estimate of the parameter of interest (Van et al. 2014). When computing a Bayesian point estimate, we want to find a single value that conveys information about the centre of the posterior distribution, i.e. what value of the parameter is most likely in some sense. The two commonly used point estimates for the parameters of the model are; posterior mean and the posterior median (the mean and median of the posterior density). Moreover, these values are easier to compute even in complicated problems (Bolstad and Curran 2016).

Posterior Standard Deviation

The posterior standard deviation in the Bayesian approach is equivalent to the standard error of frequentist methods. As the name implies, it is the standard deviation of the posterior density. However, the interpretation of these two values is quite different. The posterior standard deviation is a direct statement about the uncertainty in the true value of the parameters and the standard error is a statement about the uncertainty of the estimated value of parameters (Bolstad and Curran 2016).

Credible Intervals

In Bayesian inference, the equivalent of the confidence interval is the credible interval. Its main purpose is to describe and summarize the uncertainty related to the parameters. A credible interval is an interval in which an unobserved parameter has a given probability. In confidence intervals, we also treat the parameter as a fixed value, and the bounds are random variables, while in credible intervals, the estimated parameter

is treated as a random variable while the bounds are considered fixed. A credible Bayesian interval for a parameter is defined as $[c_p, c_{1-p}]$ where c_p and c_{1-p} are estimated as the p^{th} and $(1-p)^{th}$ quantiles of the posterior distribution of the parameter, respectively (Van et al. 2014). The posterior density is a true probability density; we can compute quantiles and percentiles of the parameter.

The simplest credible interval is bounded by the 2.5^{th} and 97.5^{th} percentiles known as 95 % credible interval. This interval is also called symmetric credible interval because it removes equal probability (2.5 %) from both tails of the distribution (Bolstad and Curran 2016). We use a 95 % credible interval to check the significance of the estimated parameters of the model including the significance of association parameter ϕ in the joint model. The interpretation of this quantity is that the 95 % credible interval contains the true value of the parameter. In other words, there is a 95 % probability that the true parameter value lies in the credible interval.

We can estimate all the exact posterior estimates if we know the posterior distribution of our interest. The determination of posterior distribution contains the calculation of complex and multi-dimensional integrals. Such problems arise in both frequentist and Bayesian approaches (Geyer and Thompson 1995, Christensen 2004). For example, when computing the normalizing constant and marginal posterior distribution integration for a particular parameter or a vector of parameters of interest from the posterior, models become too difficult to analyze analytically. Then we need simulation algorithms like MCMC. The MCMC method is commonly used to solve the complexed integration problem and to calculate posterior estimates. All of the Bayesian processes rely on MCMC.

2.5 MCMC Methods and Diagnostics

Markov Chain Monte Carlo (MCMC) method is particularly useful in Bayesian inference because of the focus on sampling from posterior distributions, which are often difficult to work with via analytic examination. A Markov chain is a sequence of events whose distribution depends only on the outcome of the previous event. MCMC allows to approximate aspects of posterior distributions that cannot be directly calculated. The essential feature in MCMC methods is, if the simulation algorithm is implemented correctly, the Markov chain has the guarantee to converge to the targeted distribution. In other words, a Markov chain improves its approximation to the accurate underlying distribution at each step in the simulation. The properties of the Markov chain are discussed in detail by Meyn and Tweedie (2012), Feller (1957) and Breiman (1968). Non-measure-theoretic treatment of all stochastic processes including Markov chains was proposed by Pekoz (1997) and Karlin and Taylor (1975).

Bayesian inference is based on Monte Carlo samples (MCMC) drawn from the posterior distribution

using an MCMC algorithm such as the Gibbs sampler (Givens and Hoeting 2005) or Metropolis-Hastings (Hastings 1970). The MCMC method is a typical simulation method for consecutive sampling from posterior distributions and computing posterior quantities of interest from a targeted distribution. Each sample depends on the previous sample. Monte Carlo integration is generally used to approximate the expectation by using the Markov chain samples. There are several techniques to construct Markov chains, including the Metropolis-Hastings algorithm and Gibbs sampling(Givens and Hoeting 2005). The MCMC simulation was first proposed in the literature of physics. Metropolis and Ulam (1949) and Metropolis et al. (1953) explain the well-known Metropolis algorithm and used it to produce sequences of samples from the joint distribution of multiple variables. The generalized work on the Metropolis-Hastings algorithm was done by Hastings (1970), resulting in the Gibbs sampling algorithm.

In our implementation for Bayesian inference, we consider MCMC methods with the Gibbs sampler technique (Givens and Hoeting 2005). From the MCMC method, we create samples from an arbitrary posterior density, then use these samples to approximate expectations of quantities of interest.

Gibbs Sampler

The name of Gibbs sampler algorithm was proposed by Geman and Geman (1984) and they initially applied it to analyze image data. The Gibbs sampling technique was first used by Gelfand et al. (1990) to solve the problem in the Bayesian approach. Under the formation of Gibbs sampling, the parameter vector is decompose into a number of components of possibly differing dimensions, and then each of these components is updated one by one. For a particular component, an instance of a Markov chain is generated from its full conditional distribution and create sample from them.
The Gibbs sampler algorithm works as follows (Zhao et al. 2008):
1- Choose an arbitrary initial value of $\theta^{(0)} = (\theta_1^{(0)}, ..., \theta_k^{(0)}$ at t= 0.
2- Generate each component of θ as follows:
draw $\theta_1^{(t+1)}$ from $\pi(\theta_1|\theta_2^{(t)}, ..., \theta_k^{(t)}, y)$
draw $\theta_2^{(t+1)}$ from $\pi(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, ..., \theta_k^{(t)}, y)$
.....
draw $\theta_k^{(t+1)}$ from $\pi(\theta_k|\theta_1^{(t+1)}, ..., \theta_{k-1}^{(t+1)}, y)$.
3- Set $t = t + 1$ If $t < T$, the number of desired samples, return to step 2. Otherwise, stop.

2.5.1 MCMC Performance

Whenever Markov chains are run, it is important to assess the performance of these chains. It is suggested that at least two chains are run to analyze MCMC output. Important features to assess the performance of Markov chains are describes below.

MCMC Mixing

The efficiency of an MCMC algorithm depends on its mixing ability. The mixing property of a chain refers to two characteristics. First, how quickly a chain forgets its initial values. Second, how quickly a chain can explore the full support and shape of the target distribution (Cowles and Carlin 1996).

Number of Chains

There are conflicting arguments in the literature regarding the selection of the number of chains to be used in the MCMC algorithm. Some scholars (Gelman and Rubin 1992) suggest several long chains, while others (Geyer 1992) recommend only one and long chain. Gilks et al. (1995) and Givens and Hoeting (2005) discussed the advantages and disadvantages of each method. The main idea is that the chain can reach around the mode of the target distribution and can stay there forever, even for a very long chain. In such a case, the convergence diagnostic may indicate the convergence of the chain, although the chain does not fully explore the support and shape of the target distribution. On the other hand, running multiple chains can ensure that at least one of them will explore the features of the target distribution and will wash out the influence of initial values. Usually, in practice, two or more chains are run with the idea that at least one of these chains converge to the targeted distribution and explore all the feature related to that distribution. In order to implement MCMC for our joint models, we construct two Markov chains that necessarily converge to an underlying stationary distribution.

Burn-in and Stopping Time

The dependence of a Markov chain on its starting value may remain stable even after the chain has been run for a long period. As a consequence, if the chosen initial values are far different from the posterior mode, this dependency may make the chain converge slowly (Cowles and Carlin 1996). It may take some iterations for the chains to enter into the high probability region where they are more representative of the target distribution. In practice, initial S iterations are discarded as a burn-in period to make the chain independent of its starting values and converge quickly to the target distribution (Gelman 1996). Typically, a chain should be stopped at a particular time after running the chain for a sufficiently long period to obtain good mixing. In practice, it is difficult to decide about the stopping time. Gilks et al. (1995) suggested an informal way of determining the stopping time to run several long chains and to compare the estimates (posterior means/medians) from each chain. If the estimates produced by different chains do not agree closely, then the run length should be increased.

Thinning

The slow decay of autocorrelation generally exhibits poor mixing of a chain. It is, therefore, a good practice that the inference should be based on every i^{th} iteration of chains, with i set to some value high enough that

successive draws are approximately independent (Gelman 1996). This strategy is known as thinning in the literature.

MCMC Convergence

However, it is often difficult to decide at what point it would be reasonable to believe that the samples are accurately approximate the underlying stationary distribution of the Markov chain. The convergence of a chain is to check how efficiently the chain has approximated to its stationary distribution (Cowles and Carlin 1996). To determine the convergence of the chains, formal test statistics, Gelman-Rubin R statistic (Gelman and Rubin, 1992), is a very popular and useful technique. This statistic is based on the comparison of within-chain and between-chain variances. Values of R substantially above 1 indicate a lack of convergence. Some authors suggest that $R < 1.2$ is acceptable (Brooks and Gelman 1998). We use formal test statistics, the Gelman-Rubin R statistic (Gelman and Rubin 1992) technique to check the convergence of the two Markov chains.

Graphical Tools

Convergence and mixing property of a chain can also be examined through three widely used graphical tools: trace plot, autocorrelation plot and density plot (Gelman 1996). Trace plots show how rapidly the chain is mixing. Trace plot is the realization of the chains versus the iteration number. A well-behaved chain will move away from its starting values quickly, no matter where it started and the samples will wiggle about vigorously in the supported region by the posterior density indicating no specific trend seen between two chains (Givens and Hoeting, 2005). If a clear trend is seen in the trace plot, suggests that there is a lack of good mixing in chains and stationary distribution has not achieved.

An autocorrelation plot shows the serial correlation in the chain at different lags of iteration (Givens and Hoeting, 2005). In general, the autocorrelation decreases as the lag increases. If the situation is different, thinning should be explored. A density plot is a smoothed histogram of the MCMC samples used to approximate the posterior density. Smooth kernel density plots indicate the samples are accurately approximate the underlying stationary distribution of the Markov chain (Gilks et al. 1995).

In the presence of high autocorrelation, convergence is not guaranteed and indicating that the distribution is multimodal (Cowles and Carlin 1996). We can see that multimodality by kernel density plots with multiple modes and lumpiness rather than a smooth curve. It is to be noted that such behaviour in the density plots may result due to high autocorrelation within a converged chain. In such cases, severity can be reduced by running the chain for a longer time or with heavier thinning (Gelman 1996). We use trace plots and density plots to check the good mixing and convergence of both chains in our analysis.

2.5.2 Model Selection Criterion

The most common way to compare the fit of two different models is the likelihood method, compare the values of the likelihood evaluated at the parameter estimates. The model with the larger likelihood fits the data better than the model with a smaller likelihood value. However, the likelihood value always increases when more parameters are added to the model. Therefore, considering the likelihood alone for the selection of the model is not appropriate, it always selects the model with the most parameters. Instead, we want to choose the model according to the principle of parsimony, which provides the best fit to the data with the less number of parameters.

There are many model selection criteria in Bayesian analysis, derived based on a variety of principles. These include deviance information criterion (DIC) (Spiegelhalter et al. 2003), BIC or the Schwarz criterion (Schwarz et al. 1978), and Watanabe-Akaike information criterion (WAIC) (Watanabe 2010). DIC is the most widely used tool for model comparison in Bayesian analysis (DIC, Spiegelhalter et al. 2003). It does not require maximization over the parameter space, like the AIC and BIC. A smaller DIC indicates a better fit to the data set.

DIC is one of the ways to select the most parsimonious model. It creates a balance between the likelihood and the number of parameters in the model. It is computed from (a) the deviance which is obtained by the minus twice of the value of the log of the likelihood; deviance plays a vital role in classical statistics (b) then added an estimate of the number of unique and estimable parameters in the model, called the effective number of parameters. DIC involves posterior mean that takes into account prior information and penalized likelihood. DIC can be computed as (Spiegelhalter et al. 2003)

$$DIC = E[D(\theta|data)] + p_D. \tag{2.26}$$

The first part of left hand side of the equation (2.26) presents the estimated value of Deviance statistics $= -2\,log\,(likelihood(\theta|data))$ and is used to measure the goodness of fit for the model and the second part presents effective number of parameters (p_D) . The DIC is a Bayesian alternative to the other two widely used criteria; Akaike's information criterion (AIC) and Bayesian information criterion (BIC) (Robert 2007 and Spiegelhalter et al. 2003).

Since deviance has the opposite sign to the likelihood. Therefore, the best fitting model will be the one with the lowest DIC (i.e., the highest penalized likelihood). In practice, the value of the DIC varies for a given model because of error in the MCMC sampling and noise in the data. If the difference in DIC values between the two models is > 10, then the model with a smaller value of DIC is preferred (Raftery 1995). In general, if the difference in the values of DIC of two models is five or more than five (preferably is greater

than 10) (*http://users.jyu.fi/ ∼ hemipu/itms/DIC%20web%20site%20from%20BUGS%20project.pdf*), then it presents clear evidence in favour of the model with lower DIC value on the other model with high DIC value. Note that if differences between DIC values of two model is less than five, then it provides weaker evidence to choose one model over the other model based on low DIC value assumption. Smaller differences may be the result of random variation and should not be considered as a support for one model over another.

Furthermore, DIC has several applicable properties, including that it can be calculated when non-informative or improper priors are used. Moreover, DIC can be easily computed using MCMC simulation implemented in the WinBUGS software. In this thesis, we consider DIC as a criterion for the comparison of our suggested models.

3 Joint Models of Accelerated Failure time and Longitudinal Data

Joint models and joint model analysis using maximum likelihood estimation (MLE) method were discussed in Chapter 2. In this chapter, we will present the formulation of our proposed joint models of longitudinal and time-to-event data using the log-logistic and Weibull AFT models. We adopt the approach of Hederson et al. (2000) to formulate the model, where the two processes are linked through stochastic dependence (Gamerman and Lopes 2006). A Bayesian approach is considered for statistical inference, implemented in the statistical software WinBUGS (Lunn et al. 2000).

3.1 Notation

For completeness of this chapter, we re-introduce our notation in this section. Let there be n subjects with lifetimes denoted by $T_1, T_2, \ldots, T_n$. Assuming that the data are subject to right censoring, we observe $t_i = \min(T_i, C_i)$, where $C_i > 0$ corresponds to a potential censoring time for subject i. Letting $\delta_i = \mathrm{I}(T_i \leq C_i)$ that equals 1 if $T_i \leq C_i$ and 0 otherwise, the observed data for individual i consist of $\{t_i, \delta_i, z_i\}$, $i = 1, 2, \ldots, n$, where t_i is a lifetime or censoring time according to whether $\delta_i = 1$ or 0, respectively, and z_i is a $p \times 1$ vector of covariates. Also, assume that the i^{th} subject provides a set of longitudinal quantitative measurements $\{y_{ij}, j = 1, 2, \ldots, n_i\}$ at times $\{s_{ij}, j = 1, 2, \ldots, n_i\}$, and $\boldsymbol{\theta}$ denotes the full parameter vector.

3.2 Longitudinal Sub-model

We model the longitudinal response y_{ij} at time s_{ij} by the relationship (Hederson et al. 2000)

$$y_{ij} = \mu_i(s_{ij}) + U_i(s_{ij}) + \epsilon_{ij}, \tag{3.1}$$

where $\mu_i(s_{ij})$ is the mean response, $U_i(s_{ij})$ incorporates subject-specific random effects, and $\epsilon_{ij} \sim N(0, \sigma^2)$ is a sequence of mutually independent measurement errors. We assume that the mean response at time s is characterized by a linear model

$$\mu_i(s) = \mathbf{x}_i'(s)\boldsymbol{\alpha}, \tag{3.2}$$

where $\mathbf{x}_i(s)$ is a vector of covariates (possibly time-dependent) and $\boldsymbol{\alpha}$ is the corresponding vector of regression coefficients (fixed effects). For $U_i(s)$, we assume a linear random effects model

$$U_i(s) = \mathbf{w}_i'(s)\mathbf{b}_i, \tag{3.3}$$

where $\mathbf{w}_i(s)$ is the design vector for the random effects $\mathbf{b}_i \sim N(\mathbf{0}, \Sigma_b)$.

3.3 Survival Sub-model

We consider an AFT model to describe the event hazard process at time t (Cox and Oakes, 1984), expressed as

$$h_i(t) = h_0(g_i(t)) \exp(-\mathbf{z}_i'\boldsymbol{\beta} - V_i(t)), \tag{3.4}$$

where $h_0(\cdot)$ is the baseline hazard function, $g_i(t) = \int_0^t \exp(-\mathbf{z}_i'\boldsymbol{\beta} - V_i(u))du$, $\mathbf{z}_i$ is a vector of baseline covariates with a corresponding vector of regression coefficients $\boldsymbol{\beta}$, and $V_i(t)$ is specified in a similar way to $U_i(t)$. Note that $\mathbf{z}_i$ may or may not have elements in common with $\mathbf{x}_i$ of the longitudinal model. Under this setup, the probability density function and the survival function can be written as

$$f_i(t) = f_0(g_i(t)) \exp(-\mathbf{z}_i'\boldsymbol{\beta} - V_i(t)), \tag{3.5}$$

and

$$S_i(t) = S_0(g_i(t)), \tag{3.6}$$

where $f_0(\cdot)$ and $S_0(\cdot)$ are the baseline probability density function and the baseline survival function, respectively.

3.4 Association Structure

In our implementation, dependence between the longitudinal and the time-to-event sub-models is captured through $V_i(t) = \phi U_i(t)$ so that ϕ is a measure of association induced by the fitted longitudinal values. Although association via longitudinal values is the most common way to formulate a joint model, the association structure can also be modeled using random intercepts or random slopes. Hederson et al. (2000) discussed such association structures for PH models. Addressing the association via random intercepts or random slopes for AFT models could be an area of future research.

3.5 Example: Log-Logistic Sub-Model

Using equation (3.4) and $V_i(t) = \phi U_i(t) = \phi \mathbf{w}_i'(t)\mathbf{b}_i$, the hazard function at time t_i can be written as

$$h_i(t_i) = \frac{\kappa\rho(\rho g_i(t_i))^{\kappa-1}}{1+(\rho g_i(t_i))^{\kappa}} \exp(-\mathbf{z}_i'\boldsymbol{\beta} - \phi \mathbf{w}_i'(t_i)\mathbf{b}_i), \tag{3.7}$$

where $g_i(t_i) = \int_0^{t_i} \exp(-\mathbf{z}_i'\boldsymbol{\beta} - \phi \mathbf{w}_i'(u)\mathbf{b}_i)du$ and $h_0(g_i(t_i)) = \kappa\rho(\rho g_i(t_i))^{\kappa-1}/[1+(\rho g_i(t_i))^{\kappa}]$. Similarly, the survival function can be expressed as

$$S_i(t_i) = S_0(g_i(t_i)) = [1+(\rho g_i(t_i))^{\kappa}]^{-1}. \tag{3.8}$$

Using equation (3.7) and equation (3.8), the density function of (t_i, δ_i) given $\mathbf{b}_i$ and $\boldsymbol{\theta}$ can be written as

$$f(t_i, \delta_i|\mathbf{b}_i, \boldsymbol{\theta}) = \left\{ \frac{\kappa\rho(\rho g_i(t_i))^{\kappa-1}}{1+(\rho g_i(t_i))^{\kappa}} \exp(-\mathbf{z}_i'\boldsymbol{\beta} - \phi \mathbf{w}_i'(t_i)\mathbf{b}_i) \right\}^{\delta_i} \times [1+(\rho g_i(t_i))^{\kappa}]^{-1}, \tag{3.9}$$

which plays an important role in formulating the joint model in a Bayesian framework (see Section 3.7).

3.6 Example: Weibull Sub-Model

Using equation (3.4) and $V_i(t) = \phi U_i(t) = \phi \mathbf{w}_i'(t)\mathbf{b}_i$, the hazard function at time t_i can be written as

$$h_i(t_i) = \kappa\rho(\rho g_i(t_i))^{\kappa-1} \exp(-\mathbf{z}_i'\boldsymbol{\beta} - \phi \mathbf{w}_i'(t_i)\mathbf{b}_i), \tag{3.10}$$

where $g_i(t_i) = \int_0^{t_i} \exp(-\mathbf{z}_i'\boldsymbol{\beta} - \phi \mathbf{w}_i'(u)\mathbf{b}_i)du$ and $h_0(g_i(t_i)) = \kappa\rho(\rho g_i(t_i))^{\kappa-1}$. Similarly, the survival function can be expressed as

$$S_i(t_i) = S_0(g_i(t_i)) = \exp[-(\rho g_i(t_i))^{\kappa}]. \tag{3.11}$$

Using equation (3.10) and equation (3.11), the density function of (t_i, δ_i) given $\mathbf{b}_i$ and $\boldsymbol{\theta}$ can be written as

$$f(t_i, \delta_i|\mathbf{b}_i, \boldsymbol{\theta}) = \left\{ \kappa\rho(\rho g_i(t_i))^{\kappa-1} \exp(-\mathbf{z}_i'\boldsymbol{\beta} - \phi \mathbf{w}_i'(t_i)\mathbf{b}_i) \right\}^{\delta_i} \times \exp[-(\rho g_i(t_i))^{\kappa}], \tag{3.12}$$

which plays an important role in formulating the joint model in a Bayesian framework (see Section 3.7).

3.7 Joint Model: A Bayesian Approach

Our choices of distributions for the relevant quantities allow us to write the joint model as

$$\left.\begin{aligned}
[y_{ij}|\mathbf{b}_i, \boldsymbol{\theta}] &\sim N\Big(\mathbf{x}_i'(s_{ij})\boldsymbol{\alpha} + \mathbf{w}_i'(s_{ij})\mathbf{b}_i, \sigma^2\Big), \\
[0|\mathbf{b}_i, \boldsymbol{\theta}] &\sim \text{Poisson}(-\log f(t_i, \delta_i|\mathbf{b}_i, \boldsymbol{\theta})), \\
[\mathbf{b}_i|\boldsymbol{\theta}] &\sim N(\mathbf{0}, \Sigma_b), \left[\Sigma_b^{-1}|\nu, \Psi\right] \sim \text{Wishart}(\Psi, \nu), \\
[\boldsymbol{\alpha}|\boldsymbol{\mu}_\alpha, \Sigma_\alpha] &\sim N(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), [\boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \Sigma_\beta] \sim N(\boldsymbol{\mu}_\beta, \Sigma_\beta), \\
[\phi|a_0, a_1] &\sim N(a_0, a_1), \left[\sigma^{-2}|b_0, b_1\right] \sim \text{Gamma}(b_0, b_1), \\
[\rho|c_0, c_1] &\sim \text{Gamma}(c_0, c_1), \left[\kappa^{-1}|d_0, d_1\right] \sim \text{Gamma}(d_0, d_1),
\end{aligned}\right\} \tag{3.13}$$

where the hyperparameters Ψ, ν, $\boldsymbol{\mu}_\alpha$, Σ_α, $\boldsymbol{\mu}_\beta$, Σ_β, a_0, a_1, b_0, b_1, c_0, c_1, d_0 and d_1 are all assumed known. Some remarks are necessary for the hierarchical formulation (equation 3.13).

1. Gauss-Legendre 5-point quadrature rule (Abbott 2005) is used to evaluate the integral $g_i(t_i)$ in $\log f(t_i, \delta_i|\mathbf{b}_i, \boldsymbol{\theta})$ (see Appendix B).

2. The so-called "zeros-trick" (Spiegelhalter et al., 2003) is used to specify the distribution of (t_i, δ_i), as it is not of a standard form. The idea behind this technique is that the contribution of a Poisson(ξ) observation of zero is $\exp(-\xi)$; if we set $\xi_i = -\log f(t_i, \delta_i|\mathbf{b}_i, \boldsymbol{\theta})$, $i = 1, 2, \ldots, m$, with observed data a vector of 0's, then we get the correct contributions.

3. For the gamma and Wishart distributions, we consider the parameterizations implemented in the WinBUGS software (Lunn et al. 2000). For example, b_0 and b_1 are the shape and inverse-scale parameters of the Gamma(b_0, b_1) distribution, respectively, and Ψ and ν are the inverse-scale matrix and degrees of freedom of the Wishart distribution Wishart(Ψ, ν), respectively.

3.8 Software Implementation

The MCMC algorithm is implemented in WinBUGS (Lunn et al. 2000) using fairly vague, minimally informative priors. For example, $\boldsymbol{\alpha}$ is assumed to follow a multivariate normal distribution with mean zero and variance-covariance matrix Σ_α, where Σ_α is a diagonal matrix with all the diagonal elements equal to 10^5. For a gamma prior, for example $[\sigma^{-2}|b_0, b_1] \sim \text{Gamma}(b_0, b_1)$, we take $b_0 = b_1 = 0.1$.

Note that the statistical software R is used for data manipulation and organization. The R2WinBUGS package in R (R Core Team 2018) is then used to run WinBUGS (Lunn et al. 2000) from R for MCMC samples, and the coda package (Plummer et al., 2006) is used for summarizing and plotting MCMC output.

4 Data Analyses

Our proposed methodology is illustrated with an application to aortic valve replacement surgery data. The description of the data is presented in Section 4.1, and the variables of interest and the goal of the study are highlighted in Section 4.2. We then present the analysis of the data using our proposed methodology in Section 4.3. A conclusion of our findings is given in Section 4.4.

4.1 Aortic Valve Replacement Surgery Data

The aortic heart valve replacement surgery data are publicly available in the joineR package (Philipson et al. 2018) of the statistical software R (R Core Team 2018). This is a longitudinal data set collected from an observational study. In this study, different types of heart valves were implanted on the aortic positions with the goal to detect their effects on survivals of the patients.

The study received ethical approval from the chairman of the ethics committee; note that the committee members decided to waive the need for patient consents (Lim et al. 2008). Data were collected from consecutive series of patients who underwent aortic valve replacement surgery performed by a single surgeon from 1991 to 2001. In this series, the predominant stentless valve used for the aortic valve replacement was Toronto stentless porcine valve followed by homograft valve. Moreover, Ali et al. (2003) discuss the detail of the surgical procedure used for the implementation of stentless and homograft valve.

The total number of patients who participated in the study was 256 and all the patients were studied for ten years from the year 1991 to the year 2001 with at least one year of follow-up with a serial total of 1,273 echocardiographic measurements. Echocardiography was performed on an annual basis. Patients who underwent two or more procedures were censored from the time point of the second procedure to ensure that every patient was analyzed only once. The study leads to the fact that three heart valve functions are associated with the improvement of survival of patients with a stentless or a homograft valve replacement.

The three heart valve functions are described as

- grad: valve gradient at follow-up visit
- lvmi: left ventricular mass index (standardized) at follow-up visit

- ef: ejection fraction at follow-up visit

In this thesis, for the longitudinal process, we consider log.lvmi (natural logarithm transformation of lvmi) as a longitudinal outcome (internal covariate) .

The collected dataset is unbalanced; all the subjects do not have the same number of observations. Moreover, along with the longitudinal outcomes, there are several baseline covariates available and also, the information about survival data is available. Demographic, preoperative, and mortality data were obtained from individual hospital notes, death certificates, and autopsy reports.

4.2 Study Variables of Interest

Recall that our objectives of this thesis are; developing joint models using AFT event processes, writing the model in a hierarchical framework for Bayesian implementation, developing computational algorithms to fit joint models, implemented in WinBUGS (a Bayesian software) and R and finally to analyze the aortic valve data using our approach and investigate the effects of baseline covariates on both the longitudinal response and the event process. In the subsequent sections, first, we describe the variables of our interest, then describe the proposed joint models based on data and finally present and discuss the results from the analysis of the aortic valve data.

4.2.1 Longitudinal Outcome

log.lvmi: Natural log transformation of left ventricular mass index (standardized) at follow-up visit

log.lvmi is a continuous variable and defined as
lvmi = (left ventricular mass) / (body surface area)
Left ventricular mass (LVM) was calculated from M-mode recordings with measurement unit is grams and indexed to body surface area, which is measured in meters^2.
LVMI is calculated in two steps (for more detail, see Lim et al. 2008), first LV mass is computed by using the following formula,

$$\text{LV mass} = 1.o4([\text{IVS}_d + \text{LVID}_d + \text{PWT}_d]^{-3} - [\text{LVID}_d]^3) - 13.6.$$

where LV mass is left ventricular mass in grams, IVS_d is end-diastolic interventricular septum, LVID_d is end-diastolic left ventricular internal diameter, PWT_d is enddiastolic posterior wall thickness. Then LV mass was indexed to body surface area which resulted in the left ventricular mass index (LVMI).

4.2.2 Longitudinal Time-varying Covariate

time: follow-up time

"Time" is a continuous variable measured in the number of years, the time values are obtained from observed time points (follow-up time points). Surgery date is considered as the time of origin (years).

4.2.3 Survival Outcome

status: censoring indicator

The variable "status" indicates, either the event (death) occurs or not. It is a categorical variable with two categories. We coded this variable in the analysis as

$$\text{status}(\delta) = \begin{cases} 1 & \text{if died} \\ 0 & \text{if lost to } follow \text{ up} \end{cases} \tag{4.1}$$

where the status "0= lost to follow up" is considered as the reference category.

fuyrs: event time

"fuyrs" is a continuous variable and provides us with the information about maximum follow-up time (years) with surgery date as the time of origin. This variable contains information about the time of occurrence of the event for the analysis of our study.

4.2.4 Longitudinal and Survival Baseline Covariates

hs: implanted aortic prosthesis type

In this study, the baseline covariate "hs" provides us with information about the type of heart valve implanted during surgery based on tissue type. Evidence from the literature concludes that the choice of the type of heart valve is a significant factor associated with the risk of death after the aortic heart valve replacement surgery and also has an impact on the life expectancy of the patient (Siniawski et al. 2003). We investigate the effects of hs on log left ventricular mass index (log.lvmi) and the risk of death.
hs is a categorical variable with two categories and we coded hs as

$$\text{hs} = \begin{cases} 1 & \text{if homograft} \\ 0 & \text{if stentless procine tissue} \end{cases} \tag{4.2}$$

where "0 = stentless porcine tissue" is considered as the reference category.

lv: preoperative left ventricular ejection fraction (LVEF)

About the impact of lv for the risk of death after valve surgery, literature has inconsistent conclusions (Forman 1980, D'Onofrio et al. 2017, Goldberg et al. 2013 and Gaudino et al. 2004). We investigate the effects of "lv" on internal covariate loglvmi and the risk of death after valve replacement surgery. "lv" has three categories, we consider "3=poor" as the reference category and create two dummy variable (lv1 and lv2). We use the following coding for lv in the analysis.

$$\text{lv1} = \begin{cases} 1 & \text{if good} \\ 0 & \text{if otherwise} \end{cases} \tag{4.3}$$

$$\text{lv2} = \begin{cases} 1 & \text{if moderate} \\ 0 & \text{if otherwise} \end{cases} \tag{4.4}$$

redo: previous cardiac surgery

Information collected from the hospital notes about the patient with previous cardiac surgery was named as "redo". It is a categorical variable and has two categories.

Previous cardiac surgery (redo) is considered as an important factor for long term survival after valve replacement surgery (Shehada et al. 2017). We investigate the association of "redo" on log.lvmi as well as on the risk of death in the study. We coded redo as

$$\text{redo} = \begin{cases} 1 & \text{if Yes} \\ 0 & \text{if No previous cardiac surgery} \end{cases} \tag{4.5}$$

where "0= No previous cardiac surgery" is the reference category.

dm: preoperative diabetes

Information collected about the patient with preoperative diabetes was named as "dm". It is a categorical variable and has two categories. Study shows that the patient with preoperative diabetes is associated with significantly worse outcomes after valve replacement operation (Nakamura et al. 2016). We investigate the effects of preoperative diabetes (dm) on log.lvmi as well as on the risk of death. We use the following coding for "dm",

$$\text{dm} = \begin{cases} 1 & \text{if has diabetes} \\ 0 & \text{if has no dibetes} \end{cases} \tag{4.6}$$

where "0= has no diabetes" is the reference category.

age: age of the patient at day of surgery (years)

"age" is a continuous variable. Results from the literature showed, age has a strong association with lvmi (Villa et al. 2006 and Nakamura et al. 2016) as well as on the survival rate (Langanay et al. 2012 and Tjang et al. 2007) after heart valve replacement surgery. We investigate the association of age on log.lvmi and the risk of death.

sex: gender of patient

The impact of "sex" has been identified to influence the left ventricular mass regression (Villa et al. 2006), and also has an impact on survival (Kulik et al. 2009) after heart valve surgery. We investigate the effects of sex on loglvmi as well as on the risk of death after aortic heart valve replacement surgery. In the analysis, we coded "sex" as,

$$\text{sex} = \begin{cases} 1 & \text{if Female} \\ 0 & \text{if Male} \end{cases} \tag{4.7}$$

Male is considered a reference group.

4.2.5 Model Building Based on Data

The formulation of our longitudinal sub models is already discussed in section 3.2 of this thesis. Here we will implement this formulation according to our data. Let y_{ij} denote the logarithm of j^{th} left ventricular mass index (log.lvmi) for patient i, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, n_i$. Then the linear mixed-effect model for the longitudinal response at time $s_{ij} = \text{time}_{ij}$ is defined as (using equation (3.1), equation (3.2) and equation (3.3))

$$(\text{log.lvmi})_{ij} = \mathbf{x'_{ij}}\boldsymbol{\alpha} + \mathbf{w'_{ij}}\mathbf{b_i} + \epsilon_{ij}, \tag{4.8}$$

where

$$\mathbf{x'_{ij}}\boldsymbol{\alpha} = \alpha_1 + \alpha_2\ \text{time}_{ij} + \alpha_3\ \text{hs}_i + \alpha_4\ \text{lv1}_i + \alpha_5\ \text{lv2}_i + \alpha_6\ \text{redo}_i + \alpha_7\ \text{dm}_i + \alpha_8\ \text{age}_i + \alpha_9\ \text{sex}_i, \tag{4.9}$$

and

$$\mathbf{w'_{ij}}\mathbf{b_i} = b_{i0} + b_{i1}\ \text{time}_{ij}. \tag{4.10}$$

Recall that joint density functions for both proposed survival sub-models given $\mathbf{b_i}$ and $\boldsymbol{\theta}$ including shared random effect ($\phi\ (\mathbf{w}'_i(t_i)\mathbf{b}_i)$) are defined by equation (3.9) and equation (3.12) and written as

$$f(t_i, \delta_i|\mathbf{b}_i, \boldsymbol{\theta}) = \left\{ \frac{\kappa\rho(\rho g_i(t_i))^{\kappa-1}}{1 + (\rho g_i(t_i))^{\kappa}} \exp(-\mathbf{z}'_i\boldsymbol{\beta} - \phi\mathbf{w}'_i(t_i)\mathbf{b}_i) \right\}^{\delta_i} \times [1 + (\rho g_i(t_i))^{\kappa}]^{-1}. \tag{4.11}$$

and

$$f(t_i, \delta_i|\mathbf{b}_i, \boldsymbol{\theta}) = \left\{\kappa\rho(\rho g_i(t_i))^{\kappa-1}\exp(-\mathbf{z}_i'\boldsymbol{\beta} - \phi\mathbf{w}_i'(t_i)\mathbf{b}_i)\right\}^{\delta_i} \times \exp[-(\rho g_i(t_i))^{\kappa}]. \tag{4.12}$$

where

$$\mathbf{z_{ij}}'\boldsymbol{\beta} = \beta_1\ \text{hs} + \beta_2\ \text{lv1}_i + +\beta_3\ \text{lv2}_i + \beta_4\ \text{redo}_i + \beta_5\ \text{dm}_i + \beta_6\ \text{age}_i + \beta_7\ \text{sex}_i, \tag{4.13}$$

and

$$g_i(t_i) = \int_0^{t_i} \exp(-\mathbf{z}_i'\boldsymbol{\beta} - \phi\mathbf{w}_i'(u)\mathbf{b}_i)du. \tag{4.14}$$

Note that the integration described in equation (4.14) has no closed form solution. Therefore, we use the 5-point Gauss-Legendre Quadrature rule (Abbott 2005) to solve this Integration (for more detail about Gauss-Legendre Quadrature rule, see Appendix B).

4.3 Data Analysis and Results

In this section, first, we present the descriptive analysis, then we present and discuss the results of both proposed Bayesian joint models (described in chapter 3).

4.3.1 Descriptive Statistics

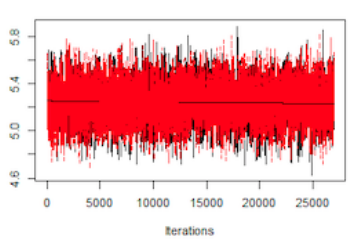
Table 4.1 presents a descriptive analysis of all the categorical and continuous variables of interest from the study. Descriptive analysis (Table 4.1) shows till at the end of the study, only 54 (21.09 %) patients experience the event (death) and 202 (78.91 %) patients either lost the follow-up or did not experience the event (censored observations). Stentless porcine tissue type valve was implanted in 48.05 % of the patients and 51.95 % patients ' heart valve was replaced by homograft type valve. The average measurement of log.lvmi was 5 gm/m^2 (S.D = 0.38) and the average age of the patients who participated in the study was 66 years (S.D = 12.37). The mean follow-up time was 3.14 years (S.D = 2.47) and the mean survival time was 5.32 years (S.D = 2.52) after aortic heart valve replacement surgery. Moreover, among 256 patients who participated in the study, approximately 71 % were male and 29 % were female.

Table 4.1: Characteristics of categorical and continuous variables of heart valve data (N = 256)

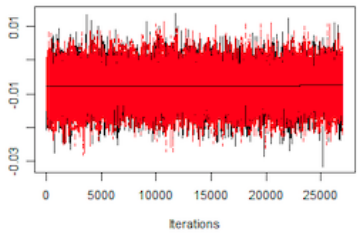
Categorical Variables	**levels**	**Frequency** (%)
status (survival response)		
	lost at follow up	202 (78.91 %)
	died	54 (21.09 %)
hs		
	stentless porcine tissue	133 (51.95 %)
	Homograft	123 (48.05 %)
lv		
	good	147 (57.42 %)
	moderate	87 (33.98 %)
	poor	22 (8.59 %)
redo		
	no previous cardiac surgery	230 (89.84 %)
	have orevious cardiac surgery	26 (10.16 %)
dm		
	no diabetes	244 (95.31 %)
	has diabetes	12 (4.69 %)
sex		
	Male	183 (71.48 %)
	Female	73 (28.52 %)
Continuous Variables	**Mean**	**Standard Deviation**
log.lvmi (longitudinal response)	5.00	0.38
time: Follow up time	3.14	2.47
fuyrs: time (survial response)	5.32	2.52
age	66.00	12.37

4.3.2 Results from Log-Logistic AFT (Model 1) and Weibull AFT (Model 2) in Bayesian Joint Model Fits

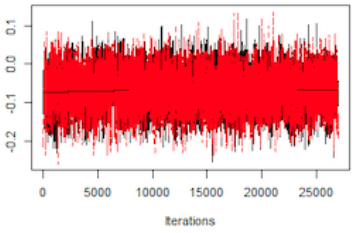
We implement MCMC methods to check the efficiency of both proposed Bayesian joint models in the analysis. For the MCMC method, the simulation of the posterior distribution is made using the Gibbs sampler algorithm. First, we construct two Markov chains to approximate the posterior density and produced two realizations each of 150,000 iterations. A burn-in of 15,000 iterations is considered, i.e., we discarded the initial 15,000 iterations to get the good mixing and convergence of chains. Moreover, the Bayesian inferences of our models are based on every 5^{th} iteration of the chain called thinning (Gelman 1996). From the thinning process, we get 27,000 total number of iterations (sample size) for each chain. MCMC diagnoses are also assessed. Diagnostic assessment of MCMC usually implies from the graphical assessment (Roy 2019) of the behaviour of the sample with respect to each fitted parameter.



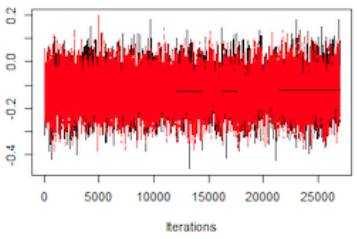
(a) Intercept



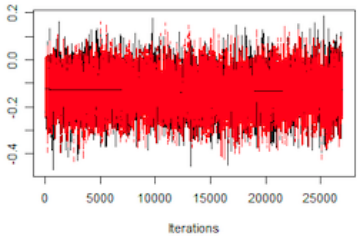
(b) time



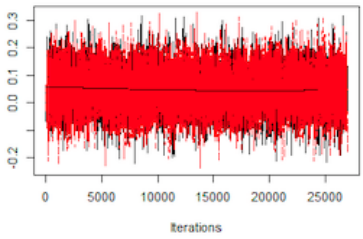
(c) hs



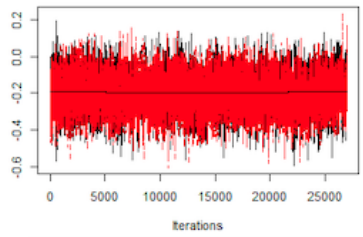
(d) lv1



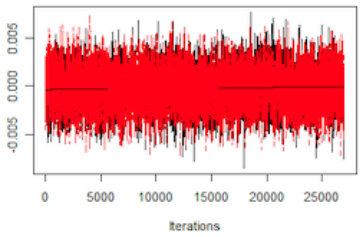
(e) lv2



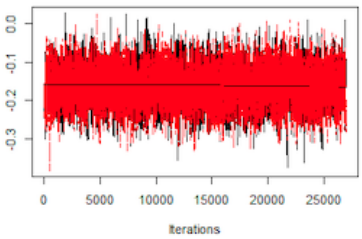
(f) redo



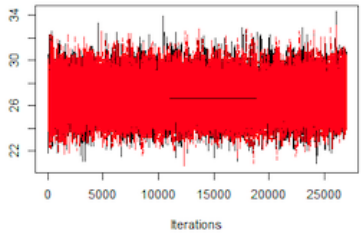
(g) dm



(h) age

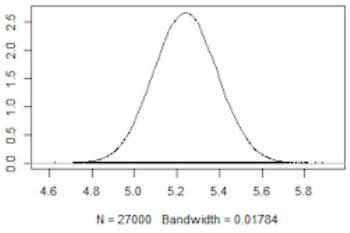


(i) sex

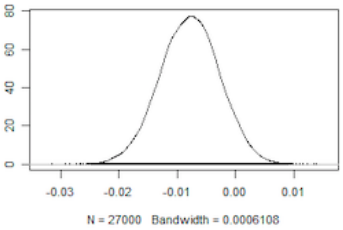


(j) σ^{-2}

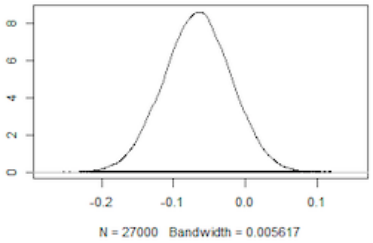
Figure 4.1: Trace plots of longitudinal sub-model parameters in Bayesian joint log-logistic AFT model fit.



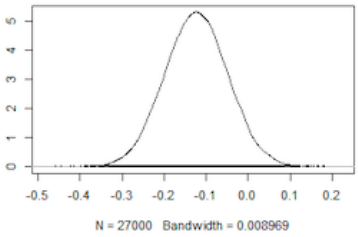
(a) Intercept



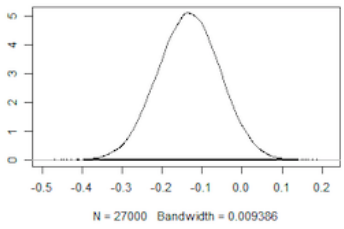
(b) time



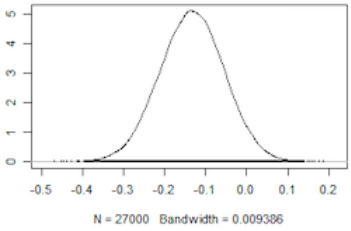
(c) hs



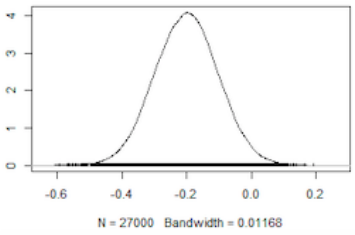
(d) lv1



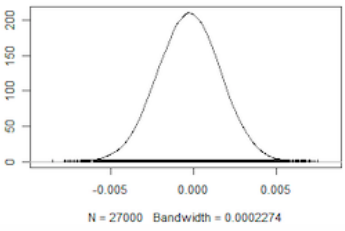
(e) lv2



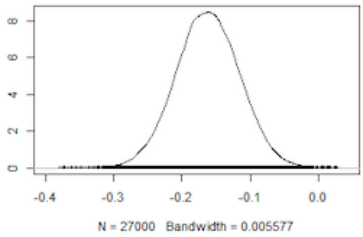
(f) redo



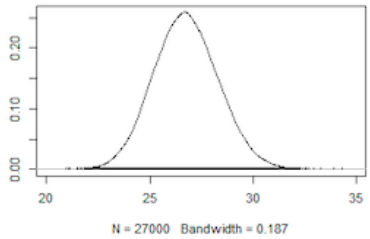
(g) dm



(h) age

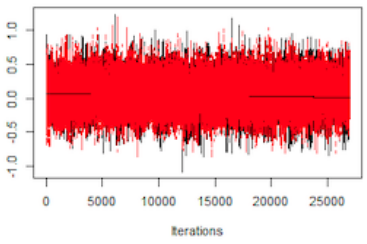


(i) sex

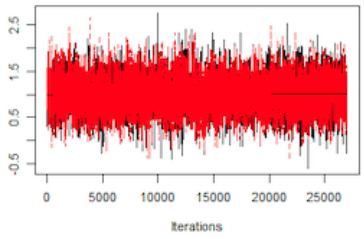


(j) σ^{-2}

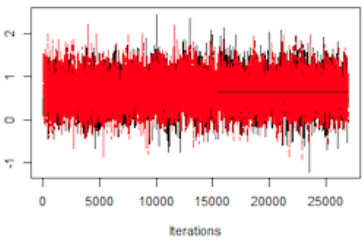
Figure 4.2: Density plots of longitudinal sub-model parameters in Bayesian joint log-logistic AFT model fit.



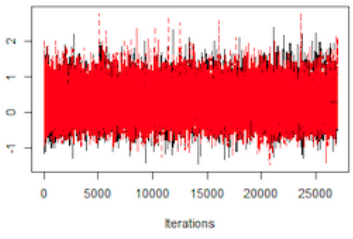
(a) hs



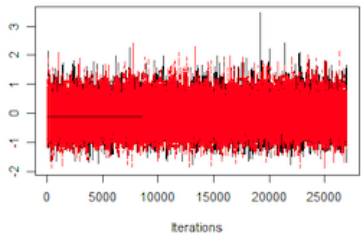
(b) lv1



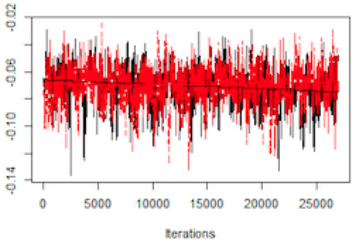
(c) lv2



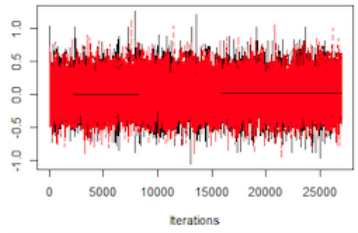
(d) redo



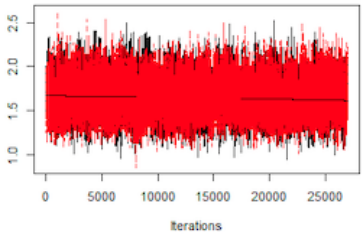
(e) dm



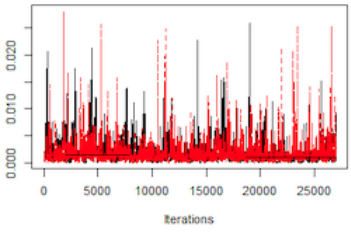
(f) age



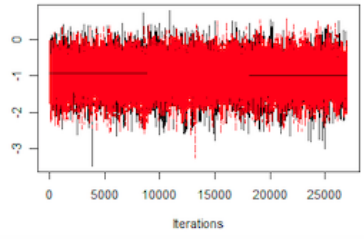
(g) sex



(h) shape=κ

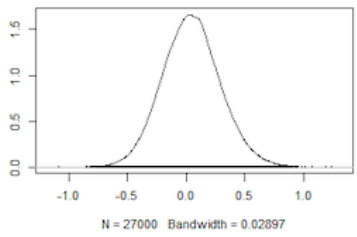


(i) rate=ρ

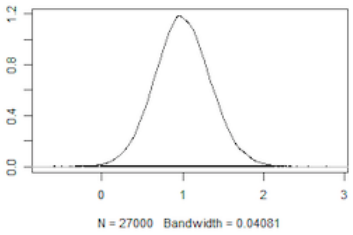


(j) Association=ϕ

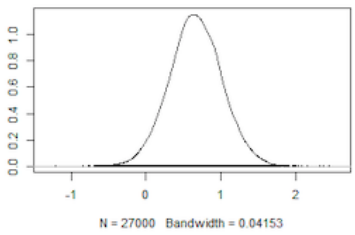
Figure 4.3: Trace plots of log-logistic AFT survival sub-model parameters in Bayesian joint model fit.



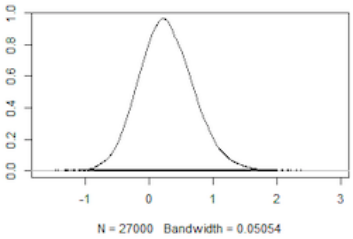
(a) hs



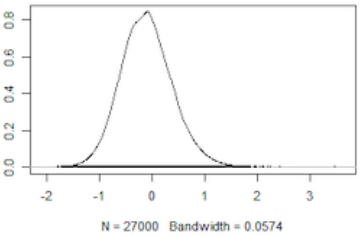
(b) lv1



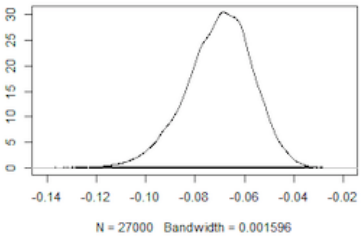
(c) lv2



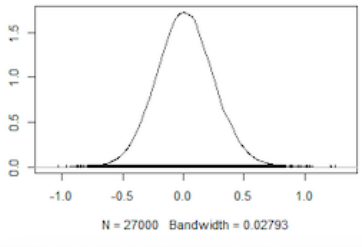
(d) redo



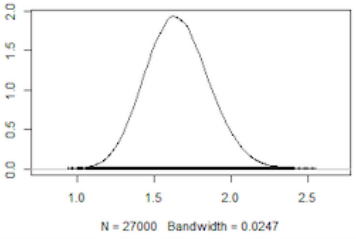
(e) dm



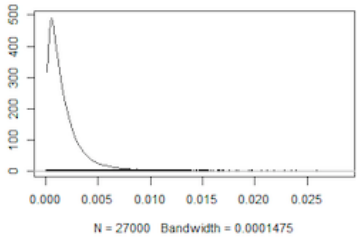
(f) age



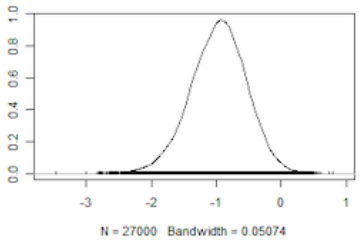
(g) sex



(h) shape=κ

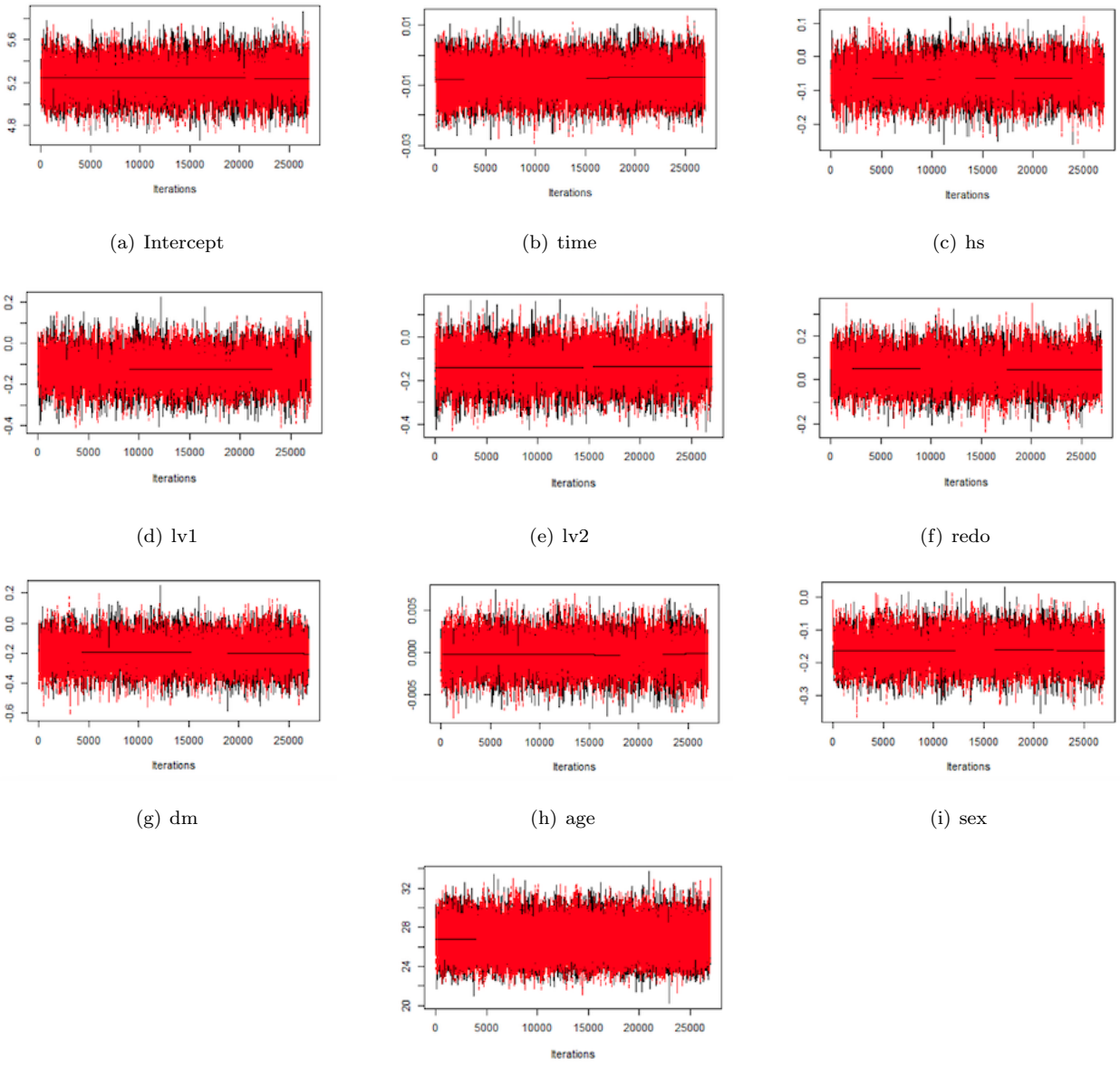


(i) rate=ρ



(j) Association=ϕ

Figure 4.4: Density plots of log-logistic AFT survival sub-model parameters in Bayesian joint model fit.



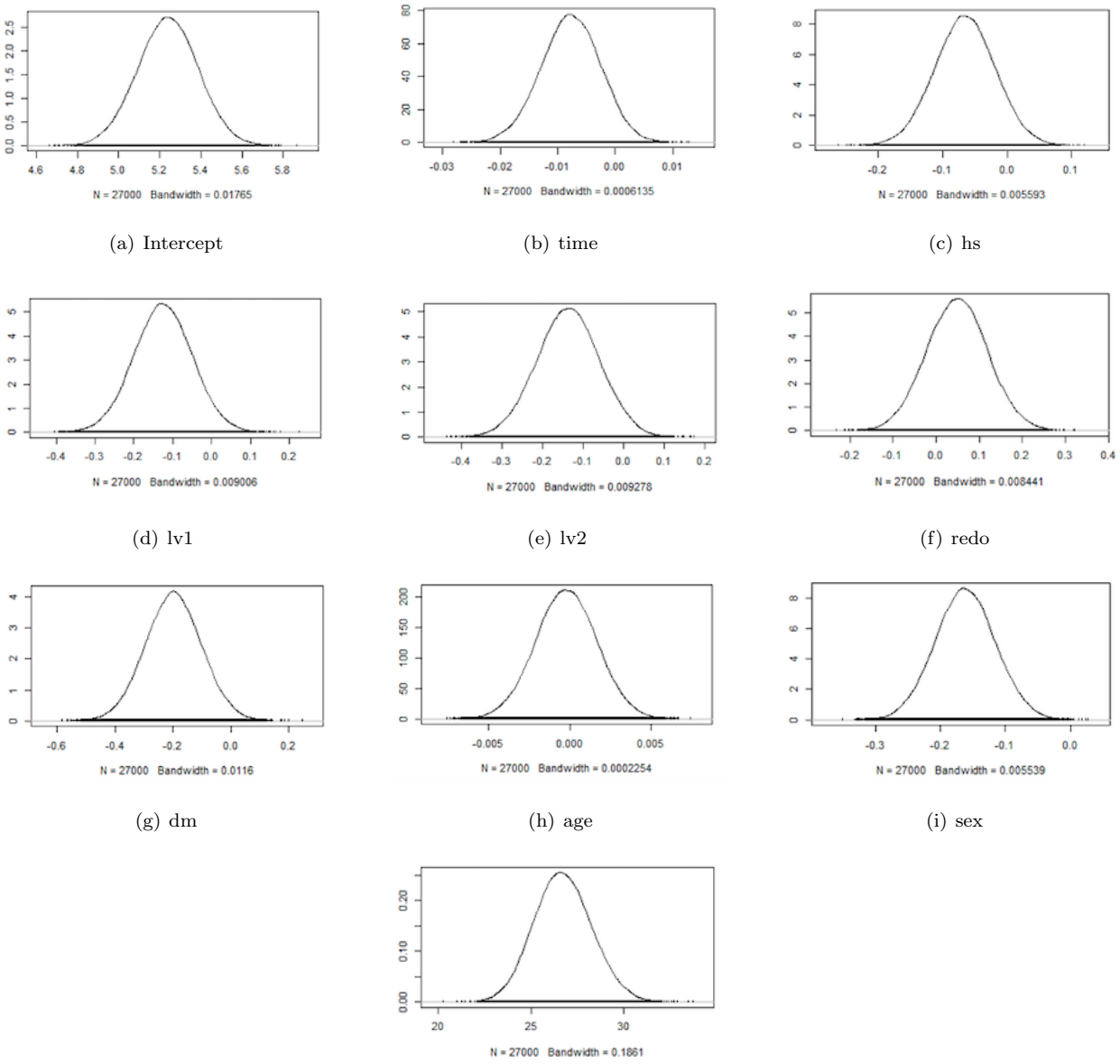
(a) Intercept (b) time (c) hs

(d) lv1 (e) lv2 (f) redo

(g) dm (h) age (i) sex

(j) σ^{-2}

Figure 4.5: Trace plots of longitudinal sub-model parameters in Bayesian joint Weibull AFT model fit.



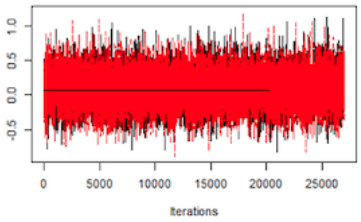
(a) Intercept (b) time (c) hs

(d) lv1 (e) lv2 (f) redo

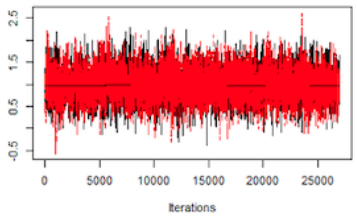
(g) dm (h) age (i) sex

(j) σ^{-2}

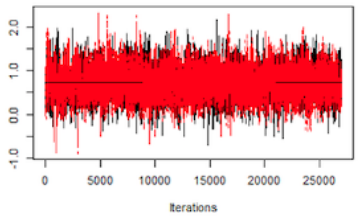
Figure 4.6: Density plots of longitudinal sub-model parameters in Bayesian joint Weibull AFT model fit.



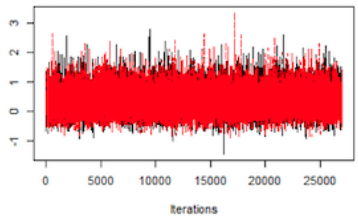
(a) hs



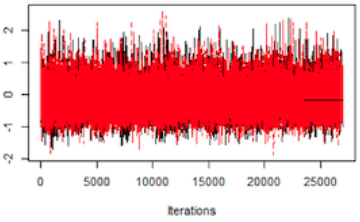
(b) lv1



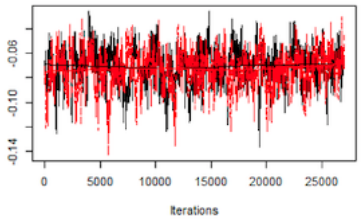
(c) lv2



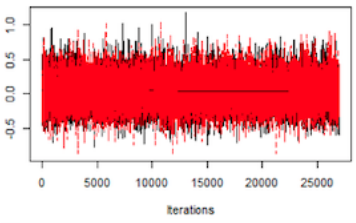
(d) redo



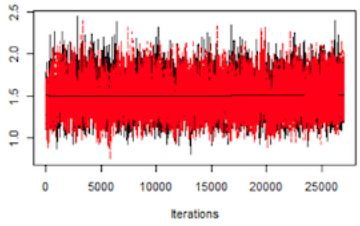
(e) dm



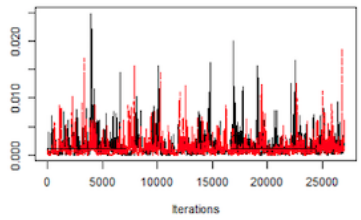
(f) age



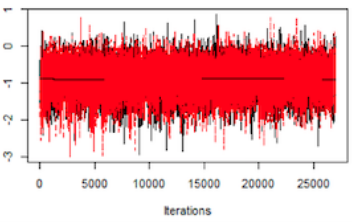
(g) sex



(h) shape=κ

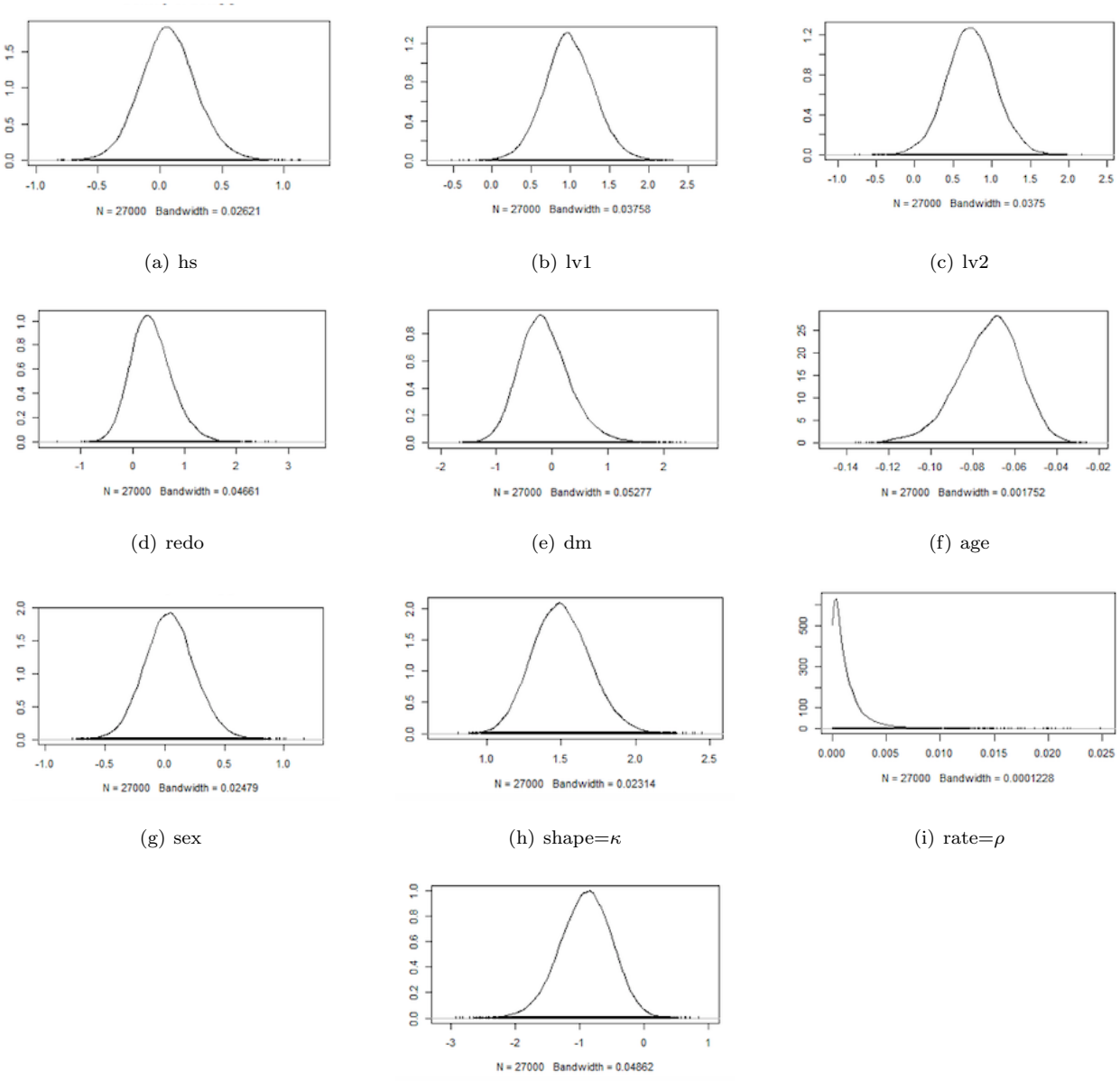


(i) rate=ρ



(j) Association=ϕ

Figure 4.7: Trace plots of Weibull AFT survival sub-model parameters in Bayesian joint model fit.



(a) hs (b) lv1 (c) lv2

(d) redo (e) dm (f) age

(g) sex (h) shape=κ (i) rate=ρ

(j) Association=ϕ

Figure 4.8: Density plots of Weibull AFT survival sub-model parameters in Bayesian joint model fit.

MCMC Diagnostics of Bayesian Joint Model 1 (section 3.5)

Trace plots of each estimated parameter for longitudinal and log-logistic AFT survival sub-model in Bayesian joint model fit are presented in Figure 4.1 and Figure 4.3, respectively. These trace plots exhibit rapid up-and-down variation with no long-term trends or drifts. The lack of any trend in the trace plots of both chains, indicating chains forget their initial values very quickly and quickly explore the full support and the shape of the targeted distribution. It also indicates that the convergence in targeted distribution takes place

rapidly. Thus, we have good mixing and convergence of both chains.

Density plots of each estimated parameter for longitudinal and log-logistic AFT survival sub-model in Bayesian joint model fit are presented in Figure 4.2 and Figure 4.4, respectively. From these plots, we see the smooth kernel density functions without any lumpiness and multiple modes, indicating no sign of multimodality. Therefore, the samples accurately approximate the underlying stationary distribution of the Markov Chain.

MCMC Diagnostics of Bayesian Joint Model 2 (section 3.6)

Trace plots of each parameter for longitudinal and Weibull AFT survival sub-model in Bayesian joint model fit are presented in Figure 4.5 and Figure 4.7, respectively. Similar to joint model 1 trace plots (Figure 4.1 and Figure 4.3), these trace plots also exhibit rapid up-and-down variation with no long-term trends or drifts. The lack of any trend in the trace plots of both chains, indicating chains forget their initial values very quickly and quickly explore the full support and the shape of the targeted distribution. It also indicates that the convergence in distribution takes place rapidly. Thus, we have good mixing and convergence of both Markov chains.

Density plots of each parameter for longitudinal and Weibull AFT survival sub-model in Bayesian joint model fit are presented in Figure 4.6 and Figure 4.8, respectively, We see the smooth kernel density functions without any lumpiness and multiple modes like the density plots for joint model 1 (Figure 4.2 and Figure 4.4,), indicating no sign of multimodality. Therefore, the samples accurately approximate the underlying stationary distribution of both Markov Chains.

Moreover, the formal test statistics known as Gelman-Rubin statistic also suggests good mixing and convergence of the chains for both joint models, the value of statistics for all the parameters and quantities of interest is less than 1.2 and approximately equal to one (see Appendix C).

We present the interested posteriors summaries of all the parameters of log-logistic AFT (model 1) and Weibull AFT (model 2) in the Bayesian joint model fits in Table 4.2.

Results from Bayesian joint model 1 (Table 4.2)

The results of longitudinal process for Bayesian joint log-logistic AFT model show that the overall mean in the longitudinal process is significant with estimates 5.240 and 95 % credible interval (4.950, 5.532). The covariates having statistically significant effects on log transformation of left ventricular mass index (log.lvmi) measurement are: patient with preoperative diabetes (dm) with an estimate of -0.199 and excluding 0 in 95 % credible interval (-0.390, -0.009) and sex (female) with an estimate of -0.162 and excluding 0 in 95 % credible interval (-0.254, -0.071). Other covariate including the follow-up time (time), implanted aortic

prosthesis valve based on the type of tissue (hs), the preoperative left ventricular ejection fraction with good (lv1) and moderate level (lv2) relative to poor level, previous heart surgery (redo) and age of the patient have no significant effect on log.lvmi measurements (Table 4.2: 95 % credible interval for all the parameters corresponding to these covariates including 0).

Moreover, the negative sign of the coefficient of preoperative diabetic patients indicates, the diabetic patients have less log transformation of left ventricular mass index measurement value compares to non-diabetic patients. Patients with preoperative diabetes have about 18 % (exp (-0.199) = 0.82) less value of log.lvmi measurement compare to non-diabetes patients, after controlling for the other factors. Female patients are found to have less log transformation of left ventricular mass index measurement value than the male patients and the value of log.lvmi measurement reduced by about 15 % (exp (-0.162) = 0.85) with time in female patients than the male patients, after controlling for the other factors.

For survival (time to event) process, 95 % credible intervals indicate significant effects for the preoperative left ventricular ejection fraction (lv) with good level (lv1) compare to poor level and age on the risk of death (95 % credible intervals for lv1 and age are (0.323, 1.697) and (-0.100, -0.046), respectively, both of which excluding 0) with an estimate of 0.997 and -0.070, respectively. The covariates; choice of heart valve based on type of tissue (hs), the left ventricular ejection fraction with moderate level (lv2) compared to the poor level, previous heart surgery (redo), patient with preoperative diabetes (dm) and sex of the patients have no significant effect on the risk of death (Table 4.2: 95 % credible interval for all the parameters corresponding to these covariates including 0).

The estimated value for covariate preoperative left ventricular ejection fraction (lv1) indicates that the patients with good level of preoperative left ventricular ejection fraction have the high risk of death compared to the patients with poor level of preoperative left ventricular ejection fraction. In the patient with good level of preoperative left ventricular ejection fraction, the risk of death increase by a factor of 2.7 (exp (0.997) = 2.7) compare to the patient with poor level of preoperative left ventricular ejection fraction, after controlling for the other factors. After the surgery, one year increase in age reduces the risk of death by about 7 % (exp (-0.070) = 0.93) after controlling for the other covariates. In other words, we conclude that the poor level of preoperative left ventricular ejection fraction and age are protective factors for the patients after aortic valve replacement surgery, after controlling for the other factors.

For joint Log-logistic AFT analysis of time to event and longitudinal processes, 95 % credible interval quantifies the significance of strong association between longitudinal process (Internal covariate; log.lvmi) and the survival process (relative risk of death) as zero does not belong to credible interval (-1.851, -0.16). Moreover, value of association parameter (ϕ) indicates, one unit increase in the measurement of Internal

covariate (log.lvmi) reduces the relative risk of death about 62 % (exp (-0.963) = 0.382), after controlling for the other factors, i.e., increase value of log.lvm after surgery is a protective factor and helps to increase the survival of patients after surgery, after controlling for the other factors.

Table 4.2: Bayesian parameters estimations of the proposed joint models fits

		Log-Logistic AFT (Model 1) (DIC= 240.7)			**Weibull AFT (Model 2) (DIC = 222.2)**		
	Parameters	Posterior mean	Standard Deviation	95 % Credible Interval	Posterior mean	Standard Deviation	95 % Credible Interval
Longitudinal Process							
	α_1 (Intercept)	5.240	0.149	4.950, 5.532	5.239	0.147	4.950, 5.526
	α_2 (time)	-0.008	0.005	-0.018, 0.002	-0.008	0.005	-0.018, 0.002
	α_3 (hs)	-0.067	0.047	-0.159, 0.024	-0.066	0.047	-0.157, 0.026
	α_4 (lv1)	-0.123	0.075	-0.270, 0.024	-0.125	0.075	-0.273, 0.023
	α_5 (lv2)	-0.133	0.078	-0.287, 0.019	-0.136	0.078	-0.289, 0.015
	α_6 (redo)	0.049	0.072	-0.090, 0.190	0.049	0.070	-0.089, 0.188
	α_7 (dm)	-0.199	0.098	-0.390, -0.009	-0.198	0.097	-0.388, -0.008
	α_8 (age)	-0.0003	0.002	-0.004, 0.004	-0.0002	0.002	-0.004, 0.004
	α_9 (sex)	-0.162	0.047	-0.254, -0.071	-0.162	0.046	-0.252, -0.071
	σ^{-2}	26.74	1.560	23.77, 29.88	26.74	1.552	23.80, 29.88
Survival Process							
	β_1 (hs)	0.044	0.247	-0.430, 0.549	0.069	0.226	-0.366, 0.529
	β_2 (lv1)	0.997	0.347	0.323, 1.697	0.987	0.322	0.360, 1.637
	β_3 (lv2)	0.663	0.351	-0.027, 1.359	0.732	0.321	0.099, 1.370
	β_4 (redo)	0.271	0.433	-0.531, 1.178	0.364	0.403	-0.346, 1.239
	β_5 (dm)	-0.093	0.492	-1.007, 0.941	-0.131	0.462	-0.945, 0.887
	β_6 (age)	-0.070	0.014	-0.100, -0.046	-0.072	0.015	-0.105, -0.046
	β_7 (sex)	0.016	0.237	-0.440, 0.490	0.043	0.212	-0.360, 0.472
	ϕ(Association)	-0.963	0.429	-1.851, -0.16	-0.912	0.408	-1.751, -0.151
	κ(shape)	1.652	0.206	1.266, 2.073	1.502	0.193	1.145, 1.900
	ρ (rate)	0.002	0.002	0.000, 0.007	0.001	0.002	0.000, 0.006

Results from Bayesian joint model 2 (Table 4.2)

Results from longitudinal process for the Bayesian joint Weibull AFT model fit are similar to the longitudinal process for the Bayesian joint log-logistic AFT model fit (Table 4.2). Under the Bayesian joint Weibull AFT model fit from longitudinal process, the results show that the overall mean is significant in the longitudinal process with estimates 5.239 and 95 % credible interval (4.950, 5.526). The covariates having significant effects on log transformation of left ventricular mass index measurement are: patient with preoperative diabetes (dm) with an estimate of -0.198 and excluding 0 in 95 % credible interval (-0.388, -0.008) and

covariate sex (female) with an estimate of -0.162 and excluding 0 in 95 % credible interval (-0.252, -0.071). The covariates, visiting time (time), the choice of heart valve on type of tissue (hs), the left ventricular ejection fraction with good (lv1) level and moderate level (lv2) compare to poor level, previous heart surgery (redo), and age of the patients have no significant effect on log.lvmi (Table 4.2: 95 % credible interval for all the parameters corresponding to these covariates including 0).

Moreover, like joint model1, the negative coefficient of patients with preoperative diabetes indicates, the diabetic patients have lower log transformation of left ventricular mass index measurement value compare to non-diabetic patients. At baseline, the diabetic patient have about 18 % (exp (-0.198)=0.82) less value of log.lvmi measurement compare to the non diabetes patients after controlling for the other factors. Female patients have approximately 15 % (exp (-0.162)=0.85) less value of log.lvmi measurement than the male patients with the time, after controlling for the other factors.

Our results from the survival process for the Bayesian joint Weibull AFT model fit are similar to the results of the survival process for the Bayesian joint log-logistic AFT model fit except for the result of preoperative left ventricular ejection fraction (lv) covariate with moderate level (lv2) compare to poor level. Results from survival process show that the covariates having significant effects on the occurrence of event (death) are: the preoperative left ventricular ejection fraction with good level (lv1) compare to poor level with an estimate of 0.987 and excluding 0 in 95 % (0.360, 1.637), the preoperative left ventricular ejection fraction with moderate level (lv2) compare to poor level with an estimate of 0.732 and excluding 0 in 95 % (0.099, 1.370) and the age of the patient with an estimate of -0.072 and excluding 0 in 95 % (-0.105, -0.046). The choice of heart valve on type of tissue (hs), previous heart surgery (redo), preoperative diabetes (dm) and sex of the patients are all not significant predictors associated with the risk of death after valve replacement surgery (Table 4.2: 95 % credible interval for all the parameters corresponding to these covariates including 0).

The estimated value for the covariates lv1 and lv2 indicate, patient with good level of preoperative left ventricular ejection fraction (lv1) and the patients with a moderate level of preoperative left ventricular ejection fraction (lv2) are at a high risk of death after the surgery compared to the patients with poor level of preoperative left ventricular ejection fraction. Moreover, in the patients with good level of lv the risk of death increased by a factor of 2.7 (exp (0.987) = 2.7) compare to the patient with poor level of left ventricular ejection fraction, after controlling for the other factors and for the patient with moderate level of left ventricular ejection fraction the risk of death increased by a factor of 2.1 (exp (0.732) = 2.1) compare to the patient with poor level of left ventricular ejection fraction, after controlling for the other factors. moreover, after the surgery, one year increase in age reduces the risk of death by about 7 % (exp (-0.072) = 0.93), after controlling for the other covariates.

For joint Weibull AFT analysis of time to event and longitudinal processes, 95 % credible interval quantifies the significance of strong association between longitudinal process (Internal covariate; log.lvmi) and the survival process (relative risk of death) as zero does not belong to the corresponding credible interval (-1.751, -0.151). Moreover, value of association parameter (ϕ) indicates, one unit increase in the measurement of Internal covariate (log.lvmi) reduces the relative risk of death by about 60 % (exp (-0.912)=0.40), after controlling for the other factors.

Furthermore, the estimated values of ancillary parameters (shape and rate) of both AFT models are: For log-logistic distribution, the shape parameter estimation is 1.652 with 0.95 credible interval (1.266, 2.073) and the estimated rate parameter value is 0.002 with 0.95 credible interval (0.000, 0.007). For Weibull distribution, the estimated shape parameter value is 1.502 with 0.95 credible interval (1.145, 1.900) and the rate parameter estimation is 0.001 with 0.95 credible interval (1.145, 1.900).

Note that the results of both proposed survival sub-models indicate that the main baseline covariate hs (implanted aortic prosthesis type of valve; stentless porcine tissue or homograft) is not significantly associated with occurrence of event (death) (Table 4.2: 95 % credible interval including zero). In other words, there is no significant difference is seen between the selection of a heart valve for surgery based on tissue type either stentless porcine tissue or homograft in order to prolong the risk of death after surgery. One possible reason for this nonsignificant result is, only 21 % patients experience the event (death) until the end of the study (Table 4.1).

Finally, from the results (Table 4.2), the DIC value for log-logistic fit is 240.7 and the value of DIC for Weibull fit is 222.2. Therefore, based on DIC values we suggest that the Weibull AFT model perform better in terms of goodness of fit and the principle of parsimony than the Log-Logistic AFT model for the analysis of aortic valve replacement surgery data set (Difference in DIC values = 240.7 – 222.2 = 18.5 $>$ 10).

4.4 conclusion

In this chapter, to check the performance of our proposed Bayesian joint models (detail description in chapter 3) for the longitudinal and survival processes, we implement these proposed models for aortic heart valve replacement surgery data from joineR package in R software. We create the code for Bayesian approach and implement it via WinBUGS and R software. Moreover, we can not ignore the joint modelling approach due to missing observations in the data and association between longitudinal and survival processes.

5 Discussion/Conclusion

The inspiration of biostatistical research arises from clinical and biomedical studies. The data collected from these studies have always been getting attention due to their particular features that need special consideration when doing an analysis. New statistical methods have developed over time to handle an analysis of such data coming from these sources. A typical clinical study often involves collecting repeated measurements on a biomarker (e.g., lvmi measurements) along with an observation of the time to the occurrence of an event (e.g., death), resulting in a joint modeling setup, a model becomes increasingly popular in clinical studies.

In general, joint modeling is a collection of statistical methods to properly handle a longitudinal response while investigating its effects on time to the occurrence of an event simultaneously. (Wulfsohn and Tsiatis 1997 and Tsiatis et al. 1995). Joint models can be formulated with a probability distribution (parametric models) or without assuming a probability distribution (Cox model or semi-parametric Cox PH model) for time-to-event process. However, selection for the joint model framework is quite limited in the literature. The best choice for the selection of longitudinal model can be made based on the observed longitudinal data, and the best survival model can be selected based on the survival data, using standard model selection procedures for these models.

Parametric models are pivotal in the joint modeling of longitudinal and time-to-event data. A non-parametric or semi-parametric model usually leads to an underestimation of standard errors of the parameter estimates in the joint analysis (Hsieh et al. 2006, Rizopoulos 2012). Two general classes of models are in common use for regression analysis of time-to-event data: accelerated failure time (AFT) models and proportional hazards (PH) models. Parametric PH models are commonly considered to describe the time-to-event process of joint modeling. An alternative framework involves considering AFT models, which is particularly useful when the PH assumption is in question. For AFT models, Weibull, Log-logistic, and Lognormal distribution are widely used parametric models. Weibull distribution is the only distribution which is closed for both AFT and PH regression models.

In this thesis, we proposed a joint model framework that consists of a continuous longitudinal outcomes for longitudinal process and a parametric AFT model for time-to-event process and linked both processes via shared random effects, in which a characteristic of the longitudinal process defined as a function of the

random effects is included in the survival model (Wulfsohn and Tsiatis 1997).

We developed two Bayesian joint AFT models involving Log-Logistic and Weibull distributions in Chapter 3. We introduced a link between the parametric AFT survival process and the longitudinal process via one parameter of association (ϕ) corresponding to shared random effects proposed by Henderson et al. 2000. We also developed algorithms for computationally intensive Bayesian approach and implemented it via WinBUGS and R software. Finally, standard available software package R2WinBUGS was used to link WinBUGS and R software for the analysis of aortic heart valve replacement surgery data set. Bayesian analysis provides inferences that are conditional on the data (Statisticat L. L. C 2015). Unlike frequentist methods, the Bayesian analysis does not depend on asymptotic approximation for statistical inference (Senn 2003). Availability of software makes Bayesian implementation for complicated models relatively more straightforward and simple than frequentist methods (involve multiple integrals over random effects).

Analysis was performed with an application to aortic heart valve replacement surgery data (available in joineR package in R software) to illustrate the performance of our two proposed models (Chapter 3) with the aim to explore the effect of time-independent and time-dependent covariates on continuous longitudinal outcome (log.lvmi), and comparing the efficiency of two types of valves based on tissue type (Stentless porcine tissue or Homograft) implanted during surgery and association between internal covariate (longitudinal response: log.lvmi) and the occurrence of an event (death) after the surgery.

From the results (Table 4.2) about the significant effect of the covariates for the longitudinal process of both suggested joint models concluded that only two covariates, diabetes (dm) and sex were significantly associated with log transformation of left ventricular mass index (log.lvmi) measurement. Patients with diabetic had about 18 % less value of log.lvmi measurement compared to non-diabetes patients, after controlling for the other factors. Female patients have about 15 % less value of log.lvmi measurement than the male patients, after controlling for the other factors.

Both survival processes in our analysis showed that the covariates, lv (preoperative left ventricular ejection fraction) and age of the patients had a significant effect on the risk of death after surgery. The impact of covariate lv (preoperative left ventricular ejection fraction) on the survival of patients after the surgery had inconsistent conclusions in the literature (Forman 1980, D'Onofrio et al. 2017, Goldberg et al. 2013, Tjang et al. 2007 and Gaudino et al. 2004). From a study, the nonsignificant effect of lv was seen in the survival rate after surgery (Forman 1980 and D'Onofrio et al. 2017). The minimal effect of preserved lv on postoperative morbidity was noticed and significantly improved survival rate was seen in 6 to 8 years compared with their a reduced lv (Goldberg et al. 2013). Evidence from some studies showed that reduced left ventricle ejection fraction was associated with reduced survival rate (Gaudino et al. 2004 and Tjang

et al. 2007). Our results about lv were different from above-discussed results from the literature and we concluded that the relative risk of death was increased if the preoperative level of lv is good or moderate compared to the poor preoperative level of lv.

Form the dataset used in our study, the average age of the patients participated in the study was 66 years and from both models and the results showed that the risk of death decreased by 7 % with one year increase in the age of patient after aortic heart valve replacement surgery, controlling for other factors. Our results coincide with the results of the studies in the literature (Sundt et al. 2000, Sharabiani et al. 2016 and Langanay et al. 2012). According to Sundt et al. (2000), after surgery the late survival rate was good even in the patients aged greater than 80 years old. Sharabiani et al. (2016) suggested that long-term survival following surgical AVR in over 65 years old patients was excellent and it was about 8 years more than the matched general population and from Langanay et al. 2012 results, even in older patients the mortality rate was low after surgery. Therefore, we suggest that older patients should not be denied surgery due to their advanced age alone.

Note that, from the survival process of both proposed models, our results about the main covariate hs (implanted aortic prosthesis type of valve: stentless porcine tissue or homograft tissue) are different from the results of studies (Siniawski et al. 2003 and Marathe et al. 2019). According to Siniawski et al. (2003), the stentless valves perform better to increase the life expectancy after surgery than homograft valves based on sixty day mortality rate and Marathe et al. (2019) also suggested that stentless valve is a better choice than homograft valve to increase the survival rate. But the results from our study showed that there is no significant effect with respect to valve type implanted during surgery on the risk of death (Table 4.2: 95 % credible interval include zero), i.e., there is no significant difference between the efficiency of two types of valves implanted during surgery based on tissue type (Stentless porcine tissue or Homograft) associated with reducing the risk of death in the patients after surgery. One possible reason for this nonsignificant result is that only 21 % patients experienced the event (death) until the end of the study (Table 4.1).

Moreover, the association between the two processes, longitudinal and time to event outcome was assessed and found statistically significant strong association between longitudinal response log.lvmi measurement and risk of death. From the results we concluded, one gm/m^2 increase in the value of log.lvmi after the surgery reduces the relative risk of death by about 62 % (model 1,Table 4.2) and 60 % (model 2, Table 4.2), respectively, after controlling for other factors. Our result is similar to the result of a study conducted by Lindman et al. (2014). According to Lindman et al. (2014), a better quality of life trend was noticed in patients with greater LVMI regression after the AVR surgery. Since our results indicate that there is a strong significant association between internal time-dependent covariate (longitudinal outcome) log.lvmi and the risk of death after the aortic heart valve replacement surgery. Therefore, ignoring this association may lead to biased

estimates (Guo and Carlin 2004) and cannot reveal the potential association among the responses of both processes, which could be of paramount importance. Thus, a joint analysis should be considered to better understand the joint longitudinal and time to event processes rather than separate analysis of both processes for aortic valve replacement data.

Furthermore, Based on DIC values for the model selection between both Log-logistic and Weibull distribution, Bayesian joint AFT model with Weibull distribution fits the motivated data set more efficiently in the sense of goodness of fit and the principle of parsimonious as compare to the Bayesian joint AFT model with Log-Logistic distribution. Therefore, we recommend, Bayesian joint models with AFT Weibull distribution to jointly analyze the longitudinal and survival data of the aortic valve replacement surgery to predict and control the future performance of heart function lvmi of the aortic valve and the survival of the patients.

We hoped that these findings would contribute to help surgeons and cardiologists to refine the indications, timing, prognostication, and follow-up of patients before and after aortic valve replacement surgery and these findings would contribute to developing the targeted policies and programs regarding the prevention after AVR to increase the life expectancy of patients even in elders. Moreover, these findings will be beneficial to reduce the burden of cost on healthcare and the individuals who are suffering from heart valve diseases.

5.1 Limitation of Study

There are some limitations of this study that should be considered while performing the analysis.

1. The choice of poor selection of initial values may result in slow-mixing for a Markov chain. Hence, the selection of initial values should be chosen carefully to avoid lengthy runs.

2. The choice of prior distribution should be made carefully because Bayesian methods are sensitive to the choice of prior distributions. In the absence of the prior information, non-informative priors or flat priors may be the better choice.

3. Monte Carlo methods can be computationally expensive if the dimensions of the random effects b_i are very large.

5.2 Future Work

In this section, we discuss some possible future extensions of our work.

1. Addressing the association via random intercepts or random slopes for AFT models could be an area of future research.

2. In this thesis, we focused on the development of the joint model, and we did not discuss anything about the diagnostics of the longitudinal and survival sub-models to check the adequacy of the models in the joint modeling framework with AFT models. Hence on possible future work is to consider the diagnostic of the longitudinal and survival sub-models using residual plots.

3. In this study, we applied the joint modeling approach for one continuous longitudinal outcome and one event of occurrence. The future extensions of the shared parameter joint model will be considered with multivariate biomarkers outcome for longitudinal process and competing risks for survival process.

4. In recent work, we only consider the data under the right censoring. We have the plan to work with a data set involving both left-truncation and right censoring in the joint modeling framework with AFT models in the future.

5. For the model selection, we only use the Deviance Information Criterion (DIC). Considering the other model selection criterion, including Akaike's information criterion (AIC), Bayesian information criterion (BIC), and Watanabe-Akaike information criterion (WAIC) along with DIC is also a future research topic of interest.

6. In this thesis we did not discuss anything about the comparison of proposed models with other models, therefore conducting the simulation study for our proposed Joint Bayesian models with suggested parametric AFt model could be one of the possible future work.

Bibliography

[1] Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4), 701–726.

[2] Abbott, P. (2005). Tricks of the trade: Legendre-gauss quadrature. *Mathematica Journal*, 9:689–691.

[3] Abrams, D., Goldman, A., Launer, C., Korvick, J., Neaton, J., Crane, L., Grodesky, M.,Wakefield, S., Muth, K., Kornegay, S., Cohn, D., Harris, A., Luskin-Hawk, R., Markowitz, N., Sampson, J., Thompson, M., and Deyton, L. (1994). Comparative trial of didanosine and zalcitabine in patients with human immunodeficiency virus infection who are intolerant of or have failed zidovudine therapy. *New England Journal of Medicine*, 330:657–662.

[4] Albert,P. S. & Shih, J. H. (2010). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics*, 66(3), 983–987.

[5] Ali, A., Lim, E., Halstead, J., Ashrafian, H., Ali, Z., Khalpey, Z., ... & Pepper, J. (2003). Porcine or human stentless valves for aortic valve replacement? Results of a 10-year comparative study. *The Journal of heart valve disease*, 12(4), 430-5.

[6] Banerjee, S. (2005). On geodetic distance computations in spatial modeling. textitBiometrics, 61, 617-625.

[7] Bayes, M. & Price, M. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.

[8] Bennett S. (1983). Loglogistic regression models for survival data. *Journal of the Royal Statistical Society – Series C (Applied Statistics)*, 32, 165–171.

[9] Bloomfield, P. (2002). Choice of heart valve prosthesis. *Heart*, 87(6), 583-589.

[10] Boldgiv, B. (2004). Introduction to Bayesian Statistics by William M. Bolstad, 2004, 354 pages, ISBN 0-471-27020-2, A John Wiley and Sons, Inc., *Mongolian Journal of Biological Sciences*, 2(2), 77-78.

[11] Bolstad, W. M., & Curran, J. M. (2016). *Introduction to Bayesian statistics*. John Wiley & Sons.

[12] Box, G. E. P. & Tiao, G. C. (2011), *Bayesian Inference in Statistical Analysis (Vol. 40)*, John Wiley and Sons.

[13] Breiman, L. (1968). Probability. Addision-Wellsey. Reading, Ma.

[14] Bridgewater, B., Cooper, G., Livesey, S., & Kinsman, R. (2011). Society for cardiothoracic surgery in Great Britain and Ireland. Maintaining patients' trust: modern medical professionalism, *Henley-on-Thames: Dendrite Clinical Systems*.

[15] Brooks, S. P. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4), 434–455.

[16] Brown, E. R., Ibrahim, J. G. & DeGruttola V., (2005). A fexible B-spline Model for Multiple Longitudinal Biomarkers and Survival. *Biometrics* , 61(1), 64-73.

[17] Carapetis, J. R., McDonald, M.,& Wilson, N. J. (2005). Acute rheumatic fever. *The Lancet* , 366, 155-68.

[18] Carlin, B. P., & Louis, T. A. (2008). *Bayesian methods for data analysis*. Chapman and Hall/CRC.

[19] Chan, K. J., Rahman-Haley, S., Mittal, T. K., Gavino, J. A., & Dreyfus, G. D. (2011). Truly stentless autologous pericardial aortic valve replacement: an alternative to standard aortic valve replacement. *The Journal of thoracic and cardiovascular surgery*, 141(1), 276-283.

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

[21] Christensen, O. F. (2004). Monte Carlo maximum likelihood in model-based geostatistics. *Journal of computational and graphical statistics*, 13(3), 702-718.

[22] Collett, D. (2015). *Modelling survival data in medical research*. Chapman and Hall/CRC.

[23] Colosimo, E. A., Fausto, M. A., Freitas, M. A., & Pinto, J. A. (2012). Practical modeling strategies for unbalanced longitudinal data analysis. *Journal of Applied Statistics*, 39(9), 2005-2013.

[24] Conn, P. M. (2006). *Handbook of Models for Human Aging*, Elsevier Academic Press.

[25] Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), 187-220.

[26] Cox, D.R. & Oakes, D. (1984). *Analysis of survival data*. Chapman & Hall, CRC Press London. Volume 21, chapter 5, page 67.

[27] Dantan, E., Joly, P,. Dartigues, J. F. & Jacqmin-Gadda, H. (2011). Joint model with latent state for longitudinal and multistate data. *Biostatistics*, 12(4), 723–736.

[28] DeGruttola, V., & Tu, X. M. (1994). Modelling Progression of CD4-Lymphocyte Count and its Relationship to Survival Time. *Biometrics*, 50(4), 1003-1014

[29] Dempster, A.P., Laird, N.M. & Rublin, D.B., (1997). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-22..

[30] Devereux, R. B., & Reichek, N. (1977). Echocardiographic determination of left ventricular mass in man. Anatomic validation of the method. *Circulation*, 55(4), 613-618.

[31] Diaz, L. C. (2014). Joint Modelling for Longitudinal and Time-to-Event Data. Application to Liver Transplantation Data. *Master's thesis. University of Santiago de Compostela.*

[32] D'Onofrio, A., Besola, L., Rizzoli, G., Bizzotto, E., Manzan, E., Tessari, C., & Fraccaro, C. (2017). Impact of changes in left ventricular ejection fraction on survival after transapical aortic valve implantation. *The Annals of thoracic surgery*, 103(2), 559-566.

[33] Elashoff, R. M., Li, G. & Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64(3), 762–771.

[34] El-Hamamsy, I., Clark, L., Stevens, L. M., Sarang, Z., Melina, G., Takkenberg, J. J., & Yacoub, M. H. (2010). Late outcomes following freestyle versus homograft aortic root replacement: results from a prospective randomized trial. *Journal of the American College of Cardiology*, 55(4), 368-376.

[35] Erango, M.A., Goshu, A.T., Buta, G.B. & Dessiso, A.H. (2017). Bayesian Joint Modeling of Survival of HIV/AIDS Patients Using AFT Data and Longitudinal CD4 Cell Counts. *British Journal of Medicine and Medical Research*, 20: 1-12.

[36] Farewell., VT. & Prentice., RL. (1979). A study of distribution shape in life testing. *Technometerics*, 19;:69-75.

[37] Faucett, C., & Thomas, D. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*, 15(15): 1663–1685.

[38] Faucett, C.L., Schenker, N. & Taylor, J.M.G. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics*, 58(1): 37-47.

[39] Feller, W. (1957). An introduction to probability theory and its applications.

[40] Fitzmaurice, G. M., & Ravichandran, C. (2008). A primer in longitudinal data analysis. *Circulation*, 118(19), 2005-2010.

[41] Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). Applied Longitudinal Analysis. *John Wiley & Sons*, volume 998.

[42] Follmann ,D., & Wu, M. (1995). An Approximate Generalized Linear Model with Random Effects for Informative Missing Data. *Biometrics*, 51 (1): 151-168

[43] Forman, R., Firth, B. G., & Barnard, M. S. (1980). Prognostic significance of preoperative left ventricular ejection fraction and valve lesion in patients with aortic valve replacement. *The American journal of cardiology*, 45(6), 1120-1125.

[44] Gamerman, D., & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2nd ed. London: Chapman & Hall/CRC.

[45] Gaudino, M., Alessandrini, F., Glieca, F., Luciani, N., Cellini, C., Pragliola, C., ... & Possati, G. (2004). Survival after aortic valve replacement for aortic stenosis: does left ventricular mass regression have a clinical correlate?. *European heart journal*, 26(1), 51-57.

[46] GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, 388(10053), 1459-1544.

[47] Gelfand, A. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal Of The American Statistical Association*, 85(410): 398–409.

[48] Gelfand, A. E., Hills, S. E., Racine-Poon, A., & Smith, A. F. M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *Journal of the American Statistical Association*, 85: 972–985.

[49] Gelman, A. (1996). *Inference and monitoring convergence*, Chapman & Hall, London. pages 131–143.

[50] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 1, 1–19

[51] Gelman, A. & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7: 457–472.

[52] Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7 (4): 457-472.

[53] Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman & Hall/CRC.

[54] Geman, S. & Geman, D. (1984). Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

[55] Gemeda, B, Buta, Ayele, T., Goshu, Hailemichael, M. & Worku. (2014). Bayesian joint modelling of disease progression marker and time to death event of HIV/AIDS patients under ART follow-up. *British Journal of Medicine and Medical Research*, 5(8), 1034-1043.

[56] George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, 21(4), 686-694.

[57] Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical Science*, 7 (4): 473-483.

[58] Geyer, C. J., & Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90(431), 909-920.

[59] Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1995). *Introducing markov chain monte carlo.* Chapman & Hall, London, 1-19.

[60] Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.) (1995), *Markov Chain Monte Carlo in Practice.* Chapman & Hall, London, 131–143. .

[61] Givens, G. H., & Hoeting, J. A. (2005). Computational Statistics (Wiley Series in Computation Statistics).

[62] Goldberg, J. B., DeSimone, J. P., Kramer, R. S., DiScipio, A. W., Russo, L., Dacey, L. J., & Clough, R. A. (2013). Impact of preoperative left ventricular ejection fraction on long-term survival after aortic valve replacement for aortic stenosis. *Circulation: Cardiovascular Quality and Outcomes*, 6(1), 35-41.

[63] Gompertz, B. (1815). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Proceedings of the Royal Society of London* Series I, 2, 252-253.

[64] Gott, V. L., Alejo, D. E., & Cameron, D. E. (2003). Mechanical heart valves: 50 years of evolution. *The Annals of thoracic surgery*, 76(6), S2230-S2239.

[65] Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434), 883-904.

[66] Gulbins, H., & Reichenspurner, H. (2009). Which patients benefit from stentless aortic valve replacement?. *The Annals of thoracic surgery*, 88(6), 2061-2068.

[67] Guo, X. & Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician.* 58(1), 16-24.

[68] Halkos, M. E., Kilgo, P., Lattouf, O. M., Puskas, J. D., Cooper, W. A., Guyton, R. A., & Thourani, V. H. (2010). The effect of diabetes mellitus on in-hospital and long-term outcomes after heart valve operations. *The Annals of thoracic surgery*, 90(1), 124-130.

[69] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.

[70] Hatfield, A., Hodges, S., & Carlin, P. (2012). Combining longitudinal and survival information in Bayesian joint models: When are treatment estimates improved? *Biostatistics*, 1-31.

[71] He, B., & Luo, S. (2016). Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. *Statistical methods in medical research*, 25(4), 1346-1358.

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

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

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

[75] Hsieh, F., Tseng, Y.-K., & Wang, J. L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62(4): 1037–1043.

[76] *https://my.clevelandclinic.org/health/articles/17067-heart-valves*

[77] *https://www.nhlbi.nih.gov/health-topics/heart-valve-disease*

[78] *https://www.heartandstroke.ca/heart/conditions/valvular-heart-disease*

[79] *https://rdrr.io/cran/joineR/man/heart.valve.html*

[80] *http://users.jyu.fi/ ∼ hemipu/itms/DIC%20web%20site%20from%20BUGS%20project.pdf*

[81] Huang, Y., Dagne, G. & Wu, L. (2011). Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. *Statistics in Medicine*, 30(24), 2930–2946.

[82] Jeffreys, H. (1961), *Theory of Probability*. 3rd Edition, Oxford: Oxford University Press.

[83] Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42(4), 805-820.

[84] Kalbfleisch, J. D., & Ross, L. Prentice. 2002. *The statistical analysis of failure time data.*

[85] Karlin, S., & Taylor, H. M. (1975). A First Course in Stochastic Processes (New York: Academic).

[86] Kim, S., Zeng, D., Chambless, L. & Li, Y. (2012). Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in biosciences*, 4(2), 262–281.

[87] Kleinbaum, D. G. & Klein, M. (2012). *Survival analysis: A self-learning text.* Third edition. Springer-Verlag, New York.

[88] Kulik, A., Lam, B. K., Rubens, F. D., Hendry, P. J., Masters, R. G., Goldstein, W., & Ruel, M. (2009). Gender differences in the long-term outcomes after valve replacement surgery. *Heart*, 95(4), 318-326.

[89] Langanay, T., Flécher, E., Fouquet, O., Ruggieri, V. G., De La Tour, B., Félix, C., & Leguerrier, A. (2012). Aortic valve replacement in the elderly: the real life. *The Annals of thoracic surgery*, 93(1), 70-78.

[90] Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963–974.

[91] Lawless, J. F. (2011). *Statistical models and methods for lifetime data* (Vol. 362). John Wiley & Sons.

[92] Lee, E.T. & Wang, J. (2003). *Statistical methods for survival data analysis*. John Wiley & Sons.

[93] Li, Q. (2014). *Statistical Inference for Joint Modelling of Longitudinal and Survival Data* (Doctoral dissertation, University of Manchester).

[94] Liang, K.-Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

[95] Lim, E., Ali, A., Theodorou, P., Sousa, I., Ashrafian, H., Chamageorgakis, T., ... & Pepper, J. (2008). Longitudinal study of the profile and predictors of left ventricular mass regression after stentless aortic valve replacement. *The Annals of thoracic surgery*, 85(6), 2026-2029.

[96] Lindman, B. R., Stewart, W. J., Pibarot, P., Hahn, R. T., Otto, C. M., Xu, K., ... & Makkar, R. (2014). Early regression of severe left ventricular hypertrophy after transcatheter aortic valve replacement is associated with decreased hospitalizations. *JACC: Cardiovascular Interventions*, 7(6), 662-673.

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

[98] Lix, L., Ayles, J., Bartholomew, S., Cooke, C., Ellison, J., Emond, V., ... & Paterson, J. M. (2018). The Canadian Chronic Disease Surveillance System: A model for collaborative surveillance. *International Journal of Population Data Science*, 3(3).

[99] Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. . *Statistics and Computing*, 10(4): 325–337.

[100] Marathe, S. P., Bell, D., Betts, K., Sayed, S., Dunne, B., Ward, C., ... & Alphonso, N. (2019). Homografts versus stentless bioprosthetic valves in the pulmonary position: a multicentre propensity-matched comparison in patients younger than 20 years. *European Journal of Cardio-Thoracic Surgery.*

[101] McCrink, Lisa. M., Marshall, A. H., & Cairns, K. J. (2013). Advances in Joint Modelling: a review of recent developments with application to the survival of end stage renal disease patients, *International Statistical Review*, 81(2): 249-269

[102] Metropolis, N. & Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247), 335–341.

[103] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21: 1087–1092.

[104] Meyn, S. P., & Tweedie, R. L. (2012). *Markov chains and stochastic stability*, Springer Science & Business Media.

[105] Murawska, M. M. (2014). *Extensions in Joint Modeling of Survival and Longitudinal Outcomes.*

[106] Nakamura, T., Toda, K., Kuratani, T., Miyagawa, S., Yoshikawa, Y., Fukushima, S., & Sawa, Y. (2016). Diabetes mellitus impairs left ventricular mass regression after surgical or transcatheter aortic valve replacement for severe aortic stenosis. *Heart, Lung and Circulation*, 25(1), 68-74.

[107] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4): 945–966.

[108] Nishimura, R. A., Otto, C. M., Bonow, R. O., Carabello, B. A., Erwin, J. P., Guyton, R. A., ... & Sundt, T. M. (2014). 2014 AHA/ACC guideline for the management of patients with valvular heart disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*, 63(22), 2438-2488.

[109] Nkomo, V. T., Gardin, J. M., Skelton, T.N., et al. (2006). Burden of valvular heart diseases: a population-based study. *The Lancet*, 368(9540): 1005–1011.

[110] Pekoz, E., & Ross, S. M. (1997). Estimating the mean cover time of a semi-Markov process via simulation. *Probability in the Engineering and Informational Sciences*, 11(2), 267-271.

[111] Philipson, P., Sousa, I., Diggle, P., Williamson, P., Kolamunnage-Dona, R., Henderson, R., & Hickey, G. (2018). joineR: Joint Modelling of Repeated Measurements and Time-to-Event Data. R package version 1.2.4, *URL: https://github.com/graemeleehickey/joineR/*.

[112] Pibarot, P., & Dumesnil, J. G. (2009). Prosthetic heart valves: selection of the optimal prosthesis and long-term management. *Circulation*, 119(7), 1034-1048.

[113] Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news*, 6(1), 7-11.

[114] Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). Numerical recipes 3rd edition: *The art of scientific computing*. Cambridge university press.

[115] R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2012.

[116] Raftery, A. E. (1999). Bayes factors and BIC: Comment on "A critique of the Bayesian information criterion for model selection". *Sociological Methods & Research*, 27(3), 411-427.

[117] Report from the Canadian Chronic Disease Surveillance System: *Heart Disease in Canada*, 2018.

[118] Rizopoulos, D., Verbeke, G. & Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95(1), 63-74.

[119] Rizopoulos, D., Verbeke, G., & Lesaffre, E. (2009) Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3), 637–654,

[120] Rizopoulos, D., Verbeke, G., & Molenberghs, G. (2010). Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics*, 66(1), 20-29.

[121] Rizopoulos, D. D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9), 1-33.

[122] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman & Hall/CRC.

[123] Rizopoulos, D. (2012). Fast Fitting of Joint Models for Longitudinal and Event Time Data Using a Pseudo-Adaptive Gaussian Quadrature rule. *Computational Statistics and Data Analysis*, 56, 491-501.

[124] Robert CP. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Second Edition. Springer LLC.

[125] Roy, V. (2019). Convergence diagnostics for Markov chain Monte Carlo. *arXiv preprint arXiv:1909.*11827.

[126] Sahn, D. J., DeMaria, A. N. T. H. O. N. Y., Kisslo, J. O. S. E. P. H., & Weyman, A. F. (1978). Recommendations regarding quantitation in M-mode echocardiography: results of a survey of echocardiographic measurements. *Circulation*, 58(6), 1072-1083.

[127] Salkind, N. J. & Rasmussen, K. (2008). *Encyclopedia of Educational Psychology*. Volume 1.Thousand oaks, California: Sage Publications, Inc.

[128] Schluchter,M.D. (1992). Methods for the analysis of informatively censored longitudinal data. *Stat. Med*, 11(1415), 1861-1870, 1992.

[129] Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461–464.

[130] Self, S. & Pawitan, Y. (1992). Modeling a marker of disease progression and onset of disease. In AIDS Epidemiology. *Springer*, 231–255. .

[131] Senn, S. (2003). Bayesian, likelihood, and frequentist approaches to statistics. *Applied Clinical Trials*, 12(8), 35-38.

[132] Shehada, S. E., Elhmidi, Y., Puluca, N., Ozturk, O., Demircioglu, E., Wendt, D., & Thielmann, M. (2017). Impact of previous cardiac surgery in patients undergoing transcatheter aortic valve implantation: a systematic review. *The Journal of cardiovascular surgery*, 58(5), 787-793.

[133] Sharabiani, M. T., Fiorentino, F., Angelini, G. D., & Patel, N. N. (2016). Long-term survival after surgical aortic valve replacement among patients over 65 years of age. *Open heart*, 3(1), e000338.

[134] Siniawski, H., Lehmkuhl, H., Weng, Y., Pasic, M., Yankah, C., Hoffmann, M., & Hetzer, R. (2003). Stentless aortic valves as an alternative to homografts for valve replacement in active infective endocarditis complicated by ring abscess. *The Annals of thoracic surgery*, 75(3), 803-808.

[135] Song, X., Davidian, M., & Tsiatis, A., A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4), 742-53.

[136] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002), Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society, Series B*, 64(4): 583–616.

[137] Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003) *WinBUGS version 1.4 User Manual*.

[138] Statisticat, L. L. C. (2015). LaplacesDemon: complete environment for Bayesian inference. *R package version*, 15(01).

[139] Stone, C. A., & Zhu, X. (2015). *Bayesian analysis of item response theory models using SAS. Sas* Institute.

[140] Sturtz, S., Ligges, U., & Gelman, A. E. (2005). R2WinBUGS: a package for running WinBUGS from R. 1-16.

[141] Sundt, T. M., Bailey, M. S., Moon, M. R., Mendeloff, E. N., Huddleston, C. B., Pasque, M. K., ... & Gay Jr, W. A. (2000). Quality of life after aortic valve replacement at the age of¿ 80 years. *Circulation*, 102(suppl 3), Iii-70.

[142] Sweeting, M. J. & Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5), 750-763.

[143] Terrera, G. M., Piccinin, A. M., Johansson, B., Matthews, F., & Hofer, S. M. (2011). Joint Modeling of Longitudinal Change and Survival: An Investigation of the Association Between Change in Memory Scores and Death. *GeroPsych*, 24(4), 177–185.

[144] Thom, H. C. (1958). A note on the gamma distribution. *Monthly Weather Review*, 86(4),117–122.

[145] Tjang, Y. S., van Hees, Y., Körfer, R., Grobbee, D. E., & van der Heijden, G. J. (2007). Predictors of mortality after aortic valve replacement. *European journal of cardio-thoracic surgery*, 32(3), 469-474.

[146] Tsamasphyros, G., & Dimou, G. (1990). Gauss quadrature rules for finite part integrals. International Journal for Numerical Methods in Engineering, 30(1), 13-26.

[147] Tseng, Y., Hsieh, F., & Wang, JL. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92(3), 587-603.

[148] Tsiatis, A. A., DeGruttola, V., & Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, 90(429), 27-37.

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

[150] Une, D., Mesana, L., Chan, V., Maklin, M., Chan, R., Masters, R. G., & Ruel, M. (2015). Clinical impact of changes in left ventricular function after aortic valve replacement: analysis from 3112 patients. *Circulation*, 132(8), 741-747.

[151] Vahanian, A., Baumgartner, H., Bax, J., Butchart, E., Dion, R., Filippatos, G., ... & Nataf, P. (2007). Guidelines on the management of valvular heart disease: The Task Force on the Management of Valvular Heart Disease of the European Society of Cardiology. *European heart journal*, 28(2), 230-268.

[152] Vallinayagam ,V., Prathap, S. & Venkatesan, P. (2014). Parametric regression models in the analysis of breast cancer survival data. *International Journal of Science and Technology*, 3(3).

[153] Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child development*, 85(3), 842-860.

[154] Verbeke, G. (1997). Linear mixed models for longitudinal data. *In Linear mixed models in practice* (pp. 63-153). Springer, New York, NY.

[155] Villa, E., Troise, G., Cirillo, M., Brunelli, F., Dalla Tomba, M., Mhagna, Z., & Quaini, E. (2006). Factors affecting left ventricular remodeling after valve replacement for aortic stenosis. An overview. *Cardiovascular ultrasound*, 4(1), 25.

[156] Viviani, S. (2012). Mixed effect joint models for longitudinal responses with drop-out: estimation and sensitivity issues.

[157] Wang, Y. and Taylor, & J. M. G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 455(96): 895–905.

[158] Wang, H., Naghavi, M., Allen, C., Barber, R. M., Bhutta, Z. A., Carter, A., ... & Coggeshall, M. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The lancet*, 388(10053), 1459-1544.

[159] Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.

[160] Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in medicine*, 11(14-15), 1871-1879.

[161] Wu M. C. & Carroll R. J., (1988). Estimation and Comparison of Changes in the Presence of Informative Right-Censoring by Modeling the Censoring Process. *Biometrics*, 44(1): 175-188.

[162] Wu, L. (2009). *Mixed effects models for complex data*. Chapman and Hall/CRC.

[163] Wu, L., Hu, X. J., & Wu, H. (2007). Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. *Biostatistics*, 9(2), 308-320.

[164] Wu, L., Liu, W. & Hu, X. J. (2010). Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. *Biometrics*, 66 (2), 327–335.

[165] Wu, L., Liu, W., Yi, G. Y., & Huang, Y. (2012). Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*.

[166] Wulfsohn, M. S., & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53 (1), 330–339.

[167] Xu, J. & Zeger, S.L. (2001). The evaluation of multiple surrogate endpoints. *Biometrics*, 57(1): 81-87.

[168] Ye, W., Lin, X., & Taylor, J. M. G. (2008). Semiparametric modeling of longitudinal measurements and time to-event data two-stage regression calibration approach. *Biometrics*, 64(4), 1238–1246.

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

[170] Yun, K. L., Sintek, C. F., Fletcher, A. D., Pfeffer, T. A., Kochamba, G. S., Hyde, M. R., ... & Khonsari, S. (1999). Aortic valve replacement with the freestyle stentless bioprosthesis: five-year experience. *Circulation*, 100(suppl 2), II-17.

[171] Zhao, X., Yu, C., & Tong, H. (2008, May). A Bayesian approach to weibull survival model for clinical randomized censoring trial based on MCMC simulation. *In 2008 2nd International Conference on Bioinformatics and Biomedical Engineering* (pp. 1181-1184). IEEE.

Appendix A

Parametric Distributions: Table 2.1

A.1 Exponential Distributions

The exponential distribution is the simplest model, and It has only one parameter ρ. ρ is the rate parameters and the scale parameters can be obtained by, $\lambda = \rho^{-1}$. Since this distribution has no information about the shape, therefore not flexible enough to describe hazard shapes for time-to-event data. However, the large value of ρ indicates short survival and high risk, where the small value of ρ indicates long survival and low risk.

A.2 Weibull Distribution

The Weibull distribution (Weibull and Stockholm 1951) is a generalization of the exponential distribution. Weibull distribution has two parameters, shape ($\kappa > 0$) and rate ($\rho > 0$) parameters. In survival analysis, Weibull distribution is the most widely used distribution.Weibull distribution can model different types of data (either skewed or symmetric distributional shape). It is particularly popular in engineering applications. The model from Weibull distribution is flexible and accommodates monotone hazard shapes. In this thesis, we use Weibull distribution to compare our proposed log-logistic distribution, both distribution under the AFT model setting.

A.3 Log-Logistic Distribution

The log-logistic distribution is particularly useful to model unimodal (i.e., non-monotone) hazard functions. This distribution has two parameters, rate$\rho > 0$ and $\kappa > 0$. ρ is the rate parameter and κ is the shape parameter. The hazard function of the log-logistic distribution is monotone decreasing for $\kappa \leq 1$ and unimodal (non-monotone) for $\kappa > 1$.

The log-logistic distribution widely used when the mortality increase and reaches a peak after some finite time point and then slowly decrease (Bennett 1983). For example, the log-logistic model can be used to describe the lifetimes of breast cancer patients (peak mortality of breast cancer patients occurs after about three years (Langlands et al. 1979)). In this thesis, our focus is on log-logistic distribution under the AFT model setting.

A.4 Logistic Distribution

Logistic distribution; $-\infty < \mu < \infty$ is the location parameter and $\sigma > 0$ is the scale parameter of distribution. Note that: log-logistic is a survival distribution, but logistic is not.

A.5 Log-normal Distribution

Another popular model to model the unimodal (i.e., non-monotone) hazard function is, log-normal distribution. $\Phi(.)$ for the distribution is the cumulative distribution function of standard normal distribution and parameters of distribution are defined as; $-\infty < \mu < \infty$ and $\sigma > 0$.
Log-Logistic and log-normal distributions have the same shape for the hazard functions.

A.6 Gamma Distribution

The gamma distribution is not widely used in survival statistical as the Weibull, log-logistic and log-normal distributions because, from a computational point of view, the survivor and the hazard functions of the gamma distribution are intractable. Despite the limitation of gamma distribution it has been used to model lifetimes of technical systems with repeated repairing after failure, rainfall data in meteorology, and insurance claims and loan data in business (Thom 1958).
For gamma distribution the rate parameter is $\rho^{-1} > 0$ and shape parameter is $\kappa > 0$. For gamma distribution the hazard function is monotone increasing function if $\kappa > 1$ and monotone decreasing for $0 < \kappa < 1$.

A.7 Gompertz Distribution

Benjamin Gompertz, a British actuary, introduced a law of mortality, which is nowadays called the Gompertz law of mortality (Conn 2006). He assumed the exponential increase of mortality with age. In the distribution, t is the follow-up time and $a > 0$ are the distribution parameters.

Appendix B

Gauss-Legendre Quadrature Rule for Numerical Integration

In numerical analysis, a quadrature rule is used for an approximation of the definite integral of a function, This approximation is usually known as a weighted sum of function values at some specified points within the domain of integration.

An n-point Gaussian quadrature rule, named after German mathematician and physicist Carl Friedrich Gauss, is a quadrature rule constructed to yield an exact result for polynomials of degree $(2n-1)$ or less by a suitable choice of the nodes x_j and weights w_j for $j = 1, ..., n$. The domain of integral for such a rule is taken as [-1,1] and defines as

$$\int_{-1}^{1} f(x)dx = \sum_{j=1}^{n} w_j f(x_j)$$

which is exact for polynomials of degree 2n-1 or less. This exact rule is known as the Gauss-Legendre quadrature rule. The quadrature rule will only be an accurate approximation to the integral above if f(x) is well-approximated by a polynomial of degree 2n-1 or less on [- 1,1]. The quadrature nodes are defined as the roots of a polynomial on [-1, 1] from a class of orthogonal polynomials and the weights are defined as

$$w_j = \frac{2}{(1-x_j^2)[(P_n'(x_j)]^2}$$

Where P_n represents the Legendre polynomials.

Moreover, if the domain of integration is different from [-1, 1], first we mustl transform the domain and then apply the rule to approximate the integration value. For example if the domain of integral is [a, b] then we can do this transformation as explained below

$$\int_a^b f(x)dx = (\frac{b-a}{2}) \int_{-1}^{1} f(\frac{b-a}{2}x + \frac{b+a}{2})dx \tag{B.1}$$

The nodes and weights for 5-point Gauss-Legendre quadrature rule are given in the following Table B.1 (for more details see Press et al. 2007).

Table B.1: Nodes and Weights of 5-point Gauss-Legendre quadrature

Nodes x_i	0.9061798459	-0.9061798459	0.5384693101	-0.5384693101	0.00
Weights w_i	0.2369268851	0.2369268851	0.4786286705	0.4786286705	0.5688888888

To evaluate the integral in equation (4.14), we use 5-point Gauss-Legendre rule. First we change the domain of integral [0, t] so that its range is from [-1, 1]. for this purpose we substitute a = 0 and b = t_i and function f(x) = $g(x)$ in equation (B.1), the 5-point Gauss-Legendre rule gives

$$\int_0^{t_i} g_i(x)dx = (\frac{t_i}{2}) \int_{-1}^{1} g(\frac{t_i}{2}x + \frac{t_i}{2})dx = \sum_{j=1}^{5} w_j g(\frac{t_i}{2}x_j + \frac{t_i}{2}) \tag{B.2}$$

In our implementation, first we compute $\frac{t_i}{2}\ x_j + \frac{t_i}{2}$ for each individual i. Then we integrate our function $g_i(t_i)$ at these values in WinBUGS.

Appendix C

Gelman-Rubin Statistics R: For Bayesian Joint AFT Log-logistic and Weibull Distribution

Gelman-Rubin Statistics for Bayesian Joint AFT Log-logistic and Weibull Distribution is given below

Table C.1: Gelman-Rubin statistics results

	Log-Logistic	**Weibull**
Parameters	**Test statistics value**	**Test statistics value**
α_1 (Intercept)	1.00	1.00
α_2 (time)	1.00	1.00
α_3 (hs)	1.00	1.00
α_4 (lv1)	1.00	1.00
α_5 (lv2)	1.00	1.00
α_6 (redo)	1.00	1.00
α_7 (dm)	1.00	1.00
α_8 (age)	1.00	1.00
α_9 (sex)	1.00	1.00
β_1 (hs)	1.00	1.00
β_2 (lv1)	1.00	1.00
β_3 (lv2)	1.00	1.00
β_4 (redo)	1.00	1.00
β_5 (dm)	1.00	1.00
β_6 (age)	1.03	1.02
β_7 (sex)	1.00	1.00
ϕ(Association)	1.00	1.00
κ(shape)	1.01	1.00
ρ (rate)	1.01	1.03

Appendix D

WinBUGS Codes to Fit Proposed Joint Models

D.1 Code to fit joint models using the Log-Logistic distribution

```
library(nlme)
library(joineR)
library(R2WinBUGS)
model.file <-function() {
for (i in 1:n) {
for (j in n1[i]: n2[i]) {
y[j] ~ dnorm(muy[j], inv.sigSqu)
muy[j] <- inprod2(alpha[1:p1], xlong[j,1: p1]) + inprod2(u[i,1:pp1], zlong[j,1:pp1])
}
zeros[i] ~ dpois(zeros.mean[i])
zeros.mean[i] <- -l[i] + const
pred0[i] <- inprod2(beta[1:p2], xsurv[i,1: p2])
pred1[i] <- pred0[i] + phi*inprod2(u[i,1:pp2], zsurv[i,1: pp2])
for(k in 1: quad.points) {
psi0[i, k] <- c15[k]*exp(-phi*inprod2(u[i,1: pp2], xx15[i,1: pp2, k]))
}
psi[i] <- (sum(psi0[i,])*st[i]/2)*exp(-pred0[i])
logf[i] <- log(kappa) + kappa*log(rho) + (kappa-1)*log(psi[i]) - 2*log(1 + pow((rho*psi[i]),kappa)) - pred1[i]
logs[i] <- -log(1 + pow((rho*psi[i]), kappa))
l[i] <- status[i]*logf[i] + (1 - status[i]) logs[i]
u[i,1:pp1] ~ dmnorm(U0[ ], inv.Sigma[,])
}
inv.Sigma[1: pp1,1: pp1] ~ dwish(R[,], w.df)
alpha[1:p1] ~ dmnorm(alpha.mu[ ], iSigma1[,])
beta[1:p2] ~ dmnorm(beta.mu[ ], iSigma2[,])
phi ~ dnorm(prior.phi.mu, prior.phi.tau)
inv.sigSqu ~ dgamma(prior.tauz1, prior.tauz2)
kappa1~ dgamma(prior.kappa1, prior.kappa2)
rho ~ dgamma(prior.rho1, prior.rho2)
kappa <- 1/kappa1
logkappa <- log(kappa)
logrho <- log(rho)
}
bugs.model <- file.path(tempdir(), “model.file.bug”)
write.model(model.file, bugs.model)
file.show(bugs.model)

    ######## DATA ########
    data(heart.valve)
lv1 <- ifelse(heart.valve$lv==1,1,0)
lv2 <- ifelse(heart.valve$lv==2,1,0)
heartvalve1 <- cbind (heart.valve, lv1,lv2)
heartvalve1$hs <- ifelse(as.numeric(heartvalve1$hs)==2,0,1)

    ################
```

```
    lmeFit <- lme(log.lvmi ~1 + time + hs + lv1 + lv2 + redo + dm + age + sex ,
random = ~ time|num , data = heartvalve1)
summary(lmeFit)

    Fixed_Estimate <- fixed.effects(summary(lmeFit))
Random_Estimate <- as.matrix(ranef(lmeFit))

    heartvalve.id <- heartvalve1[!duplicated(heartvalve1$num), ]

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

    SurvFit <- survreg(Surv(fuyrs, status)~ hs + lv1 + lv2 + + redo + dm + age + sex ,
data = heartvalve.id, x = TRUE, dist = "loglogistic" )
summary(SurvFit)

    ########Data preparation for WinBUGS model #########

    y <- heartvalve1$log.lvmi
obs <- as.numeric(row.names.data.frame(heartvalve1))
num <- as.numeric(heartvalve1$num)
n <- length(unique(heartvalve1$num))
n1 <- as.numeric(tapply(obs,num,function(x) x[1]))
n2 <- as.numeric(tapply(obs,num,function(x) x[length(x)]))
xlong0 <- data.frame(heartvalve1, int=1)
xlong <- xlong0[, c("int","time","hs" , "lv1", "lv2", "redo", "dm" , "age", "sex")]
p1 <- ncol(xlong)
zlong <- xlong0[, c("int","time") ]
pp1 <- ncol (zlong)
SurvCov <- UniqueVariables(xlong0,c("int","hs" , "lv1", "lv2", "redo", "dm" ,
"age", "sex", "fuyrs","status"), id.col="num")
xsurv <- SurvCov[, c("hs" , "lv1", "lv2", "redo", "dm" , "age", "sex")]
zsurv <- SurvCov [, c("int", "fuyrs")]
st <- SurvCov [, "fuyrs"]
status <- SurvCov [, "status"]
p2 <- ncol(xsurv)
pp2 <- ncol(zsurv)

    ######### 5 Points Gauss-Legender quadrature rule ########

    ######### Weights ########

    c15 <- c(0.2369268850561891,0.2369268850561891, 0.4786286704993665,
0.4786286704993665, 0.5688888888888889)

    ######### Nodes ########

    t15 <- c(0.9061798459386640, -0.9061798459386640, 0.5384693101056831,
-0.5384693101056831, 0.0000000000000000)

    quad.points <- length(t15)
x150 <- sapply (st,function(st){0.5*(st*t15 + st)})
x15 <- t(x150)
zsurv2 <- cbind(zsurv[,"int"], x15)
```

```
xx150 <- cbind (zsurv2[, c (1,2)], zsurv2[, c (1,3)], zsurv2[, c (1,4)], zsurv2[, c (1,5)], zsurv2[, c (1,6)])
xx15 <- array (xx150, dim=c(n, pp2, quad.points))

    ######## Initialize longitudinal parameters ########

    u_init <- Random_Estimate
alpha_init <- Fixed_Estimate

    ######## Initialize survival parameters ########

    beta_init = SurvFit$coeff[-1]
init_kappa <- 1/(SurvFit$scale)
init_rho <- exp(-SurvFit$coeff[1])
init_phi <- 0
inv.Sigma.u_init <- diag(0.1,pp1)
data <- list(n=n,n1=n1,n2=n2,p1=p1,p2=p2,pp1=pp1,pp2=pp2,y=y, xlong=as.matrix(xlong),
zlong=as.matrix(zlong), st=st,status=status,xsurv=as.matrix(xsurv), zsurv=as.matrix(zsurv),
zeros=rep(0,n), const=0,R=diag(0.01, pp1),iSigma1=diag(0.00001, p1), iSigma2=diag(0.00001, p2),
alpha.mu=rep(0,p1), beta.mu=rep(0, p2),U0=rep(0,pp1),w.df=pp1+1, prior.phi.mu=0,
prior.phi.tau=0.0001, prior.tauz1=0.1,prior.tauz2=0.1, prior.kappa1=0.1,prior.kappa2=0.1,
prior.rho1=0.1,prior.rho2=0.1, xx15=xx15,c15=c15,quad.points=5)

    parameters.bugs <- c("alpha", "beta", "phi", "kappa", "rho", "logkappa", "logrho", "inv.sigSqu",
"inv.Sigma", "deviance", "logs", "u")

    inits.bugs <- rep(list(list(alpha=alpha_init,beta=beta_init,phi=init_phi,inv.sigSqu=1,
inv.Sigma=inv.Sigma.u_init,kappa1=1/init_kappa,rho=init_rho,u=u_init)),2)

    heartvalve1.sim <- bugs(data=data, inits=inits.bugs, parameters.to.save=parameters.bugs,bugs.model,
n.chains=2, n.iter=150000, n.burnin=15000, n.thin=5, DIC=TRUE, program="WinBUGS",
bugs.directory="C:/winBUGS14")

    summary(heartvalve1.sim )
    chain1 <- heartvalve1.sim$sims.array[,1,]
chain2 <- heartvalve1.sim$sims.array[,2,]

    mcmc.sam <- mcmc.list(mcmc(chain1),mcmc(chain2))
summary (mcmc.sam)

    gelman.diag (mcmc.sam, autoburnin=TRUE, multivariate=FALSE)

    plot (mcmc.sam,density=F,trace=T, ask=TRUE)

    plot (mcmc.sam,density=T,trace=F, ask=TRUE)

    Print(heartvalve1.sim)
```

D.2 Code to fit joint models using the Weibull distribution

```
library(nlme)
library(joineR)
library(R2WinBUGS)
model.file<- function(){
```

```
for (i in 1:n) {
for (j in n1[i]: n2[i]) {
y[j] ~ dnorm(muy[j], inv.sigSqu)
muy[j] <- inprod2(alpha[1:p1], xlong[j,1: p1]) + inprod2(u[i,1: pp1], zlong[j,1: pp1])
}
zeros[i] ~ dpois(zeros.mean[i])
zeros.mean[i] <- - l[i] + const
pred0[i] <- inprod2(beta[1: p2], xsurv[i,1: p2])
pred1[i] <- pred0[i] + phi*inprod2(u[i,1: pp2], zsurv[i,1: pp2])
for(k in 1:quad.points){
psi0[i,k] <- c15[k]*exp(- phi*inprod2(u[i,1:pp2], xx15[i,1:pp2,k]))
}
psi[i] <- (sum(psi0[i,])*st[i]/2)*exp(-pred0[i])
logf[i] <- log(kappa) + kappa*log(rho) + (kappa-1)*log(psi[i]) - pow((rho*psi[i]),kappa) - pred1[i]
logs[i] <- - pow((rho*psi[i]), kappa)
l[i] <- status[i]*logf[i]+(1 - status[i])*logs[i]
u[i,1:pp1] ~ dmnorm(U0[], inv.Sigma[,])
}
inv.Sigma[1: pp1,1: pp1] ~ dwish(R[,], w.df)
alpha[1: p1] ~ dmnorm(alpha.mu[], iSigma1[,])
beta[1:p2] ~ dmnorm(beta.mu[], iSigma2[,])
phi ~ dnorm(prior.phi.mu, prior.phi.tau)
inv.sigSqu ~ dgamma(prior.tauz1, prior.tauz2)
kappa1 ~ dgamma(prior.kappa1, prior.kappa2)
rho ~ dgamma(prior.rho1, prior.rho2)
kappa <- 1/kappa1
logkappa <- log(kappa)
logrho <- log(rho)
}

    bugs.model <- file.path(tempdir(), "model.file.bug")
write.model(model.file, bugs.model)
file.show(bugs.model)

    ######## DATA ########
    data(heart.valve)
lv1 <- ifelse(heart.valve$lv==1,1,0)
lv2 <- ifelse(heart.valve$lv==2,1,0)
heartvalve1 <- cbind (heart.valve, lv1,lv2)
heartvalve1$hs <- ifelse(as.numeric(heartvalve1$hs)==2,0,1)

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

    lmeFit <- lme(log.lvmi ~1 + time + hs + lv1 + lv2 + redo + dm + age + sex ,
random = ~ time|num , data = heartvalve1)
summary(lmeFit)

    Fixed_Estimate <- fixed.effects(summary(lmeFit))
Random_Estimate <- as.matrix(ranef(lmeFit))

    heartvalve.id <- heartvalve1[!duplicated(heartvalve1$num), ]

    ################
```

```
    SurvFit <- survreg(Surv(fuyrs, status) ~ hs + lv1 + lv2 + + redo + dm + age + sex ,
data = heartvalve.id, x = TRUE, dist = "weibull")
summary(SurvFit)

    ######## Data preparation for WinBUGS model ########

    y <- heartvalve1$log.lvmi
obs <- as.numeric(row.names.data.frame(heartvalve1))
num <- as.numeric(heartvalve1$num)
n <- length(unique(heartvalve1$num))
n1 <- as.numeric(tapply(obs,num,function(x) x[1]))
n2 <- as.numeric(tapply(obs,num,function(x) x[length(x)]))
xlong0 <- data.frame(heartvalve1, int=1)
xlong <- xlong0[, c("int","time","hs" , "lv1", "lv2", "redo", "dm" , "age", "sex")]
p1 <- ncol(xlong)
zlong <- xlong0[, c("int","time") ]
pp1 <- ncol (zlong)
SurvCov <- UniqueVariables(xlong0,c("int","hs" , "lv1", "lv2", "redo", "dm" ,
"age", "sex", "fuyrs","status"), id.col="num")
xsurv <- SurvCov[, c("hs" , "lv1", "lv2", "redo", "dm" , "age", "sex")]
zsurv <- SurvCov [, c("int", "fuyrs")]
st <- SurvCov [, "fuyrs"]
status <- SurvCov [, "status"]
p2 <- ncol(xsurv)
pp2 <- ncol(zsurv)

    ######### 5 Points Gauss-Legender quadrature rule ########

    ######## Weights ########

    c15 <- c(0.2369268850561891,0.2369268850561891, 0.4786286704993665,
0.4786286704993665, 0.5688888888888889)

    ######## Nodes ########

    t15 <- c(0.9061798459386640, -0.9061798459386640, 0.5384693101056831,
-0.5384693101056831, 0.0000000000000000)

    quad.points <- length(t15)
x150 <- sapply (st,function(st){0.5*(st*t15 + st)})
x15 <- t(x150)
zsurv2 <- cbind(zsurv[,"int"], x15)
xx150 <- cbind (zsurv2[, c (1,2)], zsurv2[, c (1,3)], zsurv2[, c (1,4)], zsurv2[, c (1,5)], zsurv2[, c (1,6)])
xx15 <- array (xx150, dim=c(n, pp2, quad.points))

    ######## Initialize longitudinal parameters ########

    u_init <- Random_Estimate
alpha_init <- Fixed_Estimate

    ######## Initialize survival parameters ########

    beta_init = SurvFit$coeff[-1]
init_kappa <- 1/(SurvFit$scale)
```

```
init_rho <- exp(-SurvFit$coeff[1])
init_phi <- 0
inv.Sigma.u_init <- diag(0.1,pp1)
data <- list(n=n,n1=n1,n2=n2,p1=p1,p2=p2,pp1=pp1,pp2=pp2,y=y, xlong=as.matrix(xlong),
zlong=as.matrix(zlong), st=st,status=status,xsurv=as.matrix(xsurv), zsurv=as.matrix(zsurv),
zeros=rep(0,n), const=0,R=diag(0.01, pp1),iSigma1=diag(0.00001, p1), iSigma2=diag(0.00001, p2),
alpha.mu=rep(0,p1), beta.mu=rep(0, p2),U0=rep(0,pp1),w.df=pp1+1, prior.phi.mu=0,
prior.phi.tau=0.0001, prior.tauz1=0.1,prior.tauz2=0.1, prior.kappa1=0.1,prior.kappa2=0.1,
prior.rho1=0.1,prior.rho2=0.1, xx15=xx15,c15=c15,quad.points=5)

    parameters.bugs <- c("alpha", "beta", "phi", "kappa", "rho", "logkappa", "logrho", "inv.sigSqu",
"inv.Sigma", "deviance", "logs", "u")

    inits.bugs <- rep(list(list(alpha=alpha_init,beta=beta_init,phi=init_phi,inv.sigSqu=1,
inv.Sigma=inv.Sigma.u_init,kappa1=1/init_kappa,rho=init_rho,u=u_init)),2)

    heartvalve1.sim <- bugs(data=data, inits=inits.bugs, parameters.to.save=parameters.bugs,bugs.model,
n.chains=2, n.iter=150000, n.burnin=15000, n.thin=5, DIC=TRUE, program="WinBUGS",
bugs.directory="C:/winBUGS14")

    summary(heartvalve1.sim)

    chain1 <- heartvalve1.sim$sims.array[,1,]
chain2 <- heartvalve1.sim$sims.array[,2,]

    mcmc.sim <- mcmc.list(mcmc(chain1),mcmc(chain2))
summary(mcmc.sim)

    gelman.diag(mcmc.sim, autoburnin=TRUE, multivariate=FALSE)

    plot(mcmc.sim,density=F,trace=T, ask=TRUE)

    plot(mcmc.sim,density=T,trace=F, ask=TRUE)

    Print(heartvalve1.sim)
```