Statistical Inferences Using Competing Risks Data

by

Afaf Alzahrani

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2018

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In this thesis, we analyse competing risks data using the two-parameter bathtub (TPBT) distribution. The hazard rate of the TPBT distribution can be either increasing or a bathtub-shaped, which allows it to be a good fit for several data sets. In competing risks data, it is assumed that the object (system) is under attack of many risks (causes of failure) that compete to destroy it. In this study, we assume that the system will be destroyed by only one cause and all risks are independent. We discuss two models. The first does not allow covariates while the second does. We used the maximum likelihood and Bayes methods to estimate the model parameters, the relative risks and some of the reliability measures of the system. The likelihood equations of the unknown parameters have no analytic solution and numerical methods will be used to get the maximum likelihood estimations. Also, the posterior distribution of the parameters is not in a convenient form, therefore we used Markov Chain Monte Carlo (MCMC) method to simulate random draws from the posterior distribution and then use it to obtain the Bayes estimates of the parameters, the relative risks and the system's reliability measures. Furthermore, to study the performance of the two estimation techniques used, we provided a simulation study. This paper is illustrated on two real datasets.

LIST OF ABBREVIATIONS AND SYMBOLS USED

AIDS	Acquired immune deficiency syndrome
CCR5	C-C chemokine receptor type 5
CR	Competing Risks model
CRR	Competing Risks Regression model
CXCR4	C-X-C chemokine receptor 4
HIV	Human immunodeficiency virus infection
MCMC	Markov chain Monte Carlo method
MLE	Maximum likelihood estimator
SI	Syncytium Inducing (SI) HIV phenotype
TPBT	Two-parameter bathtub distribution

ACKNOWLEDGMENTS

I would like to thanks the Saudi Bureau for giving me the scholarship for all the length of my study. Also, I would like to thank my supervisor Dr. Sarhan for his support and advice during my research. He is the great supervisor and has excellent teaching techniques. I am so proud to work with him because I learned a lot from him about Bayesian inference. I want to say thanks to my family especially my dad and my mom for their support which made me so happy during my master degree. Also, I would like to say thanks for all my friends', especially who have communicated with me.

Chapter 1: Introduction

The statistical analysis of lifetime data, which has been referred to as survival time or failure time, is essential in many fields, including the biomedical, engineering, and social sciences (Lawless, 2011). The most common applications of lifetime distribution are to study humans' diseases, treatment, and components (Lawless, 2011). As an example of disease, human immunodeficiency virus (HIV) is a blood-borne pathogen transmitted primarily through unprotected sexual intercourse (Bertozzi et al., 2006). It has been found that there is a much higher risk of HIV infection in men who have sex with men (MSM) than in heterosexual couples possibly due to the increased prevalence of anal intercourse (Mayer et al., 2013). The mucosal tissue in the rectum is thinner than the vaginal mucosa and susceptible to micro-tears due to friction during intercourse, which elevates the risk of HIV viral infection (Zhou et al., 2013). Moreover, the rates of unprotected intercourse and having multiple sexual partners appear to be higher in the MSM population, serving as additional risk factors for HIV transmission (Newcomb et al., 2014). Since 1970, the statistical analysis of lifetime data has been developed, including the methodology, theory, and fields of application (Lawless, 2011).

In the competing risks model, it is assumed that the data contain the time-to-event and an indicator presents the type of event, which can be either the cause of failure or censored. There are two types of the life data including the complete data and censored data. The complete data consider all the information about time-to-failure for each subject in the lifetime test. In the case of censored data, there are different types of censoring scheme, such as right-censoring, left-censoring, and interval-censoring; however, Type-I and Type-II censoring schemes are the most commonly used in practice. On the one hand, the Type-I censoring schemes can describe the situation when the experiment continues up to a pre-specified time. On the other hand, the type-II censoring schemes can describe the situation when the experiment continues until a pre-specified number of failures occur.

For statistical inferences, many methods can be used to analyze lifetime data for interpreting research accurately and drawing appropriate conclusions based on the competing risks model assumptions, by using both parametric and non-parametric setups. To apply the parametric setups, suppose each (survival time or failure time) follows a specific parametric lifetime distribution. Statistical inference uses different competing risks models to determine how each subject is at the risk of failure due to different possible causes (Laake & Fagerland, 2015) when the occurrence of one cause of failure precludes all other causes of failure from occurring. The competing risk system can represent risk factors such as death, disease, treatment, etc.

1.1 Statement of the problem

The hazard rate function plays an important role in analysing the lifetime data because it can be either increasing-shaped or bathtub-shaped, which allows it to be a good fit for several data sets. Distribution with bathtub-shaped hazard function is common when following an individual life from actual birth to death. In competing risks data, it is assumed that the subject (system) is under attack by many risks (causes of failure) that compete to destroy it. In this study, we assume that only one cause can destroy the object and all risks are independent, but when the object has not received any attack during the study period, it will be considered as a censored observation. The maximum likelihood and Markov Chain Monte Carlo (MCMC) methods will be used to estimate the unknown parameters included in competing risks models without and with covariates when the risks follow TPBT distribution with different parameters and the relative risk rates of each cause of failure in the presence of all other causes. Also, some of the reliability measures of the system can be used to analyze the lifetime data such as cumulative distribution function (CDF) and probability density function (PDF), which can be defined as,

$$F(t) = P(T \leq t) = \int_0^t f(u)du$$

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}$$

The survival and hazard functions can be defined as,

$$S(t) = P(T > t) = \int_t^\infty f(u)du = 1 - F(t)$$

$$h(t) = \lim_{\Delta t \to 0} \left\{ \frac{\mathrm{P(t \leq T < t + \Delta t | T \geq t)}}{\Delta \mathrm{t}} \right\}$$

The relationships between the reliability measures of the system, which are common in practical situations, can be written as following:

The relationship between the hazard function and survival function can be obtained as

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{f(t)}{1 - F(t)} \\ &= -\tfrac{d}{dt}\ln(1 - F(t)) \\ &= -\frac{d}{dt}\ln(S(t)) \end{aligned}$$

The cumulative hazard function

$$\mathrm{H(t)} = \int_0^t \mathrm{h(u)du}$$

$$= -\ln\big(1 = F(t)\big) = \, -\ln S(t)$$

So,

$$S(t) = e^{-H(t)}$$

The probability density function (PDF) can be obtain by using the following formula

$$f(t) = \frac{dF(t)}{dt} \;\; = - \, \frac{dR(t)}{dt}$$

$$f(t) = h(t)e^{-H(t)}$$

So, in order to be able to evaluate and compare the results from MCMC and the maximum likelihood methods, a simulation study will be used. Additionally, this paper will include two real-life data sets. The first data set is from the Amsterdam Cohort Studies on HIV infection and AIDS. The second data set is the survival time of electrical appliances. All analyses and the simulation study are performed using the statistical software R. Under specific assumptions, the competing risks models without and with covariates will be explained to analyse competing risks data using the cause-specific hazard function.

1.1.1 Competing risks without covariates

The cause-specific hazards function is an essential measure in competing risks setting. The hazard function of the TPBT distribution will be used because it can be either increasing when the shape of $\lambda \geq 1$ or bathtub-shaped when the shape of $\lambda < 1$, to analyze competing risks data. In these cases, the data consider only the time to event and the status. As an example, the data from the Amsterdam Cohort Studies on HIV infection and AIDS will be analyzed. In the case of this study, it will be assumed that in competing risks system there are only two causes of failure, and these causes of failure are assumed to be independent. As an example, the high risk of human immunodeficiency virus (HIV) infection for men who have sex with men (MSM) leads to physical health problems such as Acquired Immunodeficiency Syndrome (AIDS) and Syncytium Inducing (SI) HIV phenotype which are considered as causes of failure.

1.1.2 Competing risks with covariates

Under the competing risks setting, Cox regression model will be used to estimate the effect of covariates on the cause-specific hazard function. For example, the data from the Amsterdam Cohort Studies on HIV infection and AIDS will be analysed. In this case, we have only two causes of failure, which are assumed to be independent and known as Acquired Immunodeficiency Syndrome (AIDS) and Syncytium-Inducing (SI) HIV phenotype each with coefficients of the CCR5 genotype on HIV infection and age at HIV infection.

1.2 Literature Review

Fundamentally, the competing risks problem has been discussed in a lot of statistical literature. The competing risks problem from different angles has been discussed in various (see, e.g. Crowder, 2001; Pintilie, 2006; Beyersmann et al., 2012; Marubini & Valsecchi, 1995; Kalbfleisch & Prentice, 2002; Klein & Moeschberger, 2003). Pintilie (2011) states that since the 18th century, competing risks problem has been part of survival analysis that is used to analyze an event. Hence, Daniel Bernoulli

(1760) was the first one who used the competing risks model to estimate the mortality rate when smallpox appeared as the cause of death (Chiang, 1991).

Competing risks problem arises in different fields such as medical and engineering that have been discussed by many authors (Haller, Schmidt & Ulm, 2013). For example, according to Pepe & Mori (1993), competing risks is widely used in medical research particularly in studying cancer. The event of interest is the time from treatment initiation to tumor-related death, but if death occurs from other causes such as cardiovascular disease, it is considered as a competing event. Another example can be seen in engineering studies, where the analysis of a series of systems shows if one component in the system fails, it leads to the failure of all systems.

According to Pintilie (2006) and Beyersmann et al (2012), in statistical literature, there are two approaches that can be used to present competing risks data such as the latent failure times approach and bivariate variables approach. To apply the latent failure times approach to present competing risks problem, their random variables $T_1, T_2, \dots, T_K$ are assumed for the time to different risks that lead to failure. This approach considers minimum time to failure, $T = \min\{T_1, T_2, \dots, T_K\}$, which means only the observed time to the first cause of failure. Consequently, as Gichangi & Vach (2005) put it, it is an indicator of variable to present the type of the observed event.

Another approach that can be used to represent competing risks data is a bivariate random variable for each individual subject on life test, which includes two random variables (T, C) T is represent the lifetime of subject and C is an indicator for the types of case of failure, C$\in\{1,2,\dots,K\}$ or if subject does not fail by these risks and will have censored, then δ is an indicator represents the event type

$$\delta = \begin{cases} 1, & failure \\ 0, & \text{censored} \end{cases}$$

In fact, there are different reasons that can be explained why censored time event occurs. For example, as Noordzij & Leffondré et al. (2013) explained occasionally during the study, a patient, might lost to follow-up before the end of the study period, for any plausible reason such as migration. Additionally, the experiment period may end before each individual subject has experienced the event of interest or may have experienced a different type of event that is not being considered in the study, which makes the follow-up difficult (Noordzij et al., 2013).

Noordzij et al. (2013) further explain that researchers have used competing risks analysis in different studies to give more information about causes of failure that can occur. These can help them to make a design to face any problem, especially in medical studies. The important quantities in competing risks problem are the cause-specific hazard (CSH) and the cumulative incidence function (CIF). Additionally, according to Gray (1988) and Pepe (1991), various authors mention differences between the effects of covariates on the CIF and CSH functions when focusing on a specific risk.

In this study, we will consider only two cases of the cause-specific hazard function to present competing risks data. In practical medical applications, for the analysis of competing risks data, typically, the cause-specific hazard function is used to estimate each risk. Similarly, in the case of studying regression model, the covariates are dependent on the cause-specific hazard function that is used to present the competing risks data. Recently, various medical researchers applied the CSH function with the independent assumption to analyse competing risks data, in case, the CSH function is based on regression that is used to study the association between covariates. For example, according to the World Health Organization (WHO), many people who have obesity will have increased risks of other diseases which negatively affect their health. According to Berrington de Gonzalez et al. (2010), Flegal et al. (2013), Gupta et al. (2014), Haque et al. (2014), and Wu et al. (2014), the relationships between obesity and other diseases such as breast cancer, cardiovascular disease, diabetes and muscular disorders are significant. As another example, Putter et al. (2007) analyzed the data from the Amsterdam Cohort Studies on HIV infection and AIDS, and the total participants in this study are 329 men who have sex with men by using non-parametric setups. These data can be considered as competing risks data with two risks as AIDS and syncytium-inducing (SI) HIV phenotype. There are some individuals in the study left with no infection switch or death. Those are considered as censored observations. It estimated the relative risk rates of each cause of failure in the presence of all other by using the Kaplan–Meier estimate. Additionally, the regression approaches was used to estimate the effect of covariates on the cause-specific hazard function and the cumulative incidence function.

For statistical inferences, many authors have applied the maximum likelihood method to estimate unknown parameters of the competing risks model such as Chen (2000) and Sarhan et al. (2010). Another approach in statistical literature is Bayesian analysis, the ideas of which date back to Thomas Bayes in the 18th-century, which are known as "Bayes' theorem" for deriving the posterior distribution. Bayesian methods started in 1990, and with time developed to the point that they have many more applications in the leading-edge research.

According to Ashby & Smith (2000), generally, the goal of using Bayesian inference is to obtain the posterior distribution to do all analyses on the unknown parameters of interest. In Bayesian epidemiology, many applications cover the cancer disease, routine data, case-control and cohort studies. Breslow (1990) and Ashby & Hutton (1996) further elaborate on this method. According to them, in Bayesian applications of epidemiology, all the analysis focuses on the description of the design study and explain the relationship between the variables, which can help in understanding the problem especially when the number of covariates is large as Ashby & Smith (2000) state.

1.3 Outline of Dissertation

In chapter 2, the competing risks problem will be explored with reference to two cases. In the first case, the competing risks model, which is dependent on the cause-specific hazard function, will be considered. In the second case, the regression model will be used to present competing risks data when the cause-specific hazard function is based on covariates. Furthermore, the important factors in competing risks analysis will be considered. Some parametric distributions will also be discussed and the approximate estimate for each case to present competing risks data will be included. In chapter 3, the concepts of Bayesian inference will be described. In section 3.2, the prior distributions for each unknown parameters in the competing risks model are assumed and the posterior distribution is obtained. In section 3.3, the prior distributions for each unknown parameters in the competing risks regression model are assumed, and the posterior distribution was assumed. Similarly, Normal approximation and MCMC methods are explained in section 3.4 and 3.5. The simulation study will be applied using the MCMC and the maximum likelihood methods in section 3.6. In chapter 4, a particular problem of HIV infection in the Amsterdam Cohort Studies, about men who have sex with men (MSM), will be explained. The competing risks models without and with covariates, which will be discussed in chapter 2, will be used to analyse the AIDSSI dataset. Additionally, setting up a Bayesian problem in R for both competing risks and regression models to present competing risks data will be described. In section 4.2, we will describe the AIDSSI dataset for both cases of this study. The normal approximation and MCMC methods will be applied for both cases. In section 4.3, the data set "survival time of electrical appliances" will be analysed under the cause specific hazard function. All results of both data sets will be discussed in section 4.4. In chapter 5, the main ideas for competing risks problem on both cases will be summarized and some future work on competing risks problem will be proposed with reference to bladder cancer studies.

Chapter 2: Competing risks problem

2.1 Introduction

Competing risks problem arise when a subject is under risk of K different types of cause of failure. The risk of K different types of cause of failure compete to attack or kill the subject, but the subject in life test experiment can fail only once by one of any risks. The causes of failure can be independent or dependent based on the assumptions of the study. Figure 2.1 shows the general competing risks system of one subject under multiple risks. We will explore the competing risks problem under two cases. In the first case, we consider the competing risks model, when dependent only on the cause-specific hazard function. In the second case, we consider regression model, where two different approaches can be used to present the competing risks data, such as the cause-specific hazard function (CSHF) and the cumulative incidence function (CIF), where both approaches depend on covariates (Dignam et al., 2012). Regression approaches are used to analyse competing risks data in epidemiologic research, (Lau, Cole & Gange, 2009). Additionally, competing risks regression models are used in clinical cancer research (e.g. Dignam, Zhang & Kocherginsky; Chappell, 2012).

In the competing risks model, it is assumed that the data contain the time-to-event and an indicator presents the type of event, which can be either the cause of failure or censored. In the case of censored data, there are different types of censoring, such as right-censored, left-censored, and interval-censored, however Type-I and Type-II censoring schemes are the most commonly used in practice.On one hand, the Type-I censoring schemes can describe the situation when the experiment continues up to a pre-specified time. While the type-II censoring schemes can describe the situation when the experiment continues until a pre-specified number of failures occur.

So, in the case of regression model, the data have similar assumptions as in competing risks model but with covariates to exploit the effect on the lifetime. Several authors have studied competing risks model by using both parametric, and non-parametric setups. To apply parametric setups, suppose each (survival time or failure time) follows a specific parametric lifetime distribution, such as the exponential, gamma, Weibull, generalized exponential or exponentiated Weibull; there are many studies about the parametric setups (see e.g. Berkson & Elveback, 1960; Cox, 1959; David & Moeschberger, 1978; Kundu & Basu, 2000; Park, 2005; Kundu & Sarhan, 2006; Sarhan, 2007; Sarhan et al., 2010). To apply the non-parametric setups do not assume the survival time or failure time follow specific distribution. Numerous studies in the non-parametric method have been carried out by several researchers (e.g. Efron, 1987; Kaplan & Meier, 1958; Peterson, 1977). The analysis of competing risks data in both cases that are mentioned above in various studies, such as the clinical, epidemiologic, demographic, basic science, and industrial literature, (Prentice et al., 1978).

In section 2.2, the concept of competing risks setting, which is important to use in analysis of competing risks data will be discussed. In section 2.3, the properties of the two-parameter bathtub distribution will be discussed. In section 2.4, a competing risks problem for each case under specific assumptions will be explained. In section 2.5, the relative risk rate will be discussed under competing risks model and competing risks regression model. In section 2.6, the maximum likelihood estimation will be described in general and the likelihood functions for competing risks model and competing risks regression model will be provided.

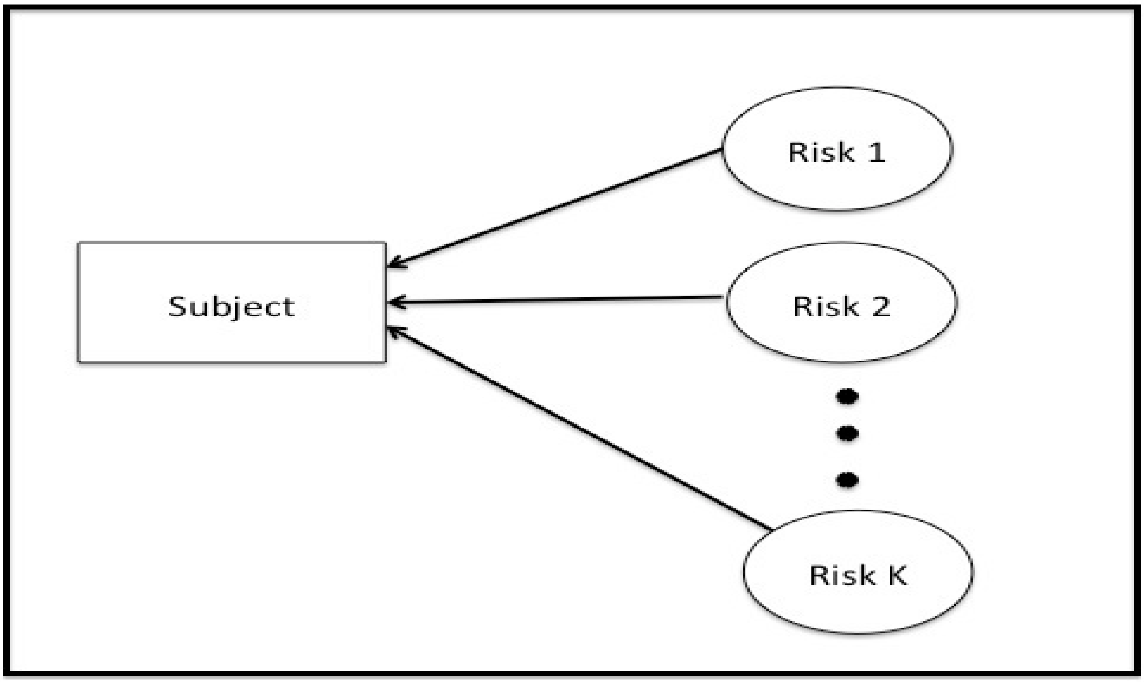


Figure 1: The competing risks model with K different types of risks.

2.2 Measurement in competing risks setting

Each object in an environment can fail due to several risks but in a competing risks system only one specific cause of failure can occur. In case of failure, we observe a pair of two quantities (T, C); T is a positive random variable that denotes the time to failure and C represents the type of cause of failure.

In this section, we describe the most important concepts that are used to analyse competing risks data. The most important functions that can be used to analyse competing risks data are the survival function and hazard rate function. The cause-specific density represents the probability of the risk that has an effects a subject in the competing risks model that leads to the death at time t by cause j, j=1,…,K,

$$f_j(t) = \lim_{\Delta t \to 0} \{\frac{P(t \leq T < t + \Delta t, C = j)}{\Delta t}\}$$

$$= \mathrm{h_j(t)\,S(t)}$$

The total probability density function of the individual under K causes of failure (risks) is

$$\mathrm{f(t)} = \sum_{\mathrm{j=1}}^{\mathrm{k}} \mathrm{f_j(t)}.$$

The cause-specific hazard rate for event type j, $\mathrm{h_j(t)}$, provides an individual's probability of failing from an cause j in a small time interval t to $\mathrm{t + \Delta t}$. Additionally, the cause-specific hazard rate describes the failure by cause j at time t for an individual, given that the individual has survived up to time t. That is,

$$\mathrm{h_j(t)} = \lim_{\Delta \mathrm{t} \to 0} \left\{ \frac{\mathrm{P(t \leq T < t + \Delta t, C = j | T \geq t)}}{\Delta \mathrm{t}} \right\}$$

The cause-specific hazards for all K risks at time t sum up to the overall hazard rate for failing from any cause at t

$$\mathbf{h_{ov.}(t)} = \sum_{\mathbf{j=1}}^{\mathbf{k}} \mathbf{h_j(t)}$$

The cumulative cause-specific hazard rate for cause j at time t is the integral over the cause-specific hazard function from zero to t

$$\mathrm{H}_{j}\mathrm{(t)} = \int_0^{\mathrm{t}} \mathrm{h_j(u)\,du}$$

The overall survivor function, $S_{ov.}$(t), denoting the probability of being free from any failure up to time t, depends on the (cumulative) cause-specific hazard functions for all K types of risks, which sums up to the overall (cumulative) hazard rate

$$\mathrm{S}_{ov.}\mathrm{(t)} = \mathrm{e}^{-\mathrm{H}_{ov.}\mathrm{(t)}}$$

2.3 Causes distributions

Let T_j be the time at which the risk from cause j might hit the individual, j=1,...,K. The random variable T_j has a lifetime distribution with a probability density function (pdf), say $f_j(t)$, hazard rate function $h_j(t)$, survival function $s_j(t)$ and cumulative distribution function $F_j(t)$. These all four functions are related to each other. One we know one of them, we can get the others. Below are the relationships between these functions:

Given the pdf $f_j(t)$, the cdf is,

$$\boldsymbol{F_j(t) = P(T_j \leq t) = \int_0^t f_j(u)\, du},$$

the survival function is

$$\boldsymbol{s_j(t) = P(T_j > t) = 1 - F_j(t)}$$

the hazard function is

$$\boldsymbol{h_j(t) = \frac{f_j(t)}{s_j(t)}} \quad .$$

Or, given the cdf $\boldsymbol{F_j(t)}$, then the pdf is

$$\boldsymbol{f_j(t) = \frac{dF_j(t)}{dt}}$$

then we can use all above relations to set the rest of function. In this study, the properties for the two-parameter bathtub distribution (TPBT) will be introduced. According to Sarhan et al. (2010) in "Reliability Engineering and System Safety" the Chen distribution is used in competing risks model to estimate unknown parameters using maximum likelihood method. In this study, we will use the same distribution but called the two-parameter bathtub distribution (TPBT). Additionally, Chen et al. (2000) explained that when analyzing lifetime data the hazard rate can be either increasing or have bathtub shape. In the case of the competing risks problem, we assumed independent identical distribution for each lifetime due to cause j, j=1,2…,k.

2.3.1 Two-parameter bathtub distribution

The two-parameter bathtub distribution is defined to have a two-parameter shape α_j and λ_j , $j = 1,2,..,K,$ and the probability density function of the TPBT distribution (pdf), $\boldsymbol{f_j(t)},$ is

$$\boldsymbol{f_j(t) = \alpha_j\lambda_j\, t^{\lambda_j-1}\, e^{t^{\lambda_j}}\;\; e^{\alpha_j(1-e^{t^{\lambda_j}})}} \quad \boldsymbol{,t > 0, \alpha_j, \beta_j > 0}$$

According to Chen (2000), the hazard function, $\boldsymbol{h_j(t)}$, has a bathtub shape when $\boldsymbol{\lambda_j < 1}$ or a increasing shape when $\boldsymbol{\lambda_j \geq 1}$

$$\boldsymbol{h_j(t) = \alpha_j\lambda_j\, t^{\lambda_j-1}\, e^{t^{\lambda_j}}}$$

The survival function, $\boldsymbol{S_j(t)}$, is

$$\boldsymbol{S_j(t) = e^{\alpha_j(1-e^{t^{\lambda_j}})}}$$

The cumulative hazard distribution function, $\boldsymbol{H_j(t)}$, is

$$\boldsymbol{H_j(t) = 1 - \; e^{\alpha_j(1-e^{t^{\lambda_j}})}}$$

2.4 Competing risks setting

In next two subsections, the assumptions for both the competing risks and the competing risks regression models will be discussed, which is helpful for presenting the competing risks data and analysing a variety of real lifetime data sets.

2.4.1 Cause specific hazard rate

The competing risks problem can occur in various areas, particularly in electrical engineering, biomedical and biological studies, where the subject (or system) could be attacked by or break down due to more than one risk which may be present at the same time (Sarhan, 2007). For example, in biomedical research, competing risks models are very commonly used, particularly in cancer studies. Each patient will be under two risks, either the relapse or death in remission, which lead to failing treatment (Klein, 2006). Many researchers are interested in estimating a certain risk in the presence of other risk factors. Statistically, this procedure is known as the competing risks model. In this study, we analysed the competing risks model where multiple independent risks are assumed to compete for the failure of an individual, but each subject individually on life test can fail only by one risk. To analyze competing risks data, this study assumed the data contain three random variables: the time to event (failure or censored), an indicator $\delta = 1$ for failure and 0 for censoring and cause failure (in the failure case). In some cases, the cause of failure might be unknown and this case is called as the incomplete data (Sarhan, 2007). In this study we assume that the causes of failure is known.

To apply the parametric setups suppose, each (survival time or failure time) follows a distribution, such as the exponential, gamma, Weibull, generalized exponential and exponentiated Weibull. There are many studies about the parametric setups (see e.g. Berkson & Elveback, 1960 ; Cox, 1959; David & Moeschberger, 1978; Kundu & Basu, 2000; Park, 2005; Kundu & Sarhan, 2006; Sarhan, 2007; Sarhan et al., 2010).

In this study, we study the competing risks model in consider an incomplete and censored observations when each risk follows the two-parameter bathtub distribution. The two-parameter bathtub distribution was selected because the hazard rate can have a bathtub shape when the shape parameter $\lambda < 1$, and an increasing shape when the shape parameter λ ≥1, which is more helpful and efficient to use for analyzing the lifetime data sets, (Chen, 2000).

Generally, suppose n identical and independent subjects are put on life test. Let T_i, a random variable represent the lifetime of subject i, i=1,2,…,n, and C_i ,a random variable, represent the causes of failure. The causes of failure can be either independent or dependent of the lifetime, but this study, assumes that the object will be destroyed by only one cause and all risks are independent. Each subject is at the risk of failure due to different possible causes, but the occurrence of one cause of failure precludes all other causes of failure from occurrence. Additionally, each subject who enters to experiment but does not receive any causes of failure by the end of study or become lost to follow up before the end of the study period is censored. In this situation, let indicator δ_i equal a value of one if any one of the causes of failure is observed but equal a value of zero for a censored time. Assume a random variable, T_{ij}, which represents the time of the failure due to cause j, j=1,...,k. Only observe the minimum life time, T_i = min $\{T_{ij}\}$ where j=1,2,…,k. When the subject fails in competing risks system, there are three observable quantities (T_i, C_i, δ_i), T_i is the lifetime of subject i and C_i is the cause of failure, $C_i \in \{1,2,\ldots,k\}$. So, when the cause of failure is observed, $\delta_i = 1$, otherwise we will use only one observable quantity, $\delta_i = 0$, which represents censored time. Furthermore, assume that T_{ij} the time of the failure due to cause j, j=1,...,k., follow specific distribution and the probability density function $f_j(t)$, therefore, the pdf of $T_i$can be written in terms of f_j and S_j as:

$$\boldsymbol{f(t) = \sum_{j=1}^{k} f_j(t) \prod_{\substack{\ell=1 \\ \ell \neq j}}^{k} S_\ell(t)}$$

Using $\boldsymbol{f_j(t) = h_j(t) S_j(t)}$, we get

$$\boldsymbol{= \sum_{j=1}^{k} h_j(t) \prod_{\ell=1}^{k} S_\ell(t),}$$

and the hazard rate function of T_i h$(\boldsymbol{t})$, is

$$\boldsymbol{h}(\boldsymbol{t}) = \sum_{j=1}^{k} \boldsymbol{h}_j(\boldsymbol{t})$$

and the survival rate function of T_i S$(\boldsymbol{t})$, is

$$\boldsymbol{S}(\boldsymbol{t}) = \prod_{j=1}^{k} \boldsymbol{S}_j(\boldsymbol{t})$$

More specifically, assume that T_{ij} the time of the failure of the subject i due to cause j, j=1,...,k., follows the two-parameter bathtub distribution with unknown parameters α_j and $\lambda_j, T_{ij} \sim TPBT(\alpha_j, \lambda_j)$, for $i = 1, \dots, n$ and $j = 1, \dots, k$.

2.4.2 Cause specific hazard regression

In this study, a regression model will be used to analyse competing risks data, and the definition of the cause-specific hazard function with covariates is equal to,

$$\begin{aligned} \boldsymbol{h}_j\left(\boldsymbol{t} \middle| \boldsymbol{\beta}^{(j)}, \boldsymbol{Z}\right) &= \boldsymbol{h}_{j0}(\boldsymbol{t})\boldsymbol{exp}(\boldsymbol{\beta}_{1\mathrm{j}}\boldsymbol{Z}_1 + \boldsymbol{\beta}_{2\mathrm{j}}\boldsymbol{Z}_2 + \dots + \boldsymbol{\beta}_{pj}\boldsymbol{Z}_p) \\ &= \boldsymbol{h}_{j0}(\boldsymbol{t})\boldsymbol{exp}(\boldsymbol{\beta}^{(j)}\boldsymbol{Z}^{\boldsymbol{T}}) \end{aligned}$$

Where $\boldsymbol{h}_j\left(\boldsymbol{t} \,\middle|\boldsymbol{\beta}^{(j)}, \boldsymbol{Z}\right)$, is the hazard function at time t, $\boldsymbol{h}_{j0}(\boldsymbol{t})$ is an unspecified baseline hazard function that can be equal to any distribution, which gives the shape for the hazard function (Lim et al., 2010), $\boldsymbol{\beta}^{(j)} = (\boldsymbol{\beta}_{1\mathrm{j}}, \boldsymbol{\beta}_{2\mathrm{j}}, \dots, \boldsymbol{\beta}_{pj})$ is a vector of regression coefficients, and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \dots, \boldsymbol{Z}_p)^T$, is the vector of all covariates. The cumulative hazard function and the survival function are:

$$H_j(t|\beta^{(j)}, Z) = H_{j0}(t)\, exp(\beta_{1j}Z_1 + \beta_{2j}Z_2 + \cdots + \beta_{pj}Z_p)$$

$$= H_{j0}(t)\ \exp(\beta^{(j)}Z^T)$$

$$S_j(t|\beta^{(j)}, Z) = [S_{j0}(t)]^{exp(\beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p)}$$

$$= [S_{j0}(t)]^{exp(\beta^{(j)}Z^T)}$$

Using the above formulae, we can get the pdf of T_{ij}, using the following relation:

$$f_j(t|\beta^{(j)}, Z) = h_j(t|\beta^{(j)}, Z)\, S_j(t|\beta^{(j)}, Z)$$

In this study, the purpose of using the competing risks regression model is to evaluate the relationship of covariates to cause-specific failures (Dignam et al.,2012). In competing risks regression, we studied the relationship between a vector of covariates Z and specific causes of failure; As an example, every woman who is wearing an intrauterine device (IUD) can be under many high risks of expelling an IUD or accidental pregnancy (Kalbfleisch & Prentice, 2002).

Generally, we suppose n identical and independent subjects are put on life test. Let T_i, a random variable, represent the lifetime of subject i, i=1,2,…,n, and C_i , a random variable, represent the causes of failure.The causes of failure can be either independent or dependent, but this study, assumes that the object will be destroyed by only one cause and all risks are independent. Each subject is at the risk of failure due to different possible causes, but the occurrence of one cause of failure precludes all other causes of failure from occurrence. Additionally, each subject who enters to experiment does not receive any causes of failure by the end of study or become lost to follow up before the end of the study will be have censored. In this situation, let indicator δ_i equal a value of one if any one of the causes of failure is observed, but equal a value of zero for a censored time. assume a random variable T_{ij}, which represent the time of the failure due to cause j, j=1,...,k. Only observe the minimum of life time, T_i = min $\{T_{ij}\}$ where j=1,2,…,k. When the subject fails in competing risks system, there are three observable quantities (T_i, C_i, δ_i). T_i is the lifetime of subject and C_i is the cause of failure, $C_i \in \{1,2,\dots,k\}$. So, when the cause of failure is observed, $\delta_i = 1$, otherwise we will use only one observable quantity, δ_i= 0, which present censored time. In case of regression, let Z_i be a vector representing covariates or explanatory variables such as age, gender and group of treatment.

More specifically, assume that T_{ij} the time of the failure due to cause j, j=1,...,k., follows the two-parameter bathtub distribution with unknown parameters α_j and $\lambda_j, T_{ij} \sim TPBT(\alpha_j, \lambda_j)$, for $i = 1, \dots, n$ and $j = 1, \dots, k$.

The probability density function of the TPBT distribution (pdf) is

$$\boldsymbol{f_j(t|\beta^{(j)}, Z) = \alpha_j\lambda_j\ t^{\lambda_j-1}\ e^{t^{\lambda_j}}\ exp(\beta^{(j)}Z^T)\ \left\{e^{\alpha_j(1-e^{t^{\lambda_j}})}\right\}^{exp(\beta^{(j)}Z^T)}}$$

the hazard rate function, h(t),

$$\boldsymbol{h_j(t|\beta^{(j)}, Z) = \alpha_j\lambda_j\ t^{\lambda_j-1}\ e^{t^{\lambda_j}} exp(\beta^{(j)}Z^T)}$$

and the survival rate function $\boldsymbol{S_j(t|\beta, Z)}$

$$\boldsymbol{S_j(t|\beta^{(j)}, Z) = \left\{e^{\alpha_j(1-e^{t^{\lambda_j}})}\right\}^{exp(\beta^{(j)}Z^T)}}$$

and the cumulative distribution function (CDF) is

$$\boldsymbol{H_j(t|\beta^{(j)}, Z) = \left\{1 - e^{\alpha_j(1-e^{t^{\lambda_j}})}\right\}\ e^{\beta^{(j)}Z^T}}$$

Here, θ becomes a vector of $\alpha_j, \lambda_j, \beta^{(j)} = (\beta_{1j}, \beta_{2j}, \dots, \beta_{Pj})$, j=1,2,…,K. That is, θ is a vector of (2+P) K parameters.

2.5 The relative risk rates

The most important characteristics of the competing risks models are the relative risks; a relative risk in competing risks system is known as one risk of many competing risks. In this section, we study the failure probability distribution of each cause of failure in presence of all other risks which explains each risk that is studies due to a specific cause of failure (Bocchetti, Giorgio, Guida, & Pulcini, 2009; Sarhan, Hamilton & Smith, 2010). According to Bocchetti et al. (2009) the failure probability of cause j at time t in the presence of all other risks is defined as following

$$F_j(t) = \int_0^t h_j(y) \prod_{\ell=1}^{k} S_\ell(y)\ dy, \quad j = 1,2,\dots,K.$$

The risk due to cause j, $j = 1,2,\dots,k$, is

$$\pi_j = \lim_{t\to\infty} F_j(t) = \int_0^\infty h_j(t) \prod_{\ell=1}^{k} S_\ell(t)\ dt$$

More specifically, in case of the TPBT competing risks model that is discussed in Chapter 2, the relative risk rates, $\boldsymbol{\pi_j}$, can be derived by solving the following integral. In this study, there are only two causes of failure, which are assumed to be independent. The relative risk due to cause $j, j = 1,2,\dots,$K, is

$$\boldsymbol{\pi_j} = \int_{\boldsymbol{0}}^{\infty} \boldsymbol{\alpha_j \lambda_j} t_i^{\boldsymbol{\lambda_j - 1}}\ \boldsymbol{e}^{\boldsymbol{t_i}^{\lambda_\mathrm{j}}}\ \boldsymbol{e}^{\sum_\ell^k \alpha_\ell (\mathbf{1} - \boldsymbol{e}^{\boldsymbol{t_i}^{\lambda_\ell}})}\ dt$$

There is no analytic solution for this integral. To calculate the risks, numerical integration methods will be applied.

Special case: if all the shape parameters of the causes are equal, say $\lambda_j = \lambda, j = 1,2,\ldots,$K. The risk due to cause j can be obtained in a closed form as:

$$\boldsymbol{\pi_j} = \frac{\boldsymbol{\alpha_j}}{\sum_{\boldsymbol{\ell=1}}^{\mathbf{K}} \boldsymbol{\alpha_\ell}} \quad , \quad j = 1,2,..,\mathrm{K}.$$

In another case of this study, the TPBT competing risks regression model that is discussed in Chapter 2, the relative risk rates, $\boldsymbol{\pi_j}$, can be derived by solving the following integral. In this study, there are only two causes of failure, which are assumed to be independent. The relative risk due to cause $\boldsymbol{j}, \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \ldots,$ K, is

$$\boldsymbol{\pi_j} \;=\; \int_{\mathbf{0}}^{\infty} \boldsymbol{\alpha_j \lambda_j} t_i^{\boldsymbol{\lambda_j - 1}} \; \boldsymbol{e}^{\boldsymbol{t_i}^{\lambda_{\mathrm{j}}}} \boldsymbol{e}^{(\beta^{(j)} \mathrm{Z}^T)} \; [\boldsymbol{e}^{\sum_{\boldsymbol{\ell}}^{\boldsymbol{k}} \boldsymbol{\alpha_\ell}(\mathbf{1} - \boldsymbol{e}^{\boldsymbol{t_i}^{\boldsymbol{\lambda_\ell}}})}\,]^{\boldsymbol{e}^{(\beta^{(\ell)} \mathrm{Z}^T)}} \; dt$$

There is no analytic solution for this integral. To calculate the risks, numerical integration methods will be applied.

Special case: if all the shape parameters of the causes are equal, say $\lambda_j = \lambda, j = 1,2,\ldots,$K. The risk due to cause j can be obtained in a closed form as:

$$\boldsymbol{\pi_j} = \frac{\boldsymbol{\alpha_j e^{(\beta^{(j)} Z^T)}}}{\sum_{\boldsymbol{\ell=1}}^{\mathbf{K}} \boldsymbol{\alpha_\ell} \, \boldsymbol{e^{(\beta^{(\ell)} Z^T)}}} \quad , \quad j = 1,2,\ldots,\mathrm{K}.$$

2.6 Maximum Likelihood Estimation

According to Aldrich (1997), the first one who presented the maximum likelihood was Fisher between 1912 and 1922, but he did not finish. A few years later more detailed development by many researchers depended on his ideas. Many researchers applied the maximum likelihood estimation, which is the most frequently used method in statistical inference to estimate unknown parameters. When any experiment depends on a large sample of participants, various researchers will use the maximum likelihood estimation because that gives reasonable estimator of $\boldsymbol{\theta}$.

Assuming a random sample $\boldsymbol{x_1, \dots, x_n}$ from a distribution $\boldsymbol{f(x_i, \theta)}$, the likelihood function is denoted as *L(data|$\boldsymbol{\theta}$)*, $\boldsymbol{\theta}$, which is a present vector of unknown parameters. In general, we can write the likelihood function as following:

$$\boldsymbol{L(data|\theta) = \prod_{i=1}^{n} f(x_i, \theta)}$$

The maximum likelihood estimator for θ is the values of θ that maximize the likelihood function or logarithm likelihood function (Enders & Bandalos, 2001). The following steps show how we can get the MLE in general. Firstly, take the log-likelihood function, which denotes $\boldsymbol{\ell(\theta)}$,

$$\boldsymbol{\ell(\theta) = \sum_{i=1}^{n} log\, f(x_i, \theta)}$$

Secondly, take the first and the second derivative of $\boldsymbol{\ell(\theta)}$, depending on number of the parameters.

In case of high-dimensional parameters $\boldsymbol{\theta = (\theta_1, \theta_2, \dots, \theta_n)}$, here $\boldsymbol{\theta}$ is vector include all unknown parameters. Hence, the Fisher information $\boldsymbol{I}$ $(\boldsymbol{\theta})$ will be have a matrix. The $\boldsymbol{ij}$-th is two different entries that are $\boldsymbol{i}$-th present number of rows and $\boldsymbol{j}$-th present number of columns.

$$\boldsymbol{I(\theta)_{ij} = E_\theta \left[\frac{\partial}{\partial \theta_i} \ell(\theta) \frac{\partial}{\partial \theta_j} \ell(\theta)\right]}$$

$$\boldsymbol{= -E_\theta \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\theta)\right]}$$

After we obtained the Fisher information matrix $\boldsymbol{I(\theta)_{ij}}$, the diagonal of the Fisher information matrix $\boldsymbol{I(\theta)_{ij}}$ will provide estimates variances but above and below the diagonal will be covariance.

$$\boldsymbol{V_{\theta}\left(\widehat{\theta}_i(X)\right) \approx I(\theta)_{ii}^{-1}} \qquad \boldsymbol{Cov_{\theta}\left(\widehat{\theta}_i(X), \widehat{\theta}_j(X)\right) \approx I(\theta)_{ij}^{-1}}$$

In this case, we assumed independent sampling from the probability density function $\boldsymbol{f(x_i, \theta)}$, and $\boldsymbol{\widehat{\theta}}$ may be approximated by $\boldsymbol{\widehat{\theta} \sim N\left(\theta\ , I(\widehat{\theta}\ )^{-1}\right)}$, the confidence interval for $\boldsymbol{\widehat{\theta}_i}$ is

$$\boldsymbol{\widehat{\theta}_i \pm Z_{\frac{\alpha}{2}} \sqrt{I(\theta)_{ii}^{-1}}}$$

In the next subsections, the likelihood function for both cases will be discussed to present competing risks data.

2.6.1The likelihood function for competing risks model

Some notation will be introduced to write general the likelihood function in competing risks model. Firstly, assume n independent subject on life test and for subject i=1,...,n. Hence, observe three random variables $(\ T_i, C_i, \delta_i\ )$. T_i represents the lifetime to failure assumed to be independent identically distributed over items i=1,…,n. C_i represents the causes of failure which are assumed to be independent, but in this study only one cause of failure can occur. We supposed if δ_i=1 the causes of failure occur, but otherwise δ_i= 0 which is censored.

The observed data will be $(T_1, C_1, \delta_1), (T_2, C_2, \delta_2), \dots, (T_n, C_n, \delta_n)$ dependent on assumptions in subsection 2.4.1. In general, the likelihood function and the log-likelihood function for competing risks model can be written as following:

$$\boldsymbol{L(data|\theta)} = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \{[\boldsymbol{f(t_i)}]^{\boldsymbol{I(\delta_i=1)}}\ \ [\boldsymbol{S(t_i)}]^{\boldsymbol{I(\delta_i=0)}}\}$$

$$= \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \{[\boldsymbol{h(t_i)}]^{\boldsymbol{I(\delta_i=1)}}\ [\boldsymbol{S(t_i)}]^{\boldsymbol{I(\delta_i=1)}}\ [\boldsymbol{S(t_i)}]^{\boldsymbol{I(\delta_i=0)}}\ \}$$

$$= \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \{[\boldsymbol{h(t_i)}]^{\boldsymbol{I(\delta_i=1)}}\ [\boldsymbol{S(t_i)}]^{\boldsymbol{I(\delta_i=1)+I(\delta_i=0)}}\ \}$$

$$= \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \prod_{j=1}^{K} \{[\boldsymbol{h_j(t_i)}]^{\boldsymbol{I(C_i=\mathrm{j})}}\ \boldsymbol{S_j(t_i)}\ \}$$

Here $\boldsymbol{\theta}$ is the vector of 2K including all unknown parameters in this model, $\boldsymbol{\theta} = (\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \dots, \boldsymbol{\alpha_K}, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \dots, \boldsymbol{\lambda_K})$. The log-likelihood function is

$$\boldsymbol{\mathcal{L}(data|\theta)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \sum_{j=1}^{K} \left[\boldsymbol{I(c_i=j)}\ \boldsymbol{log}\left(\boldsymbol{h_j(t_i)}\right) + \boldsymbol{log}\left(\boldsymbol{S_j(t_i)}\right)\right]$$

Based on the model assumptions in subsection 2.4.1, the likelihood function and the log-likelihood function for the competing risks model can be written in specific form when each cause of failure follows the two-parameter bathtub distribution with unknown parameters.

$$L(data|\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{K} \left\{ e^{\alpha_j\left(1-e^{t_i^{\lambda_j}}\right)} \left[\alpha_j\lambda_j\, t_i^{\lambda_j-1}\, e^{t_i^{\lambda_j}}\right]^{I(c_i=j)} \right\}$$

$$\mathcal{L}(data|\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{k} \left\{\left[\alpha_j\left(1-e^{t_i^{\lambda_j}}\right)\right] + I(c_i=j)\ [log(\alpha_j) + log(\lambda_j) + (\lambda_j - 1)log(t_i) + t_i^{\lambda_j}]\right\}$$

2.6.2 The likelihood function for competing risks regression model

Some notation will be introduced to write general the likelihood function in competing risks regression model. Firstly, assume n independent subject on life test and for individual i=1,...,n. Let T_i as a vector represent the lifetime of units which are assumed to be independent identically distributed over items i=1,…,n. Only observe the minimum of life time, $\boldsymbol{T_i}$ = min $\{\boldsymbol{T_{ij}}\}$ where j=1,2,…,k, and $\boldsymbol{T_{ij}}$ is a vector representing the time when the failure due to cause j, j =1,...,k, and $\boldsymbol{C_i}$ is a random variable for the event type based on the assumption the relationship between the causes are independent ,but not identically distributed over causes j and known. Also, $\boldsymbol{\delta_i}$ is an indicator representing the events type if the causes of failure are observed then $\boldsymbol{\delta_i = 1}$, but otherwise $\boldsymbol{\delta_i = 0}$ which is right censored. Let $\boldsymbol{Z_i}$ be a vector representing covariates or explanatory variables.

The observed data will be $(\boldsymbol{T_1}, \boldsymbol{C_1}, \boldsymbol{\delta_1}, \boldsymbol{Z_1}), (\boldsymbol{T_2}, \boldsymbol{C_2}, \boldsymbol{\delta_2}, \boldsymbol{Z_2}), \dots, (\boldsymbol{T_n}, \boldsymbol{C_n}, \boldsymbol{\delta_n}, \boldsymbol{Z_n})$ dependent on this assumption. The likelihood function and the log-likelihood function in general cases for regression model to present competing risks data can be written as following:

$$\boldsymbol{L(data|\theta)} = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \{\, [\, \boldsymbol{f(t_i|Z_i)}]^{\boldsymbol{I(\delta_i=1)}} \quad [\boldsymbol{S}(\boldsymbol{t_i}|\boldsymbol{Z_i})]^{\boldsymbol{I(\delta_i=0)}}\}$$

$$= \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \{[\boldsymbol{h(t_i|Z_i)}]^{\boldsymbol{I(\delta_i=1)}} \, [\boldsymbol{S(t_i|Z_i)}]^{\boldsymbol{I(\delta_i=1)}} \quad [\boldsymbol{S(t_i|Z_i)}]^{\boldsymbol{I(\delta_i=0)}}\}$$

$$= \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \{[\boldsymbol{h(t_i|Z_i)}]^{\boldsymbol{I(\delta_i=1)}} \, [\boldsymbol{S(t_i|Z_i)}]^{\boldsymbol{I(\delta_i=1)+I(\delta_i=0)}} \quad \}$$

$$= \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left\{ \boldsymbol{S(t_i|Z_i)} \prod_{\boldsymbol{j=1}}^{\boldsymbol{k}} [\boldsymbol{h_j(t_i|Z_i)}]^{\boldsymbol{I(C_i=j)}} \quad \right\}$$

$$= \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \prod_{j=1}^{K} \{\boldsymbol{S_j(t_i|Z_i)} \, [\boldsymbol{h_j(t_i|Z_i)}]^{\boldsymbol{I(C_i=j)}} \quad \}$$

Here $\boldsymbol{\theta}$ is the vector of all unknown parameters in the model that includes the cause of failure distribution parameters and the coefficients of the covariates.

$$\boldsymbol{\mathcal{L}(data|\theta)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left[\boldsymbol{log\big(S(t_i|Z_i)\big)} + \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \boldsymbol{I(C_i = j)\ log\left(h_j(t_i|Z_i)\right)} \right]$$

$$= \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \sum_{j=1}^{K} \left[\boldsymbol{log\left(S_j(t_i|Z_i)\right)} + \boldsymbol{I(C_i = j)\ log\left(h_j(t_i|Z_i)\right)} \right]$$

Based on the model assumptions in subsection 2.4.2, the likelihood function and the log-likelihood function for the competing risks regression model can be written as

$$\boldsymbol{L(data|\theta)} = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \prod_{j=1}^{K} \left\{ \left\{ e^{\alpha_j(1-e^{t^{\lambda_j}})} \right\}^{exp(\beta^{(j)} Z^T)} \alpha_j \lambda_j\ t_i^{\lambda_j - 1}\ e^{t_i^{\lambda_j}} exp(\boldsymbol{\beta}^{(j)} \boldsymbol{Z^T}) \right\}$$

$$\boldsymbol{\mathcal{L}(data|\theta)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \left\{ e^{\beta^{(j)} Z^T}\ log(S_j(t_i)) + I(C_i = j)\ log\left(h_j(t_i|Z_i)\right) \right\}$$

$$= \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \left\{ e^{\boldsymbol{\beta}^{(j)} \boldsymbol{Z^T}} \left[\alpha_j \left(1 - e^{t_i^{\lambda_j}} \right) \right] + I(C_i = j)\ [\boldsymbol{\beta}^{(j)} \boldsymbol{Z^T} + log(\alpha_j) + \log(\lambda_j) + (\lambda_j - 1)\log(t_i) + t_i^{\lambda_j}] \right\}$$

Chapter 3: Bayesian Method

3.1 Introduction

In this chapter, we start by describing the main components of the Bayesian framework. Firstly, to make a Bayesian inference, suppose we have an independent random sample $\mathrm{Y}_1, \mathrm{Y}_2, \dots, \mathrm{Y}_\mathrm{n}$ from a probability distribution. Fundamentally, all of the Bayesian analyses involve assuming a prior distribution for each unknown parameter. There are two main approaches to choosing a prior distribution; the first approach is an informative prior distribution and the second approach is a non-informative prior distribution (e.g. Glickman & van Dyk, 2007). The concept of choosing prior distribution is subjective and unscientific, because we do not have enough information on unknown parameters. Typically, the primary goal of using the Bayesian statistical analysis is to obtain the posterior distribution of model parameters to carry out all the inferences. According to Bayes' theorem, the posterior probability density function of θ, given y, is given by

$$\mathrm{g}(\theta|\mathrm{y}) = \frac{g(\theta)L(y|\theta)}{p(y)}$$

where

$$p(y) = \begin{cases} \int g(\theta)L(y|\theta)d\theta, & \text{If } \theta \; is\; continuous \\ \sum g(\theta)L(y|\theta), & \text{If } \theta \; is\; discrete \end{cases}$$

and

- $g(\theta)$: The prior distribution of θ.
- $L(y|\theta)$: The likelihood function.
- $p(y)$: The marginal distribution of y.

There are different pros and cons of using the Bayesian methods in statistical inference to do all data analysis, which are explained by various authors (such as Berger, 1985; Berger & Wolpert, 1988; Bernardo & Smith, 1994; Carlin & Louis, 2000; Robert, 2001; and Wasserman, 2004; Clard and Gelfand, 2006). To use the Bayesian methods, we assume a prior distribution for all unknown parameters, but there are no methods that can be followed to select the right prior distribution, which is one of the cons of using the Bayesian analysis. However, a posterior distribution sometimes can have influence based on the selection of prior distribution.

One of the pros is that a posterior distribution can combine information for both a prior distribution and a likelihood function. Also, using Bayesian inference has the ability to consider prior opinion or external experiential evidence into the results via the prior distribution. It is not easy to obtain the integral of the posterior distribution for high-dimensional parameters, especially when the posterior distribution does not have closed form. Also, the lack of computational tools, which made scientists in various fields, reluctant to use the Bayesian approaches. However, in recent years many programs are available which researcher can use to solve this problem. For example, in R's specific packages are helpful to use for applying different methods in Bayesian analysis.

In Bayesian analysis, there are many techniques that can be applied to estimate unknown parameters, such as normal approximations, rejection sampling, importance sampling, sampling importance resampling, and Markov Chain Monte Carlo methods. Furthermore, those methods sometimes are difficult to apply for various reasons. Each technique has specific conditions that must be achieved such as the choice of the proposal density or acceptance rate. However, in this thesis only the Markov Chain Monte Carlo (MCMC) methods will be used to obtain the posterior distribution without normalizing the constant, which will be discussed in section 3.5. For researchers who are interested to apply different techniques of Bayesian analysis in various areas, many authors can be followed who have done a lot of work on Bayesian analysis (such as Albert ,2009; Berger, 2013; Bernardo and Smith, 2001; Box and Tiao, 2011; Gelman et al., 2014; O'Hagan and Forster, 2004; Press, 2009).

3.1.1 Basic elements

The essential elements in the Bayesian decision framework will be discussed in general. A loss function $\mathbb{L}(\theta, \widehat{\boldsymbol{\theta}})$ defined on $\ominus\times \mathcal{H}$ which represents the loss incurred when the decision $\widehat{\boldsymbol{\theta}}$ is taken, and the parameter is θ. The general form of the loss function takes the form

$$\mathbb{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = \boldsymbol{\gamma}(\boldsymbol{\theta})\mathbf{W}(|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}|), \tag{3.1}$$

Where W is a non-negative function of the error $|\widehat{\boldsymbol{\theta}} - \theta|$ such that W(0)=0 and γ is a positive. It is frequently assumed that the function γ in (3.1) is a constant. Accordingly, the loss function may be written as

$$\mathrm{L}(\theta, \widehat{\boldsymbol{\theta}}) = \mathrm{a}\,\mathrm{W}(|\widehat{\boldsymbol{\theta}} - \theta|),\ \ \mathrm{a} > 0. \tag{3.2}$$

Without loss of generality, we shall assume that a=1.

3.1.2 The risk function, prior and posterior risks

Let $\widehat{\boldsymbol{\theta}} \in \mathcal{H}$ be an estimator for θ. Let $\mathbb{L}$ be the loss function. The risks function of $\widehat{\boldsymbol{\theta}}$, denoted by R $(\theta, \widehat{\boldsymbol{\theta}})$, is defined as the expectation of the loss function L. That is,

$$\mathrm{R}\left(\theta, \widehat{\boldsymbol{\theta}}\right) = \mathrm{E}_{\theta}\left[\mathbb{L}\left(\theta, \widehat{\boldsymbol{\theta}}(\mathrm{y})\right)\right] = \int_{\mathrm{y}} \mathbb{L}\left(\theta, \widehat{\boldsymbol{\theta}}(\mathrm{y})\right) L(\mathrm{y}|\theta)\,\mathrm{dy}, \forall\theta \in \ominus \tag{3.3}$$

Here, E_θ means the expectation corresponding to L. For given estimator $\widehat{\boldsymbol{\theta}}$, the risk function $R\left(\theta, \widehat{\boldsymbol{\theta}}\right)$ is considered to be a function on $\ominus$.

The prior risk (or Bayes risk) of the estimator $\widehat{\boldsymbol{\theta}} \in \mathcal{H}$ with respect to the prior distribution G having density g, denoted by $R\left(g, \widehat{\boldsymbol{\theta}}\right)$, is defined as the prior expectation of the risk function $R\,(\theta, \widehat{\boldsymbol{\theta}})$. That is,

$$\mathbf{R}\left(\mathbf{g},\ \widehat{\boldsymbol{\theta}}\right) = \int_{\ominus} \mathbf{R}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}\right) \mathbf{g}(\boldsymbol{\theta})\, \mathbf{d}\boldsymbol{\theta} \qquad (\mathbf{3}.4)$$

Because of this loss function is non-negative (Bartoszewicz, 1989,p.183); therefore, we can write the prior risk $R\left(g, \widehat{\boldsymbol{\theta}}\ \right)$, using (3.4,3.5), to be

$$R\left(g, \widehat{\boldsymbol{\theta}}\right) = \int_{\ominus} \left\{ \int_{y} \mathbb{L}\left(\theta, \widehat{\boldsymbol{\theta}}\right) L(y|\theta)\, dy \right\} g(\theta)\ d\theta \qquad (3.5)$$

The joint probability density of the random vector (Y, θ) is

$$g(x, \theta) = L(y|\theta) g(\theta) = g(\theta|y)\, p(y) \qquad (3.6)$$

Using (3.6 and 3.7) the prior risk may be written as the non-conditional expectation of the loss function with respect to the joint probability density $g(\theta, y)$, denoted by E $[L\left(\theta, \widehat{\boldsymbol{\theta}}\right)]$. Namely,

$$R\left(g, \widehat{\boldsymbol{\theta}}\right) = E\left[\mathbb{L}\left(\theta, \widehat{\boldsymbol{\theta}}\right)\right] = \int_{y \times \ominus} \mathbb{L}\left(\theta, \widehat{\boldsymbol{\theta}}\right) g(y; \theta)\, dy\, d\theta \qquad (3.7)$$

Further, the prior risk can be written in the following form.

$$\mathbf{R}\left(\mathbf{g}, \widehat{\boldsymbol{\theta}}\right) = \int_{y} \{ \int_{\ominus} \mathbb{L}(\boldsymbol{\theta}, \mathbf{d})\, \mathbf{g}(\boldsymbol{\theta}|\mathbf{y})\, \mathbf{d}\boldsymbol{\theta} \}\, \mathbf{p}(\mathbf{y})\, \mathbf{dy} \qquad (3.8)$$

The posterior risk of the estimator $\widehat{\boldsymbol{\theta}} \in \boldsymbol{\mathcal{H}}$, given Y= y, with respect to being the prior density g, denoted by $\boldsymbol{\rho_g}\left(\widehat{\boldsymbol{\theta}}\right)$, is defined as the posterior expectation of the loss function $\mathbb{L}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}\right)$. That is,

$$\rho_g\left(\widehat{\boldsymbol{\theta}}\right) = E\left[\mathbb{L}\left(\theta, \widehat{\boldsymbol{\theta}}\right)|y\right] = \int_{\ominus} \mathbb{L}\left(\theta, \widehat{\boldsymbol{\theta}}\right) g(\theta|y)\, d\theta \qquad (3.9)$$

It seems from (3.9 and 3.10) that, the prior risk $\boldsymbol{R}\ (\boldsymbol{g}, \widehat{\boldsymbol{\theta}})$ is the expectation of the posterior risk $\boldsymbol{\rho_g}(\widehat{\boldsymbol{\theta}})$, with respect to the marginal distribution of Y under G.

The Bayes estimator for $\boldsymbol{\theta}$ with respect to the prior distribution G is defined as that estimator in $\mathcal{H}$, which minimizes the posterior risk, given Y. Let $\widehat{\boldsymbol{\theta}}_{\boldsymbol{B}}(\boldsymbol{Y})$ be the Bayes estimator for $\boldsymbol{\theta}$, then

$$\boldsymbol{\rho_g}\big(\widehat{\boldsymbol{\theta}}_{\boldsymbol{B}}(\boldsymbol{Y})\big) = \underset{\widehat{\boldsymbol{\theta}} \in \boldsymbol{\mathcal{H}}}{\boldsymbol{inf}}\{\textstyle\int_{\ominus} \mathbb{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})\ \boldsymbol{g}(\boldsymbol{\theta}|\boldsymbol{y})\ \boldsymbol{d\theta}\} \qquad (3.10)$$

Note that, the Bayes estimator of $\boldsymbol{\theta}$ for a given prior distribution G is not necessarily unique. If the loos function $\mathbb{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})$ is strictly convex in d for each $\boldsymbol{\theta}$, then $\widehat{\boldsymbol{\theta}}_{\boldsymbol{B}}(\boldsymbol{Y})$ is virtually unique (see, for example, Girshick & Savage, 1951; DeGroot, 1970 and Box & Tiao, 1973). We also note that the Bayes estimator $\widehat{\boldsymbol{\theta}}_{\boldsymbol{B}}$ minimizes the prior risk. In some textbooks and articles, "for example, Girshick and Savage, (1951) and DeGroot & Rao, (1963)" the Bayes estimator is defined as that estimator $\widehat{\boldsymbol{\theta}} \in \boldsymbol{\mathcal{H}}$ which minimizes the prior risk. Recall that, the Bayes estimate $\widehat{\boldsymbol{\theta}}_{\boldsymbol{B}}$ is the value of a Bayes estimator then $\widehat{\boldsymbol{\theta}}_{\boldsymbol{B}}(\boldsymbol{Y})$.

3.1.3 The Squared error loss function

When the parameter $\boldsymbol{\theta}$ is one- dimensional, the loss function can be expressed as

$$\mathbb{L}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = \mathbf{a}\,|\,\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}|^{\mathbf{b}}, \qquad (3.11)$$

Where a > 0 (that can be chosen to be a=1), and b > 0. When b $= 2$, the loss function is quadratic and is called a squared error loss function; when b=1, the loss function is proportional to the absolute value of the estimation error and is called an absolute error loss function.

The squared error loss function lends itself to mathematical. It also yields a second good approximation to the more general loss function $\boldsymbol{W}(|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}|)$, where W(.) can be differentiated at least twice. For these reasons, it is the most commonly applies in statistical estimation.

3.1.4 The Quadratic loss function

Suppose that we are interested in estimating the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \dots, \boldsymbol{\theta_k})$, where $\boldsymbol{K} \geq \mathbf{2}.$ A generalization of the squared error loss function (3.12) is the quadratic loss function $\mathbb{L}$ given by

$$\mathbb{L}(\theta, \widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{K}(\theta_i - \widehat{\theta_\imath})^2 \qquad (3.12)$$

3.2 Bayesian competing risks model

3.2.1 Prior distribution

A prior distribution of a parameter is defined as the probability distribution that appears unpredictably about the parameter before the current data is inspected. To do Bayesian inference, a prior distribution must be assumed for all unknown parameters. There are no specific methods that explain how we can choose the right prior, but some researchers decide to choose a prior depending on their knowledge about and experience of data which means their choice is highly subjective.

In this study, suppose all unknown parameters in competing risks model are independent and follow a gamma distribution. Let α_j follow a gamma distribution with a positive shape parameter $a_{j1}$and scale parameter $\mathrm{a_{j2}}$, and λ_j follow a gamma distribution with a positive shape parameter v_{j1} and scale parameter v_{j2}, for j = 1,2,…, K. Additionally, the total number of unknown parameters in this model is four, which is relevant to 2K and K is the present number of causes of failure which equals two in this study. Hence, we assume all hyper-parameters $(a_{j1}, a_{j2}, v_{j1}, v_{j2}), j = 1,2, \dots,$ K, are known.

$$\alpha_j \sim Gamma(a_{j1}, a_{j2}), \;\; j = 1,2, \dots, \mathrm{K}$$

$$\lambda_j \sim Gamma(v_{j1}, v_{j2}), \;\; j = 1,2, \dots, \mathrm{K}$$

Here, $\theta = \left(\alpha_j, \lambda_j\right), and$ is a vector including all unknown parameters and the joint prior density function of θ up to a constant as following:

$$\mathrm{g}_{CR}(\theta) \propto \prod_{i=1}^{4} g_i(\theta_i)$$

$$\mathrm{g}_{CR}(\theta) \propto \prod_{j=1}^{K} \alpha_j^{a_{j1}-1} e^{-a_{j2}\alpha_j} \ \lambda_j^{v_{j1}-1} e^{-v_{j2}\lambda_j} \qquad , \alpha_j, \lambda_j > 0$$

And log prior density function of θ can be written as following:

$$log\left(\mathrm{g}_{CR}(\theta)\right) = \sum_{j=1}^{K} \{\left(a_{j1} - 1\right) \log\left(\alpha_j\right) - a_{j2}\alpha_j + \left(v_{j1} - 1\right) \log\left(\lambda_j\right) - v_{j2}\lambda_j \ \}$$

3.2.2 Posterior distribution

In this subsection, the posterior density, $\mathrm{g}_{CR}(\theta|\text{data})$, is the product of the joint prior density function and the likelihood function. Hence, we combined the joint prior distributions in equation 3.2 and the likelihood function. The joint posterior density function of θ, up to a constant, can be written as following:

$$\mathrm{g}_{CR}(\theta|\text{data}) \propto \mathrm{g}_{CR}(\theta) \times L_{CR}(data|\theta)$$

$$\mathrm{g}_{CR}(\theta|\text{data}) \propto \prod_{j=1}^{K} \alpha_j^{a_{j1}-1} e^{-a_{j2}\alpha_j} \ \lambda_j^{v_{j1}-1} e^{-v_{j2}\lambda_j}$$

$$\times \prod_{i=1}^{n} \prod_{j=1}^{k} \{e^{\alpha_j\left(1-e^{t_i^{\lambda_j}}\right)} \quad \left[\alpha_j \lambda_j \, t_i^{\lambda_j - 1} \ e^{t_i^{\lambda_j}}\right]^{I(c_i=j)}\} \qquad (3.13)$$

The normalizing constant is

$$p(y) = \int_0^\infty \dots \int_0^\infty \prod_{j=1}^{K} \alpha_j^{a_{j1}-1} e^{-a_{j2}\alpha_j} \; \lambda_j^{v_{j1}-1} e^{-v_{j2}\lambda_j}$$

$$\times \prod_{i=1}^{n} \prod_{j=1}^{k} \; \{e^{\alpha_j\left(1-e^{t_i^{\lambda_j}}\right)} \quad [\alpha_j \lambda_j \, t_i^{\lambda_j - 1} \; e^{t_i^{\lambda_j}}]^{\,I(c_i=j)}\} \, d\alpha_j d\lambda_j \qquad (3.14)$$

Here, θ is a vector of all unknown parameters, $\theta = (\alpha_1, \alpha_2 \dots, \alpha_K, \lambda_1, \lambda_2, \dots, \lambda_K)$. Using the joint posterior distribution, we can derive (calculate) the Bayes estimate of each parameter. The Bayes estimate of θ_j, j=1,2,…,2K, is

$$\hat{\theta}_j = \int \dots \int \theta_j \; g(\theta|data) \; d\theta$$

and the Bayes estimate for any function of θ, say $v(\theta)$, is

$$\hat{v}(\theta) = \int \dots \int v(\theta) \;\; g(\theta|data) \; d\theta$$

The integrals above do not have analytic solutions, and numerical approaches are required, which will be discussed later in section 3.4 and section 3.5. To make all parameters real-valued, we can use the logarithmic transformation. That is, we use $\theta^* = \log(\theta_j)\,, j = 1,2, \dots ,2K.$ and the new transformed vector of unknown parameters, say θ^*, is; $\theta^* = (\log(\theta_1)\,, \log(\theta_2)\,, \log(\theta_3)\,, \log(\theta_4))$. We need to obtain the Jacobian term in the transformation as following:

$$J\left|\frac{\theta}{\theta^*}\right| = \begin{bmatrix} e^{\theta_1^*} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\theta_{2K}^*} \end{bmatrix} = e^{\sum_{j=1}^{2K} \theta_j^*}$$

$$\log J\left|\frac{\theta}{\theta^*}\right| = \sum_{j=1}^{2K}\theta_j^*$$

The joint posterior density function of the transformed real-valued vector of unknown parameters of θ^* is

$$\mathrm{g}_{\mathrm{CR}}(\theta^*|data) = \mathrm{g}_{\mathrm{CR}}\left(\theta = e^{\theta^*} \mid data\right) J\left|\frac{\theta}{\theta^*}\right|$$

Thus, the log-posterior density function of θ^*, given data, is

$$\log(\mathrm{g}_{\mathrm{CR}}(\theta^*|data)) = \log\left(\mathrm{g}_{\mathrm{CR}}\left(\theta = e^{\theta^*} \mid data\right)\right) + \sum_{j=1}^{2K}\theta_j^*$$

3.3 Bayesian competing risks regression model

3.3.1 Prior distributions

We assume that all parameters are independent, α_j follows gamma distribution with hyper-parameters a_{j1} and a_{j2}, and λ_j follows gamma distribution with hyper-parameters v_{j1} and v_{j2}, for j = 1,2,…,K. Furthermore, we assumed that all the regression coefficients $\beta_{\ell j}, \ell = 1,2,\ldots,\mathrm{P}; j = 1,2,\ldots,\mathrm{K}$ are independent and follow normal distribution with known mean and known variance. That is,

$$\alpha_j \sim Gamma(a_{j1}, a_{j2}),\ \ j = 1,2,\ldots,\mathrm{K}$$

$$\lambda_j \sim Gamma(v_{j1}, v_{j2}),\ \ j = 1,2,\ldots,\mathrm{K}$$

$$\beta_{\ell j} \sim Normal\left(\mu_{\ell j}, \sigma_{\ell j}^2\right), \ell = 1,2,\ldots,\mathrm{P}, j = 1,2,\ldots,\mathrm{K}$$

We use θ to represent the vector of all unknown parameters, where $\theta = (\alpha_1, \dots, \alpha_K, \lambda_1, \dots, \lambda_K, \beta^{(1)}, \dots, \beta^{(K)})$ and $\beta^{(j)} = (\beta_{1j}, \beta_{2j}, \dots, \beta_{Pj}), j = 1,2, \dots, K$. That is θ has (2+P) K components. The joint prior density function of the vector of all unknown parameters θ is

$$g(\theta) = \prod_{j=1}^{K} g_j(\theta_j)$$

$$g_{CRR}(\theta) \propto \prod_{j=1}^{K} \left\{ \alpha_j^{a_{j1}-1} e^{-a_{j2}\alpha_j} \times \lambda_j^{v_{j1}-1} e^{-v_{j2}\lambda_j} \times \quad e^{-\sum_{\ell=1}^{P}\left\{\frac{(\beta_{\ell j}-\mu_{\ell j})^2}{2\sigma_{\ell j}^2}\right\}} \right\}$$

The logarithm of the joint prior density function of θ, up to a constant, is

$$\log(g_{CRR}(\theta)) \propto \sum_{j=1}^{K} \left\{ (a_{j1} - 1)\log(\alpha_j) - a_{j2}\alpha_j + (v_{j1} - 1)\log(\lambda_j) - v_{j2}\lambda_j - \sum_{\ell=1}^{P} \left\{ \frac{(\beta_{\ell j} - \mu_{\ell j})^2}{2\sigma_{\ell j}^2} \right\} \right\}$$

3.3.2 Posterior distribution

The posterior density, $g_{CRR}(\theta|\text{data})$ is the product of the joint prior density function and likelihood function. The joint posterior density function of the vector of all unknown parameters θ, up to a constant, as

$$g_{CRR}(\theta|\text{data}) \propto g_{CRR}(\theta) \times L_{CRR}(data|\theta)$$

$$\mathrm{g}_{CRR}(\theta|\mathrm{data}) \propto \sum_{j=1}^{K} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{0}^{\infty} \alpha_j^{a_{j1}} e^{-a_{j2}\alpha_j}\ \lambda_j^{v_{j1}} e^{-v_{j2}\lambda_j}\quad e^{-\sum_{\ell=1}^{P}\left\{\frac{(\beta_{\ell j}-\mu_{\ell j})^2}{2\sigma_{\ell j}^2}\right\}} \times$$

$$\prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left\{ \left\{ e^{\alpha_j(1-e^{t^{\lambda_j}})} \right\}^{e^{\sum_{\ell=1}^{P}\beta_{\ell j}Z_{\ell j}}} \quad t_i^{\lambda_j-1}\ e^{t_i^{\lambda_j}} e^{\sum_{\ell=1}^{P}\beta_{\ell j}Z_{\ell j}} \right\} d\alpha_j d\lambda_j d\beta_{1j} \dots d\beta_{Pj} \qquad (3.15)$$

The normalizing constant is

$$p(y) = \sum_{j=1}^{K} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{0}^{\infty} \alpha_j^{a_{j1}} e^{-a_{j2}\alpha_j}\ \lambda_j^{v_{j1}} e^{-v_{j2}\lambda_j}\quad e^{-\sum_{\ell=1}^{P}\left\{\frac{(\beta_{\ell j}-\mu_{\ell j})^2}{2\sigma_{\ell j}^2}\right\}} \times$$

$$\prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left\{ \left\{ e^{\alpha_j(1-e^{t^{\lambda_j}})} \right\}^{e^{\sum_{\ell=1}^{P}\beta_{\ell j}Z_{\ell j}}} \quad t_i^{\lambda_j-1}\ e^{t_i^{\lambda_j}} e^{\sum_{\ell=1}^{P}\beta_{\ell j}Z_{\ell j}} \right\} d\alpha_j d\lambda_j d\beta_{1j} \dots d\beta_{Pj} \qquad (3.16)$$

For one of the numerical approximation we use here, we need to transform the parameters to real valued. In this case the first 2K elements in θ are the positive parameters that should be transformation. That is, $\theta_j^* = \log(\theta_j), j = 1,2, \dots, 2\mathrm{K}$.

Where

$$\theta_j = \alpha_j, j = 1,2, \dots, \mathrm{K}$$

$$\theta_{K+j} = \lambda_j, j = 1,2, \dots, \mathrm{K}$$

As before, the Jacobian is $e^{\sum_{j=1}^{2K}\theta_j^*}$.

The joint posterior density function of θ^*, given data, is

$$\mathrm{g}_{\mathrm{CRR}}(\theta^*|data) = \mathrm{g}_{\mathrm{CRR}}\left(\theta_1 = e^{\theta_1^*}, \dots, \theta_{2K} = e^{\theta_{2K}^*}, rest\ of\ the\ parameters\ as\ is|\ data\right)\ e^{\sum_{j=1}^{2K}\theta_j^*}$$

Thus, the log-posterior density function of θ^*, given data, is

$$\log(\mathrm{g}_{\mathrm{CRR}}(\theta^*|data)) = \log\left(\mathrm{g}_{\mathrm{CRR}}(\theta = e^{\theta^*}|\ data)\right) + \sum_{j=1}^{2K}\theta_j^*$$

The marginal distributions in equations (3.14) and (3.16) does not have an analytic solution; because of that, we will use numerical techniques to calculate the posterior distribution of θ without calculating the normalized constant. There are several numerical techniques that can be used to do all Bayesian analysis. In this study, we will apply Markov Chain Monte Carlo (MCMC) method to get random draws from the posterior distributions in (3.13) and (3.15) to be able to calculate the Bayesian estimate for all unknown parameters under quadratic loss function.

3.4 Normal approximation

When the log joint posterior density function of the transformed parameters is ready, we will apply optimization algorithms to obtain an approximation to the posterior distribution. For example, we will use the Nelder-Mead method to find the approximate posterior mode and the variance-covariance matrix, which can be used in identifying the proposal distribution in the Metropolis-Hastings algorithms later. To set up for a particular Bayesian inference problem in R, we must define the log posterior density by an R function. Hence, we will use the LearnBayes package in R, which contains the Laplace function. The Laplace function is an efficient technique to summarize a posterior distribution before applying the MCMC method. According to Geisser el at. (1990), the Laplace method described by Erdelyi in 1956 is a frequently used process in statistical theory. In Bayesian inference, researchers use the Laplace method to get different quantities such as approximate posterior expectations, marginal densities, and predictive densities. The aim of using the Laplace method in this study is to calculate the integration for posterior distributions because the posterior distributions for both cases do not have closed form. The Laplace function needs to define the log joint posterior density with intelligent guesses for starting values.

3.5 Markov Chain Monte Carlo Methods

In this section, we will explain briefly some sources, which present a lot of information about Markov Chain Monte Carlo (MCMC) from the past until recently. Tierney (1994) described MCMC as an important technique in Bayesian inference that can be used especially when the posterior distributions do not have a standard form to which numerical integration techniques cannot be applied.

Additionally, Albert (2011) in his book "Bayesian Computation with R" described MCMC methods, which he used to summarize posterior distribution. MCMC is a computing technique, which is used to generate samples from the posterior distribution based on constructing a Markov Chain. Presently, it is widely used in various sciences, such as statistics, biology, computer science, etc. MCMC methods include the Metropolis–Hastings algorithm, Gibbs sampling, Slice sampling, Multiple try Metropolis, and Reversible-jump, but we will focus only on the Metropolis–Hastings algorithm.

3.5.1 Metropolis-Hastings Algorithms

The Metropolis-Hastings (M-H) algorithm is named after the American physicist and computer scientist Nicholas C. Metropolis. The M-H algorithm is straightforward and practical, and can be used to obtain random samples from any complex target distribution of high dimension that is known up to a normalizing constant. The essential idea behind that algorithm is to generate a Markov Chain $\{\theta_t, t = 0,1,2,3, \dots\}$such that its stationary distribution is the target distribution.

According to (Albert, 2011) Markov Chain Monte Carlo methods are describe as a strategic way to simulate draws from a general posterior distribution. Furthermore, there are essential properties for a Markov Chain that are irreducible and periodic. We can say a Markov Chain is irreducible when it moves from any state to any other state, in one or more steps. However, we can say Markov Chain is periodic when it is given in a particular state, and can only return to the same state at a regular interval.

In this part, there are two particular types of Metropolis-Hastings algorithm that are under MCMC method, the independence chain and the random walk chain, but we only focus on the random walk chain. Suppose we are interested in simulating a sample from a posterior density $g(\theta|\text{data})$. Hence, we start with the choice of proposal density, which is in this case a multivariate normal distribution depending on the variance-covariance matrix we got from the Laplace function and the positive scale parameter we selected. We should decide starting values for unknown parameters. Let M, which can be any value, represent the number of random draws of the chain. Here, we will explain the notation for this method.

The following steps summarize Metropolis-Hastings algorithm:

1. The starting point of the chain, θ^0
2. The number of the random draws, M.
3. We need to repeat the following steps for i=1,…,M:

- Set $\theta = \theta^{i-1}$
- Generate a candidate θ^* from a proposal distribution $p(\theta^*|\theta)$
- Calculate the acceptance probability Y as Y = min{1, R} , where

$$R = \frac{g(\theta^*|.)p(\theta|\theta^*)}{g(\theta|.)p(\theta^*|\theta)}$$

- Set $\theta^i = \theta^*$ with probability Y or otherwise set $\theta^i = \theta$.

3.6 Simulation Study

In this section, we use the simulation techniques to test the performance of the methods that are applied in this study to estimate the model parameters. Without loss of generality, we will perform this simulation when K=2. There are various measures can be used to make the comparison between the maximum likelihood and the Markov Chain Monte Carlo methods.

We use the simulation of time-to-event-data using the inversion method to test the performance of the methods that are applied in this study to estimate the model parameters.

This simulation of competing risks data is conducted based on the following scheme:

1. Set the parameters' values of $\theta = (\alpha_1, \lambda_1, \alpha_2, \lambda_2)$.
2. Set different samples size n.
3. Determine the percentage of censored data, say P.
4. Simulate a random sample with different size n from the competing risks model with TPBT (α_j, λ_j), j=1,2, of the risks.
5. Calculate the point estimates and 95% confidence interval for each parameter using the MLE method.
6. Calculate the Bayes estimates and 95% probability interval for each parameter using the Markov Chain Monte Carlo method.
7. Repeat steps 3-6 N times.
8. Compute the mean squared error (MSE) and the coverage probability (CP) of the interval estimates for MCMC and MLE methods using the following formula:

- The MSE of the $\hat{\theta}$

$$MSE(\hat{\theta}) = \sum_{i=1}^{N} \frac{(\hat{\theta}^i - \theta)^2}{N}$$

Where $\hat{\theta}^i = (\hat{\alpha}_1^i, \hat{\lambda}_1^i, \hat{\alpha}_2^i, \hat{\lambda}_2^i)$ is the point estimate of θ using the generated sample in the i^{th} iteration.

- The coverage probability (CP) is

$$CP = \frac{\sum_{i=1}^{N} D_i}{N}$$

Where D_i= 1 if the confidence interval in the i^{th}iteration captures the true parameter and zero otherwise.

We applied the simulation study with 1000 iterations, and specified different sample sizes of n= 25,50,100,200 and 300. We selected the true values of the parameters $\theta = (.5, .8, .5, 1.2)$ because that will give a bathtub shape when $\boldsymbol{\lambda_1 < 1}$ and increasing shape when $\boldsymbol{\lambda_2 \geq 1}$ of the hazard function.

As results show in Table 1 and Table 2, the mean square error of each parameter is decreasing for the Markov Chain Mote Carlo and the maximum likelihood methods. Moreover, the averages of MSE decreases when sample size increases for both methods that were applied. The MSE from both methods for $\alpha_1, \lambda_1, \alpha_2$and $\lambda_2\ provides$ slightly close values when the percentage of censored data equals to zero and 5%. Additionally, as shown in Table 3 and Table 4, the coverage probability remains close to the nominal level of 95% for each parameter in the Markov Chain Monte Carlo method when the percentage of censored data is zero and 5. In general, the coverage probability increases with sample size increases.

		MSE-MCMC				
n	p	α_1	λ_1	α_2	λ_2	Averages
25	0	0.02645	0.06315	0.02916	0.09715	0.21590
50	0	0.01138	0.02381	0.01111	0.03260	0.07890
100	0	0.00549	0.01117	0.00565	0.01493	0.03724
200	0	0.00281	0.00493	0.00260	0.00697	0.01731
300	0	0.00176	0.00367	0.00170	0.00448	0.01160
MSE-MLE						
25	0	0.02510	0.05573	0.02740	0.09262	0.20085
50	0	0.01166	0.02328	0.01266	0.03323	0.08082
100	0	0.00536	0.01056	0.00511	0.01525	0.03629
200	0	0.00250	0.00457	0.00275	0.00671	0.01653
300	0	0.00166	0.00348	0.00170	0.00451	0.01135

Table 1: The mean squared errors for $\alpha_1, \lambda_1, \alpha_2$ and λ_2 by using MCMC and MLE methods at $\theta = (.5, .8, .5, 1.2)$ and p = 0.

		MSE-MCMC				
n	p	α_1	λ_1	α_2	λ_2	Averages
25	5	0.03127	0.08256	0.03022	0.10767	0.25173
50	5	0.01218	0.02559	0.01185	0.03395	0.08357
100	5	0.00569	0.01138	0.00575	0.01717	0.03999
200	5	0.00262	0.00531	0.00272	0.00754	0.01819
300	5	0.00157	0.00333	0.00176	0.00455	0.01122
MSE-MLE						
25	5	0.02477	0.07130	0.03361	0.11886	0.24854
50	5	0.01128	0.02471	0.01095	0.03361	0.08055
100	5	0.00558	0.01192	0.00535	0.01477	0.03763
200	5	0.00295	0.00534	0.00294	0.00710	0.01832
300	5	0.00224	0.00340	0.00218	0.00428	0.01209

Table 2: The mean squared errors for $\alpha_1, \lambda_1, \alpha_2$ and λ_2 by using MCMC and MLE methods at $\theta = (.5, .8, .5, 1.2)$ and p = 5.

		CP-MCMC			
n	p	α_1	λ_1	α_2	λ_2
25	0	92.7	92.3	91.5	91.6
50	0	92.6	93.2	91.9	91.8
100	0	93.0	92.8	92.5	92.9
200	0	93.5	93.0	93.2	93.7
300	0	94.3	93.2	93.2	94.6
CP-MLE					
25	0	91.3	91.5	90.1	90.3
50	0	92.8	92.1	92.7	90.1
100	0	93.0	92.5	93.0	91.0
200	0	93.7	93.9	92.8	92.5
300	0	94.0	94.4	93.5	93.8

Table 3: The coverage probability (CP) for $\alpha_1, \lambda_1, \alpha_2$ and λ_2 by using MCMC and MLE methods at $\theta = (.5, .8, .5, 1.2)$ and p = 0.

		CP-MCMC			
n	p	α_1	λ_1	α_2	λ_2
25	5	91.2	90.0	91.3	90.6
50	5	92.8	91.1	92.0	90.6
100	5	91.2	92.7	93.5	92.0
200	5	93.0	93.0	91.1	92.3
300	5	94.1	93.8	93.1	92.7
CP-MLE					
25	5	90.6	90.2	90.1	90.0
50	5	91.9	92.6	90.0	90.9
100	5	92.0	93.5	92.0	92.1
200	5	93.5	94.0	93.3	93.2
300	5	94.3	94.3	94.0	93.1

Table 4: The coverage probability (CP) for $\alpha_1, \lambda_1, \alpha_2$ and λ_2 by using MCMC and MLE methods at $\theta = (.5, .8, .5, 1.2)$ and p = 5.

Chapter 4: Applications

4.1 Introduction

In this section, the maximum likelihood and MCMC methods will be used to estimate the parameters of all the risks and some of the reliability measures of the system for both the competing risks model and competing risks regression model. Additionally, this study will include two real-life datasets. The first data set is from the Amsterdam Cohort Studies on HIV infection and AIDS, which considers a particular problem about HIV infection among men who have sex with men (MSM) (Koblin et al., 2003). There are various types of studies, which analyzed data of HIV infection using different techniques in statistics to explain the high risk of this behavior to physical health. The second data set is survival time of electrical appliances from Lawless (2003). All analysis in this chapter is by using R.

4.2 AIDSSI dataset

We will analyse AIDSSI dataset, which is available in "mstate" package in R program, about human immunodeficiency virus (HIV) infection in the Amsterdam Cohort Studies. The total participants in this study were 329 selected from the Amsterdam Cohort (see Geskus *et al.,* 2000 and 2003). This study explained high-risk behaviours, which are due to different physical health problems, among men who have sex with men. There are only two causes of failure, which are AIDS and syncytium-inducing (SI) HIV phenotype. We used the same dataset to analyse both case study 1 and case study 2. In the next subsection, we explain each case and what variables will be considered. We are interested in the time to failure whatever the cause of failure, which competes to attack or kill participants, was. Additionally, in this situation, only one cause can occur but we do not know which cause will occur first. Table 5 shows more details about each variable.

The variable	Definition
Time	This variable presents time from HIV infection to first of SI appearance and AIDS, or last follow-up
Status	This variable gives indicator for each event such as: 0 = censored, 1 = AIDS, 2 = SI appearance
CCR 5	This is C-C chemokine receptor type 5, also known as CCR5 which is has two level "WW" (wild type allele on both chromosomes), "WM"(mutant allele on one chromosome), we give indicator for each levels:0=”WW” and 1= “WM”.
Age	This variable presents age at HIV infection.

Table 5: The description of all variables listed in the AIDISS dataset.

4.2.1 Case study 1

In this case, we will explain the competing risks model that considers only the time to event and the status. Let (T, C, δ) be random vectors, where T denotes failure time, C denotes cause of failures and δ denotes indicator status. If the causes of failure are observed ($\delta = 1$), but otherwise ($\delta = 0$) which represents Type-I censoring. In the competing risks model each patient is under two risks of failure, which are known as acquired immunodeficiency syndrome (AIDS) and syncytium inducing (SI) HIV phenotype. Furthermore, we need to specify particular assumptions to explain competing risks model:

Suppose the TPBT lifetime distribution for each cause of failure $T_i = \min\{T_{i1}, T_{i2}\}$, where $T_{i1} \sim TPBT\ (\alpha_1, \lambda_1)\ and\ T_{i2} \sim TPBT\ (\alpha_2, \lambda_2)$, and only observe the T_i .

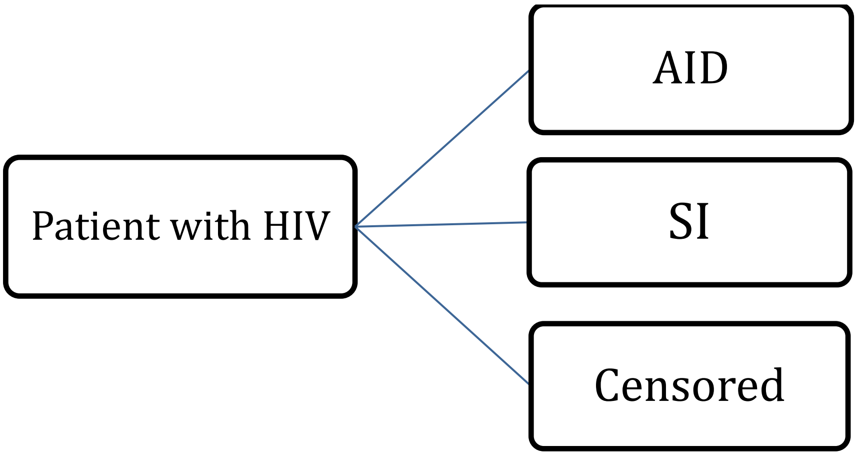


Figure 2: The diagram explains the variables of competing risks model.

4.2.1.1 The result of competing risks

In this study, we used TPBT model to analyze AIDSSI data set by using the maximum likelihood and Bayesian methods. In order to calculate the asymptotic confidence intervals for each parameter, we computed the inverse of the Fisher information matrix that approximates the variance-covariance matrix for the maximum likelihood estimates of the vector of unknown parameters $\theta = (\alpha_1, \boldsymbol{\lambda_1}, \boldsymbol{\alpha_2}, \boldsymbol{\lambda_2})$. The maximum likelihood estimates and 95% confidence intervals of the four model parameters are shown in Table 6 and Table 7.

Method	$\boldsymbol{\alpha_1}$	$\boldsymbol{\lambda_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\lambda_2}$
MLE	0.01346	0.56534	0.02122	0.50220
BE	0.013730	0.5645	0.02160	0.5004

Table 6: Estimates the parameters by maximum likelihood and Bayesian methods.

The inverse of the Fisher information matrix:

$$I^{-1} = \begin{pmatrix} 7.613204e-06 & -5.128713e-05 & 2.048718e-16 & -1.317940e-15 \\ -5.128713e-05 & 4.353662e-04 & -1.801692e-15 & 1.159028e-14 \\ 2.048718e-16 & -1.801692e-15 & 1.698079e-05 & -8.252388e-05 \\ -1.317940e-15 & 1.159028e-14 & -8.252388e-05 & 5.308760e-04 \end{pmatrix}$$

	Confidence intervals	
Parameters	**Lower bound**	**Upper bound**
$\boldsymbol{\alpha_1}$	0.008051313	0.01886739
$\boldsymbol{\lambda_1}$	0.524447107	0.60623959
$\boldsymbol{\alpha_2}$	0.013147822	0.02930126
$\boldsymbol{\lambda_2}$	0.457045017	0.54736474

Table 7: The asymptotic confidence intervals at 95% using maximum likelihood method.

Under Bayes methods, we applied MCMC technique. We suppose all unknown parameters in this model to be independent random variables and follow gamma distribution when all the hyper-parameters are known and equal to 0.001 to reflect non-informative prior. We assumed a small value for all hyper-parameters because of the ”low information” for selecting prior distribution, which is reasonable since its mean equals one while the variance is 1000.We used the normal approximation method to estimate the unknown parameters. To apply MCMC, the proposal density is a multivariate normal distribution with mean vector zero and the variance-covariance matrix was obtained from the normal approximation method and a positive value for scale parameter as 0.7.

We applied Metropolis random walk algorithm to simulate 10000 samples of draws from the posterior distribution. The acceptance rate of the draws was 51.66%. We discarded the first 50% of the draws. The rest of the draws is used to conduct inferences on the four unknown parameters by computing Bayes point estimates and the 95% credible intervals as shown in Table 6, Table 8 and Table 9. Further more, the marginal posterior distributions of the four parameters are estimated as shown Figure 3. The measures of central tendency for α_1 and α_2 both have means larger than their medians, because that histogram is right-skewed, but the measures of central tendency for λ_1 and λ_2 both have means larger than their medians, because that histogram is right-skewed.

	Probability intervals	
Parameters	Lower bound	Upper bound
α_1	0.008711196	0.019453637
λ_1	0.5256610	0.6078643
α_2	0.01423233	0.03050542
λ_2	0.4570505	0.5452463

Table 8: The 95% credible intervals for the four parameters.

Measures	α_1	λ_1	α_2	λ_2
Min.	0.005909	0.4933	0.01021	0.4321
First Quartile	0.011680	0.5488	0.01841	0.4838
Median	0.013550	0.5637	0.02133	0.4994
Mean	0.013730	0.5645	0.02160	0.5004
Third Quartile	0.015660	0.5794	0.02437	0.5172
Max.	0.027040	0.6310	0.03806	0.5802

Table 9: The measures of central tendency of all model parameters.

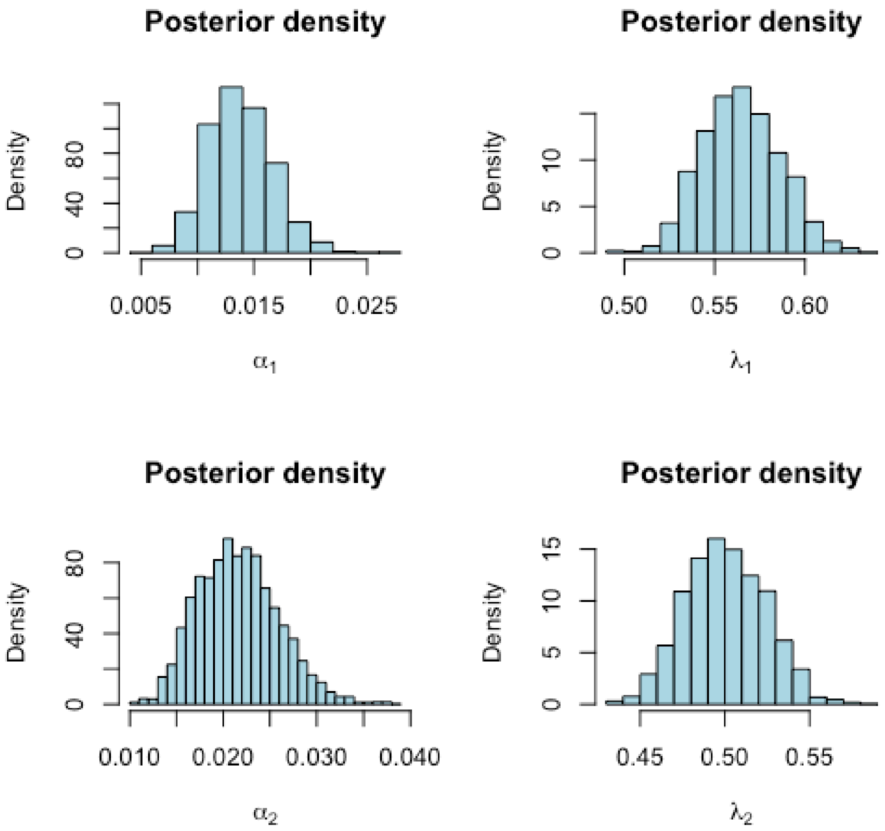


Figure 3: Estimated marginal posterior distributions of the model parameters.

Furthermore, all diagnostics test results on the simulations draws, which is instructive to determine if they approximately represent the posterior distribution of interest. Figure 4, the trace plots, shows the simulated draws of the transformed parameters with a poor choice of start values and scale factor. We plotted the same graph in Figure 5, but after discarding the early 5000 draws due to the poor choice of start values which shows better combination of simulated draws. Also, the autocorrelation plots in Figure 6, which show the lag decreases to zero very quickly, which mean the draws become more independent over time. The estimates of the relative risk rate of each cause of failure in the presence of all other causes are given in Table 12. The trace plots of the draws of $\pi_j, j = 1,2$, and the corresponding marginal posterior density functions are provided in Figure 7.

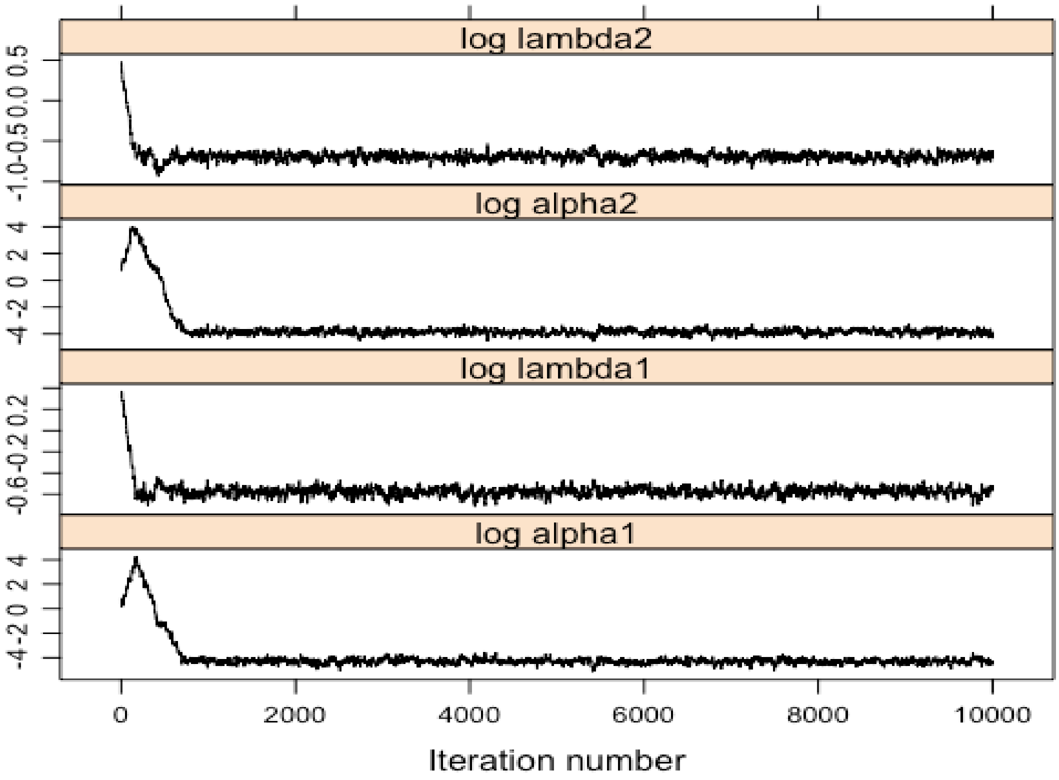


Figure 4: The trace plots of the10000 simulated draws of the transformed parameters.

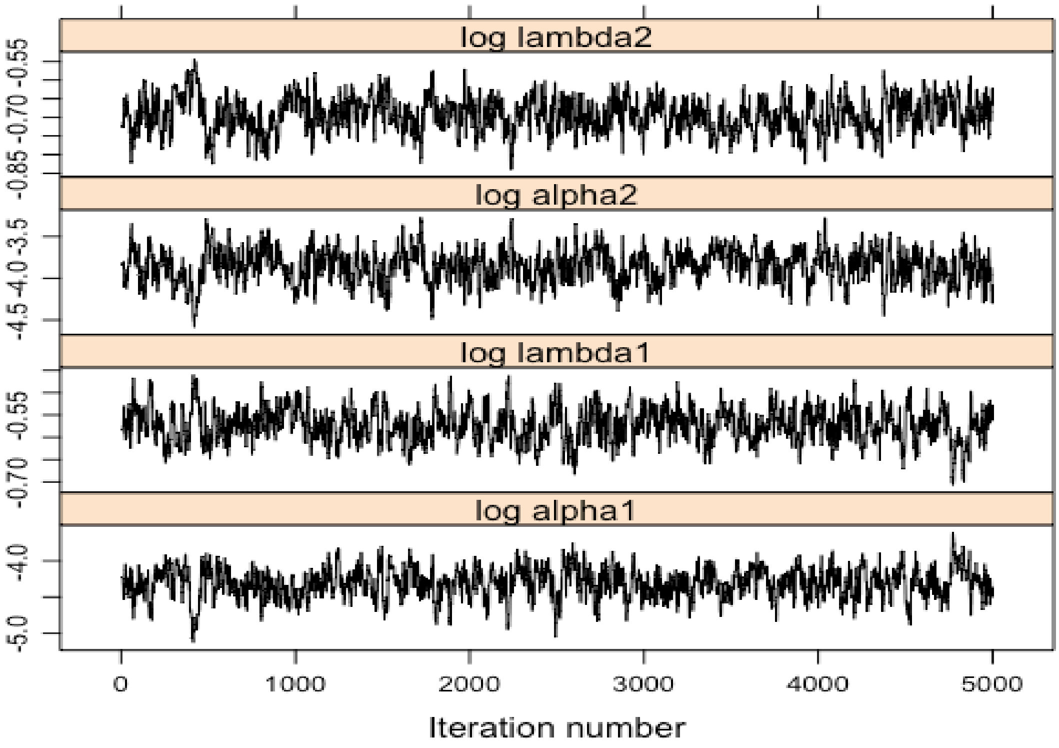


Figure 5: The simulated draws after discarding the early 50% of the draws.

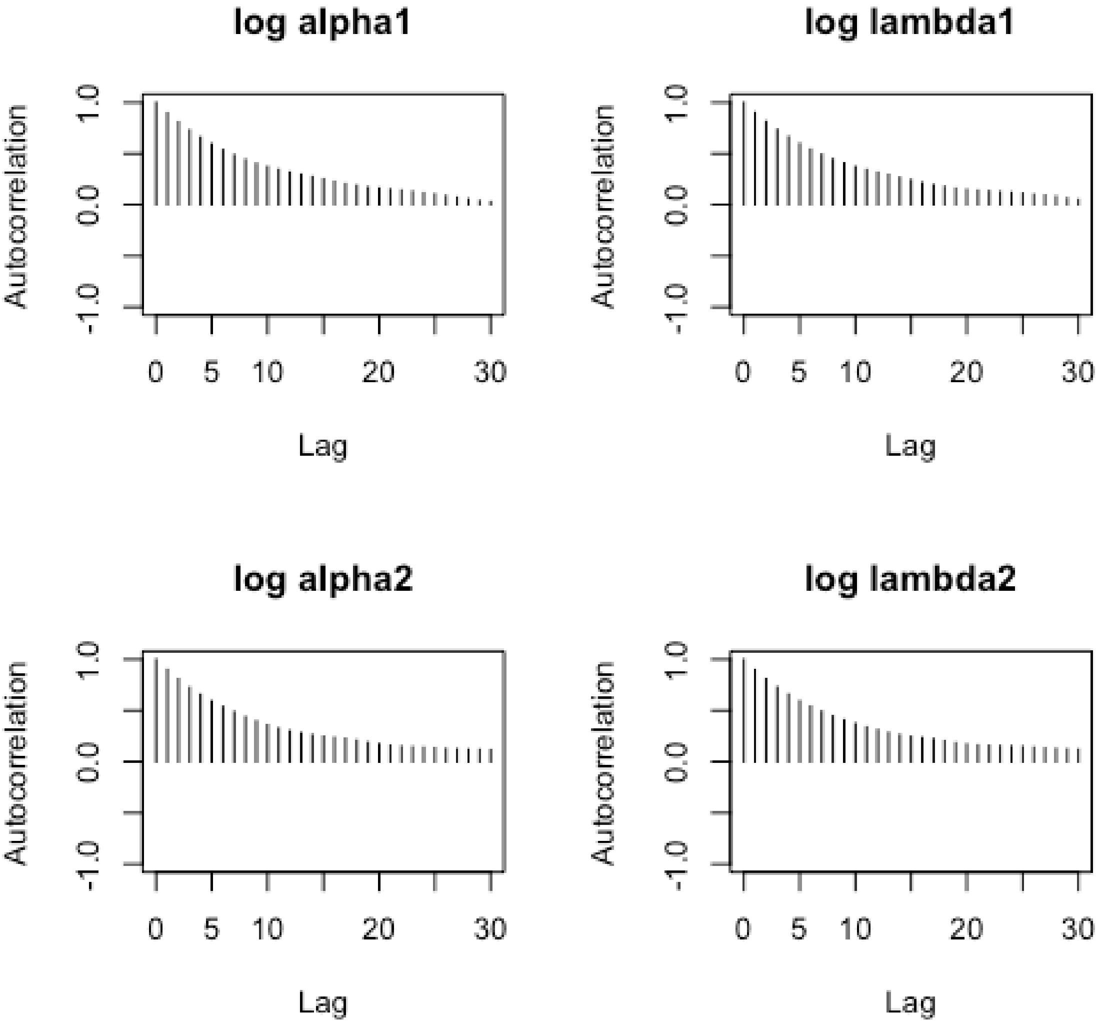


Figure 6: The autocorrelation plots of simulated draws after discarding the early 50% of the draws.

Methods	$\hat{\pi}_1$	$\hat{\pi}_2$
MLE	0.5431876	0.4568124
BE	0.5451378	0.4548611

Table 10: The estimation of the relative risk for the two causes.

Measures	$\pi_1(t)$	$\pi_2(t)$
Min.	0.4094	0.3417
First Quartile	0.5217	0.4316
Median	0.5458	0.4542
Mean	0.5451	0.4549
Third Quartile	0.5684	0.4783
Max.	0.6583	0.5906

Table 11: The measures of the central tendency for both two relative risks.

	Probability intervals	
The risks	Lower bound	Upper bound
$\hat{\pi}_1$	0.4756934	0.6131275
$\hat{\pi}_2$	0.3868725	0.5243066

Table 12: The 95% credible intervals for both two relative risks.

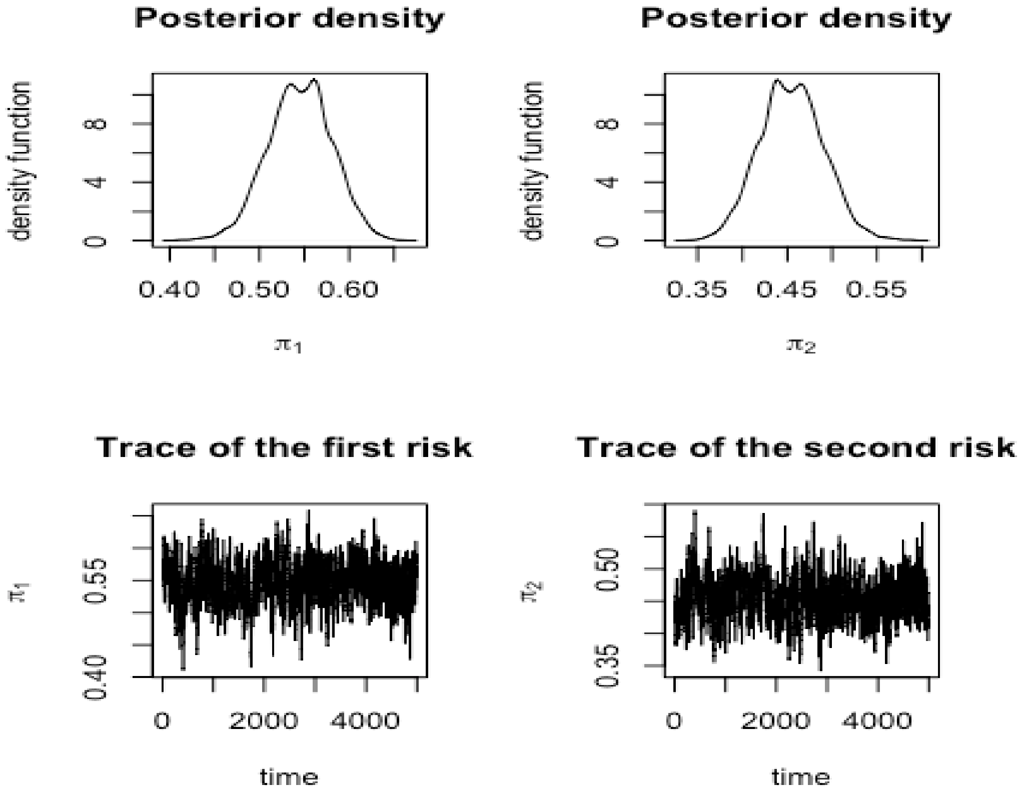


Figure 7: The marginal posterior density and trace plots for each relative risk $\pi_j(t)$,j=1,2.

Also, we estimated the hazard function and survival function of each risk at the mean of the sample (say t_0) are estimated using the maximum likelihood and MCMC methods as shown in Tables 13-16 and Figures 8 and 9. The measures of central tendency for hazard functions all have means larger than their medians, because the histogram shape are right-skewed.

Methods	$h_1(t_0)$	$h_2(t_0)$
MLE	0.06487106	0.05708413
BE	0.06469	0.05641

Table 13: The estimation of the hazard function by maximum likelihood and MCMC methods.

Measures	$h_1(t)$	$h_2(t)$
Min.	0.04484	0.03916
First Quartile	0.06065	0.05241
Median	0.06461	0.05630
Mean	0.06469	0.05641
Third Quartile	0.06863	0.06027
Max.	0.08398	0.07891

Table 14: The measures of the central tendency for the hazard functions of both two risks.

	Probability intervals	
Hazard Functions	Lower bound	Upper bound
$h_1(t)$	0.05274782	0.07667317
$h_2(t)$	0.04537678	0.06778769

Table 15: The 95% credible intervals for the hazard functions of both two risks.

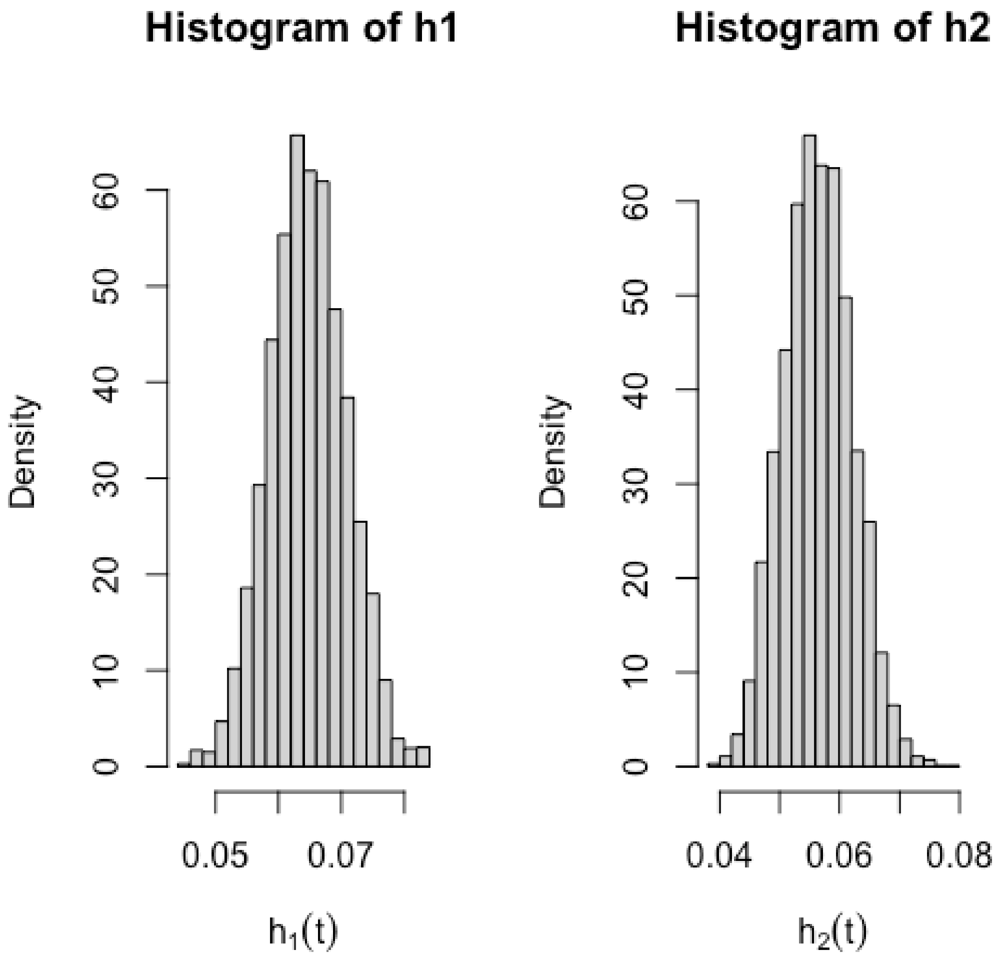


Figure 8: The posterior density for the hazard functions of both two risks.

Methods	$S_1(t)$	$S_2(t)$
MLE	0.7769016	0.7585281
BE	0.7768	0.7598

Table 16: The estimation of the survival function by maximum likelihood and MCMC methods.

Measures	$S_1(t)$	$S_2(t)$
Min.	0.7070	0.6810
First Quartile	0.7620	0.7458
Median	0.7773	0.7602
Mean	0.7768	0.7598
Third Quartile	0.7910	0.7742
Max.	0.8543	0.8237

Table 17: The measures of the central tendency for the survival functions of both two risks.

	Probability intervals	
Hazard Functions	Lower bound	Upper bound
$S_1(t)$	0.7360718	0.8188732
$S_2(t)$	0.7147731	0.8028594

Table 18: The 95% credible intervals for the survival functions of both two risks.

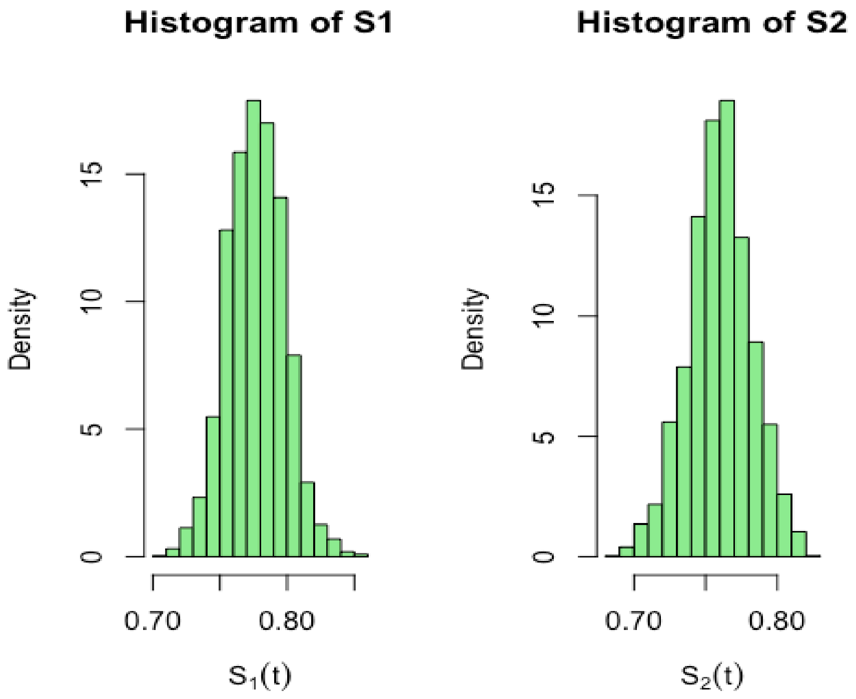


Figure 9: The posterior density for both two survival functions of the risks.

4.2.*2* Case study 2

In this subsection, we will explain each variable that is used in the competing risks regression model. For each patient the following variables are observed: the time to event, indicator status, C-C chemokine receptor type 5 (CCR5) level and age at HIV infection in years. Figure 10, the diagram explains the data for the competing risks regression model. We will use the competing risks regression model when the cause specific hazard function is based on the covariates. Hence, we assumed the shape of baseline hazard function $h_{j0}(t)$, j=1,2, equal to the two-parameter bathtub distribution.

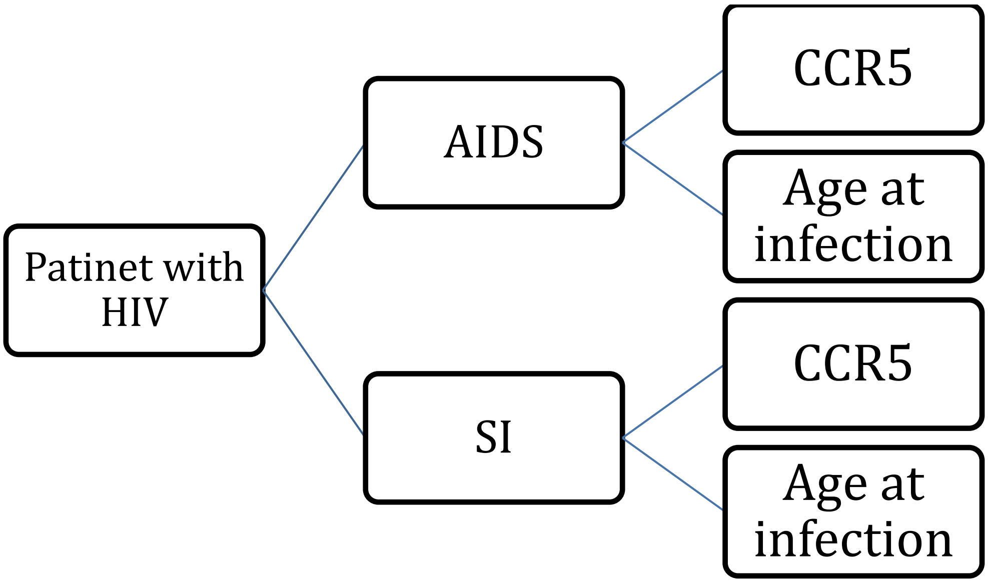


Figure 10: The diagram explains the variables of competing risks regression model.

4.2.2.1 The result competing risks regression

In this study, we used the TPBT model to analyze the AIDSSI data set by using the maximum likelihood and Bayesian methods. In order to calculate the asymptotic confidence intervals for each parameter, we computed the inverse of the Fisher information matrix to approximate the variance-covariance matrix for the maximum likelihood estimates of the vector of unknown parameters $\theta = (\alpha_1, \lambda_1, \alpha_2, \lambda_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})$. The maximum likelihood estimates and 95% confidence intervals of the four model parameters are shown in Table 19 and Table 20.

Parameters	MLE	MCMC
α_1	0.01387	0.00943
λ_1	0.27238	0.5814
α_2	0.01633	0.01446
λ_2	0.46619	0.5055
β_{11}	0.07006	-1.2846
β_{12}	0.04401	0.01666
β_{21}	0.48551	-0.3326
β_{22}	0.0071	0.0161

Table 19: Estimates the parameters by the maximum likelihood and Bayesian methods.

	Confidence intervals	
Parameters	Lower bound	Upper bound
α_1	0.00839	0.01933
λ_1	0.22543	0.3193
α_2	0.00834	0.02430
λ_2	0.41766	0.51472
β_{11}	-0.4079	0.54812
β_{12}	0.03165	0.05636
β_{21}	0.0685	0.90248
β_{22}	-0.0088	0.02319

Table 20: The 95% confidence intervals of the eight model parameters.

Under Bayes Method, we applied MCMC technique. We suppose all unknown parameters in this model assumed to be independent random variables and follow gamma distribution when all the hyper-parameters are known and equal to 0.001 that reflects non-informative prior. We assumed small value for all hyper-parameter because of the ”low information” prior that is a reasonable since its mean equal one while the variance is 1000.We used the normal approximation method to estimate the unknown parameters.

To apply MCMC, the proposal density is a multivariate normal distribution with mean vector zero and the variance-covariance matrix that obtained from the normal approximation method and a positive value for scale parameter as 0.7. We applied Metropolis random walk algorithm to simulate 10000 samples of draws from the posterior distribution. The acceptance rate of the sample draws was 21.99%. We discarded the first 50% of the draws. The rest of the draws is used to conduct inferences on the eight unknown parameters by computing Bayes point estimates and the 95% credible intervals model as shown in Table 19, Table 21 and Table 22. Furthermore, the marginal posterior distributions of the eight parameters are estimated as show in Figure 11. The measures of central tendency for α_1 and α_2 both have means larger than their medians because that the shape of histogram is right-skewed, but the rest of parameters have means equal to their medians, because that the shape of histogram is symmetric.

	Probability intervals	
Parameters	Lower bound	Upper bound
α_1	0.003541812	0.022076149
λ_1	0.5445942	0.6260562
α_2	0.004886771	0.036369654
λ_2	0.4682408	0.5628071
β_{11}	-1.8973189	-0.6908312
β_{12}	-0.01174524	0.03736942
β_{21}	-0.8086315	0.1119790
β_{22}	-0.01561741	0.04017148

Table 21: The 95% credible intervals for the eight parameters.

Measures	α_1	λ_1	α_2	λ_2	β_{11}	β_{12}	β_{21}	β_{22}
Min.	0.002151	0.5192	0.002715	0.4379	-2.380	-0.026561	-0.9663	-0.028941
First Quartile	0.006376	0.5680	0.008817	0.4932	-1.476	0.007528	-0.4865	0.007424
Median	0.008319	0.5845	0.012419	0.5104	-1.262	0.015298	-0.3162	0.015675
Mean	0.009597	0.5828	0.014261	0.5107	-1.282	0.015144	-0.3202	0.015673
Third Quartile	0.011234	0.5966	0.017064	0.5258	-1.052	0.024111	-0.1418	0.024030
Max.	0.034226	0.6421	0.090389	0.5731	-0.418	0.050399	0.4098	0.052792

Table 22: The measures of central tendency of all model parameters.

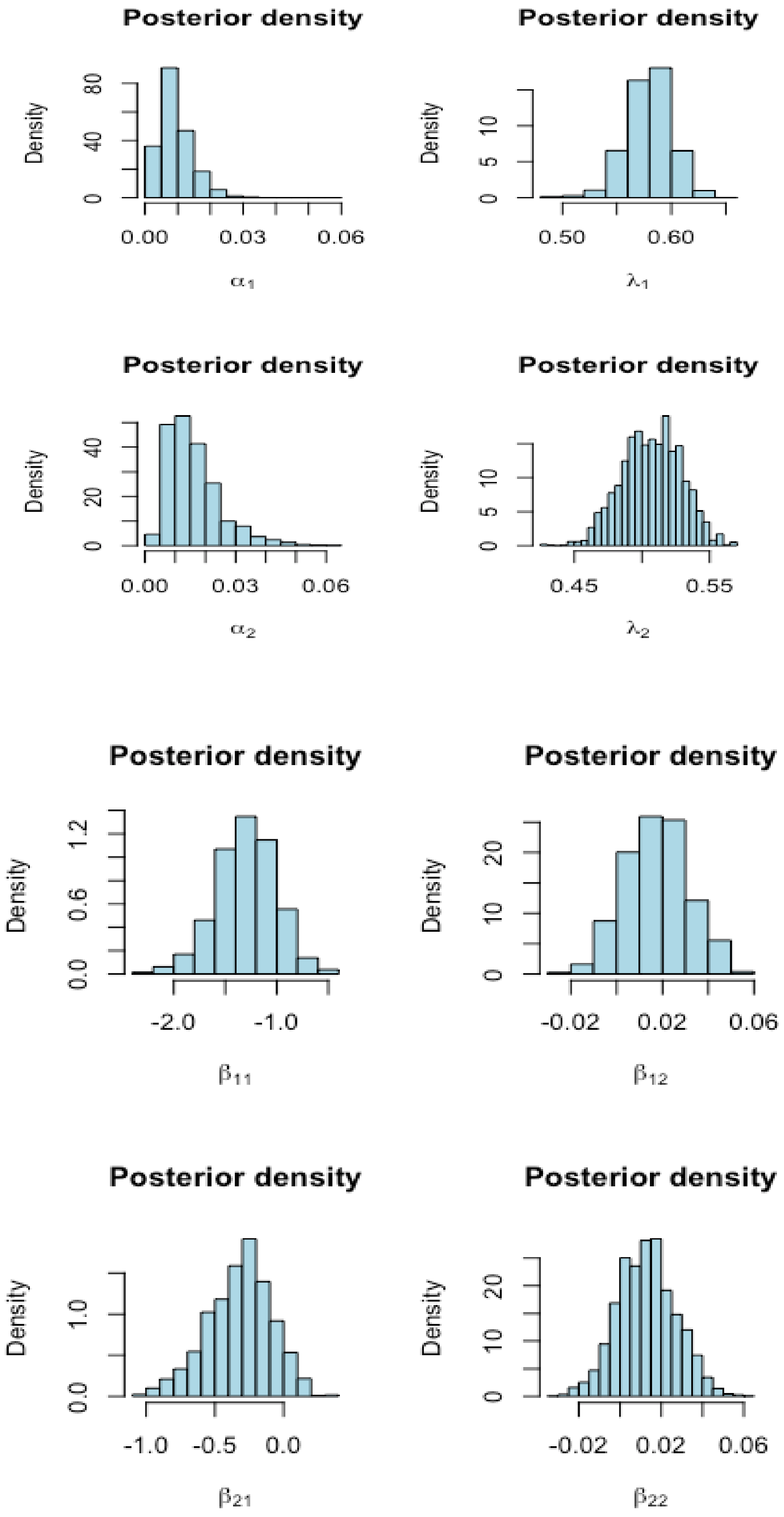


Figure 11: Estimated marginal posterior distributions of the model parameters.

Furthermore, all diagnostics test results on the simulations draws, which is instructive to determine if they approximately represent the posterior distribution of interest are provided. Figure 12, the trace plots show the simulated draws of the transformed parameters with a good choices of start values and scale factor. We plotted the same graph in Figure 13, but after discarding early the 5000 draws due to the poor choice of start values and to show better combination of simulated draws than before. Hence, we provided autocorrelation plots of simulated draws of $\alpha_1, \lambda_1, \alpha_2$ and $\lambda_2$for the random walk chain after discarding the first 50% of the draws in Figure 14. Additionally, the autocorrelation plots show the lag decreases to zero very quickly, which means that the draws become independent over time. The estimates of the relative risk rates of each cause of failure in the presence of all other causes are shown in Table 23.

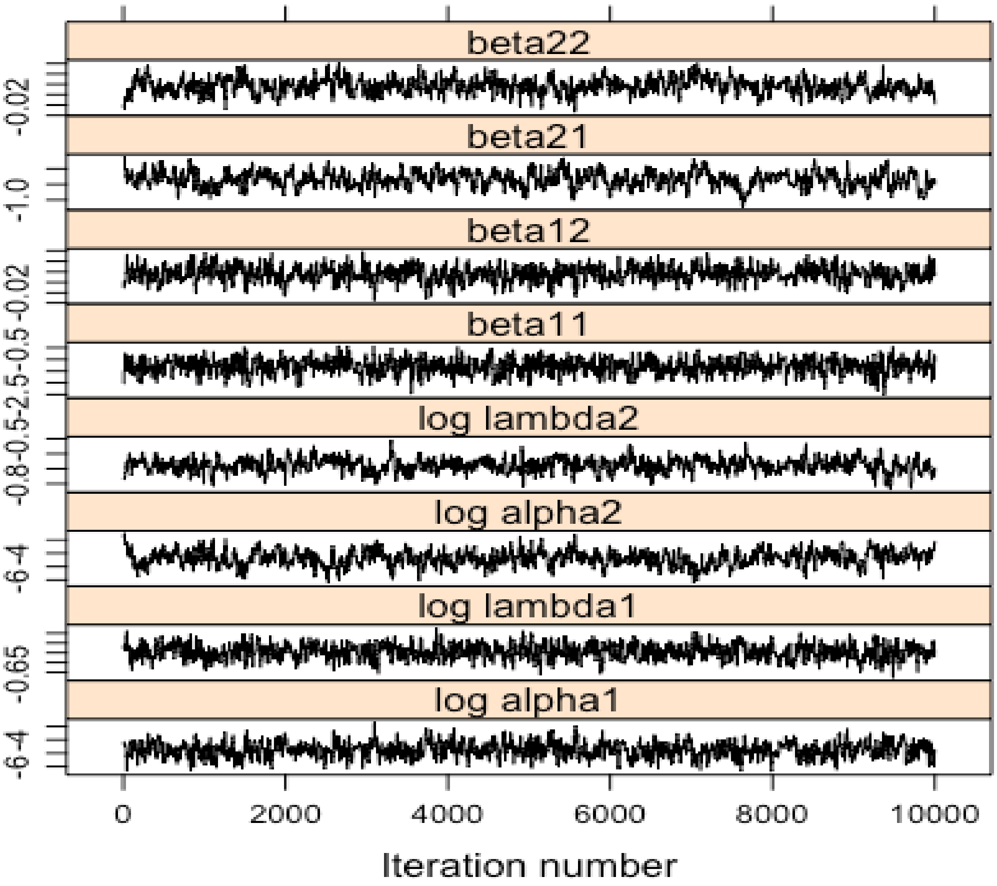


Figure 12: The trace plots of the10000 simulated draws of the transformed parameters.

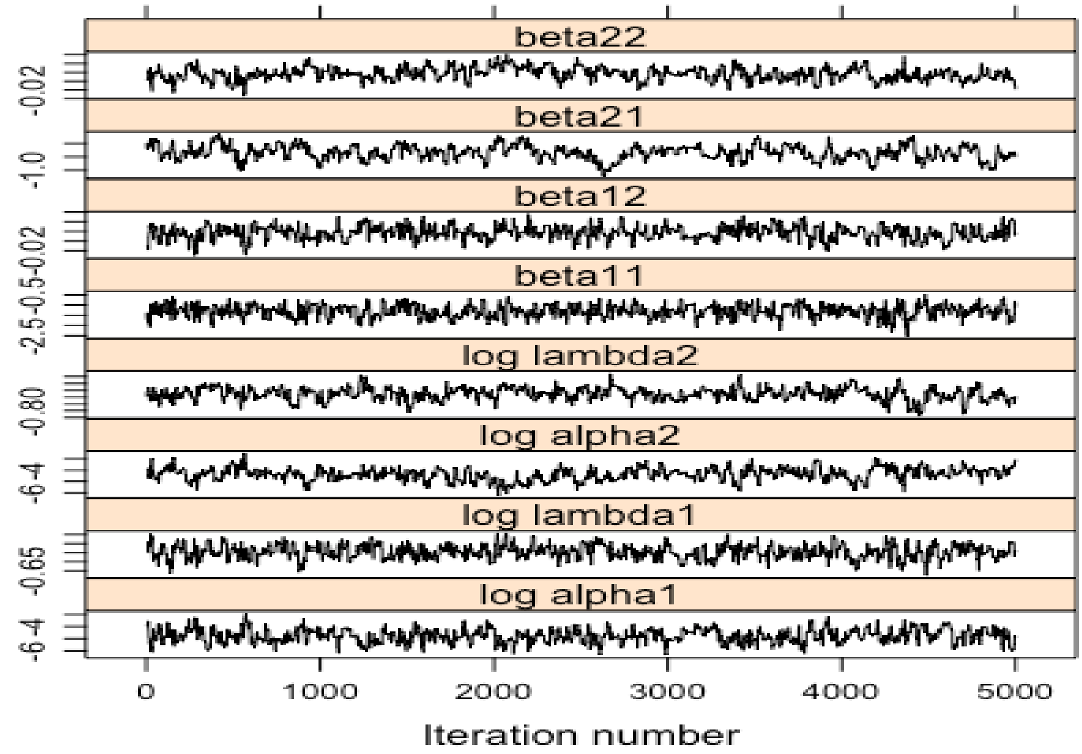


Figure 13: The simulated draws after discarding the early 50% of the draws.

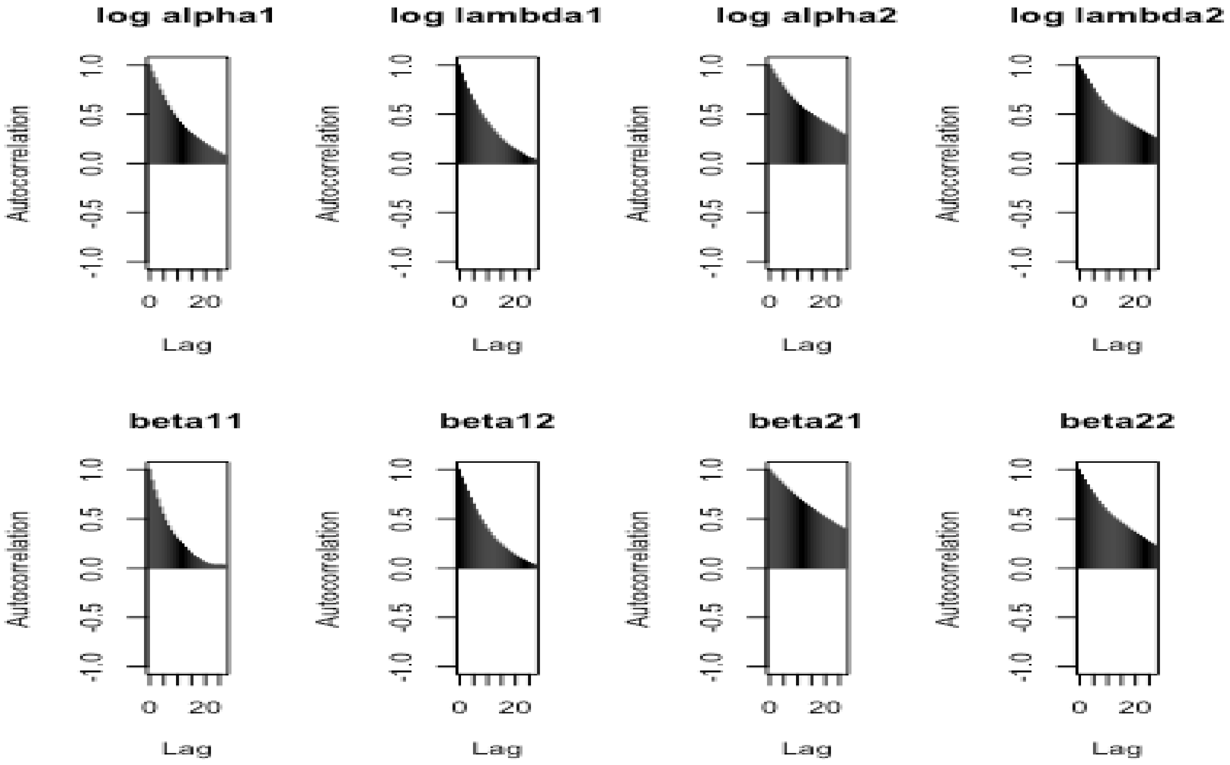


Figure 14: The autocorrelation plots of simulated draws after discarding the early 50% of the draws.

Methods	$\hat{\pi}_1$	$\hat{\pi}_2$
MLE	0.3078404	0.6921596
BE	0.4084329	0.5915671

Table 23: The estimation of the two causes at the mean of age infection and CCR5 is "WM".

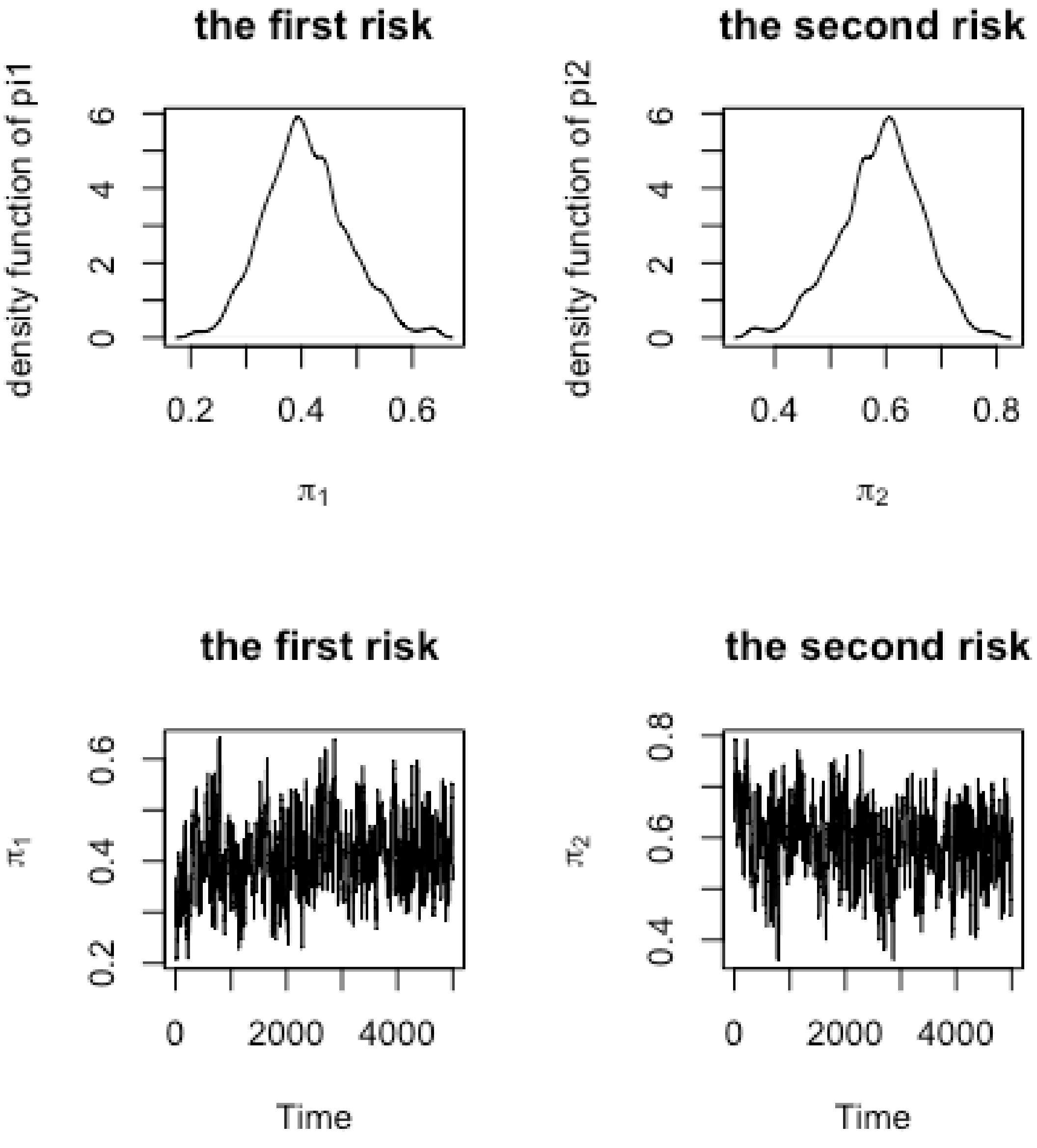


Figure 15: The marginal posterior density and trace plots for each relative risk at the mean of age infection and CCR5 is "WM".

Methods	$\hat{\pi}_1$	$\hat{\pi}_2$
MLE	0.01060627	0.9893937
BE	0.2230407	0.7769593

Table 24: The estimation of the two causes at the mean of age infection and CCR5 is "WW".

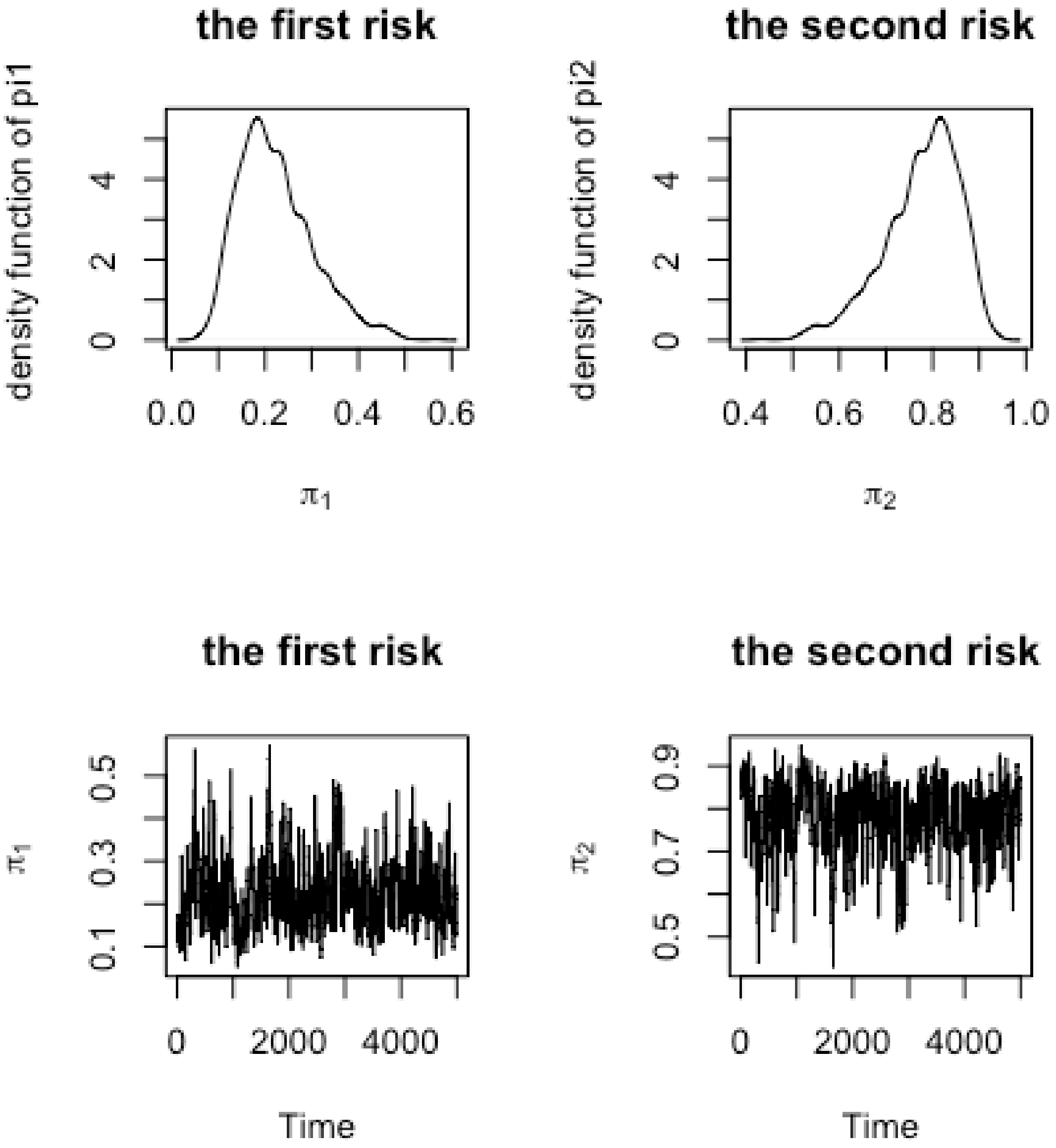


Figure 16: The marginal posterior density and trace plots for each relative risk at the mean of age infection and CCR5 is "WW".

Based on the output from the MCMC method, we selected the last 50% of the simulated draws which come from the posterior density for each parameter. As shown in Table 25, we estimated hazard function by using the maximum likelihood and MCMC methods and computed the measures of central tendency shown in Table 26 and the 95% credible intervals for the hazard function at each cause of failure shown in Table 27. Hence, we can use the measures of central tendency in Table 26 to describe the histogram shapes in Figure 17. The measures of central tendency for hazard functions all of them have means larger than the medians because that the histogram shape is right-skewed.

Methods	$h_1(t_0)$	$h_2(t_0)$
MLE	0.05323	0.033724
BE	0.06478	0.05888

Table 25: The estimation hazard function by maximum likelihood and MCMC methods.

Measures	$h_1(t)$	$h_2(t)$
Min.	0.04592	0.04185
First Quartile	0.05983	0.05475
Median	0.06443	0.05886
Mean	0.06478	0.05888
Third Quartile	0.06891	0.06309
Max.	0.09001	0.07478

Table 26: The measures of central tendency for hazard functions.

	Probability intervals	
Hazard Functions	Lower bound	Upper bound
$h_1(t)$	0.05207902	0.07798570
$h_2(t)$	0.04717179	0.07134012

Table 27: The 95% credible intervals for hazard functions.

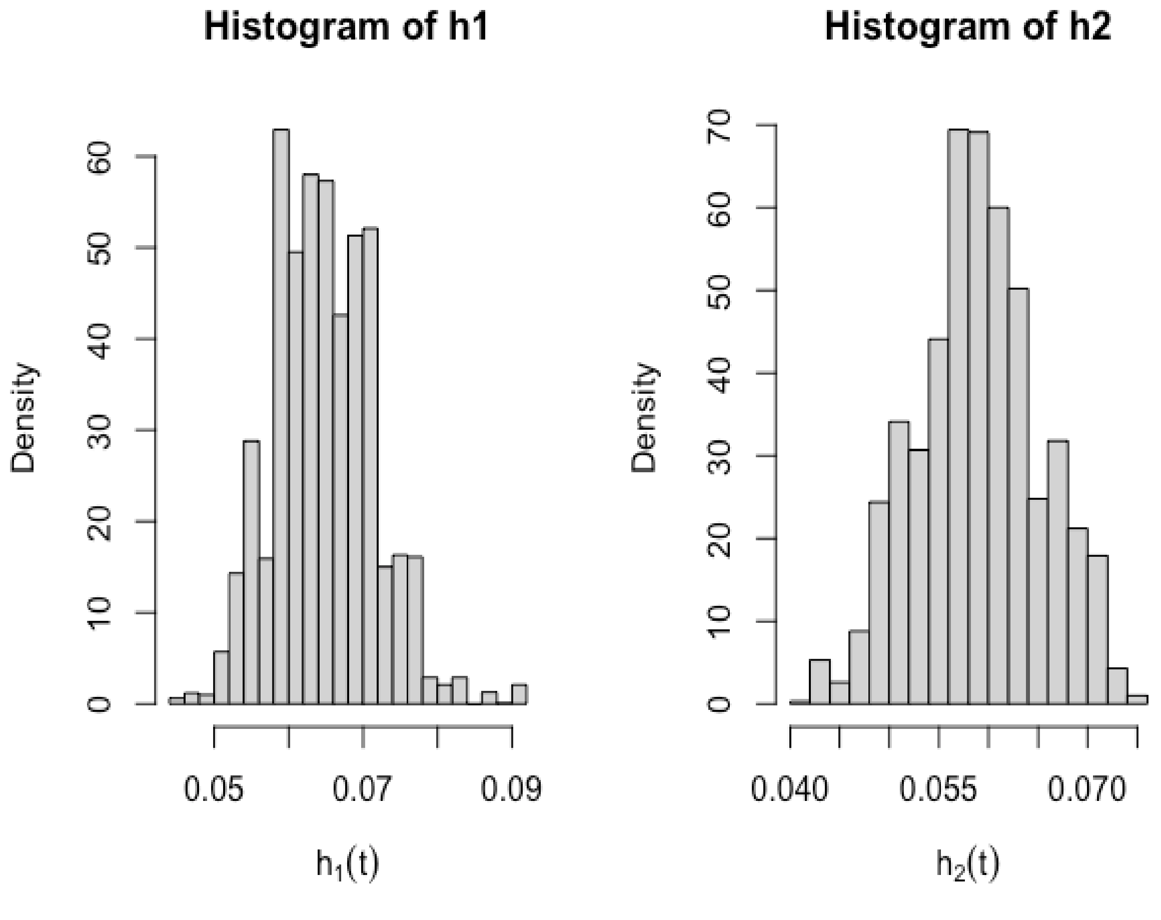


Figure 17: The posterior density for the hazard functions of both two risks evaluated at the mean of the available data.

As shown in Table 28, we estimated survival function by using the maximum likelihood and MCMC methods and computed the measures of central tendency shown in Table 29 and 95% credible intervals for the survival function for each cause of failure shown in Table 30. The measures of central tendency for survival functions both have mean equal to the median because that the histogram shape is symmetric as shown in Figure 18.

Methods	$S_1(t)$	$S_2(t)$
MLE	0.80661	0.72594
BE	0.7884	0.7585

Table 28: The point estimate of survival function by maximum likelihood and MCMC methods.

Measures	$S_1(t)$	$S_2(t)$
Min.	0.7231	0.7009
First Quartile	0.7738	0.7448
Median	0.7891	0.7576
Mean	0.7884	0.7585
Third Quartile	0.8048	0.7703
Max.	0.8577	0.8251

Table 29: The measures of central tendency for survival functions.

	Probability intervals	
Hazard Functions	Lower bound	Upper bound
$S_1(t)$	0.7436079	0.8326959
$S_2(t)$	0.7140227	0.8010096

Table 30: The credible intervals at 95% for survival functions.

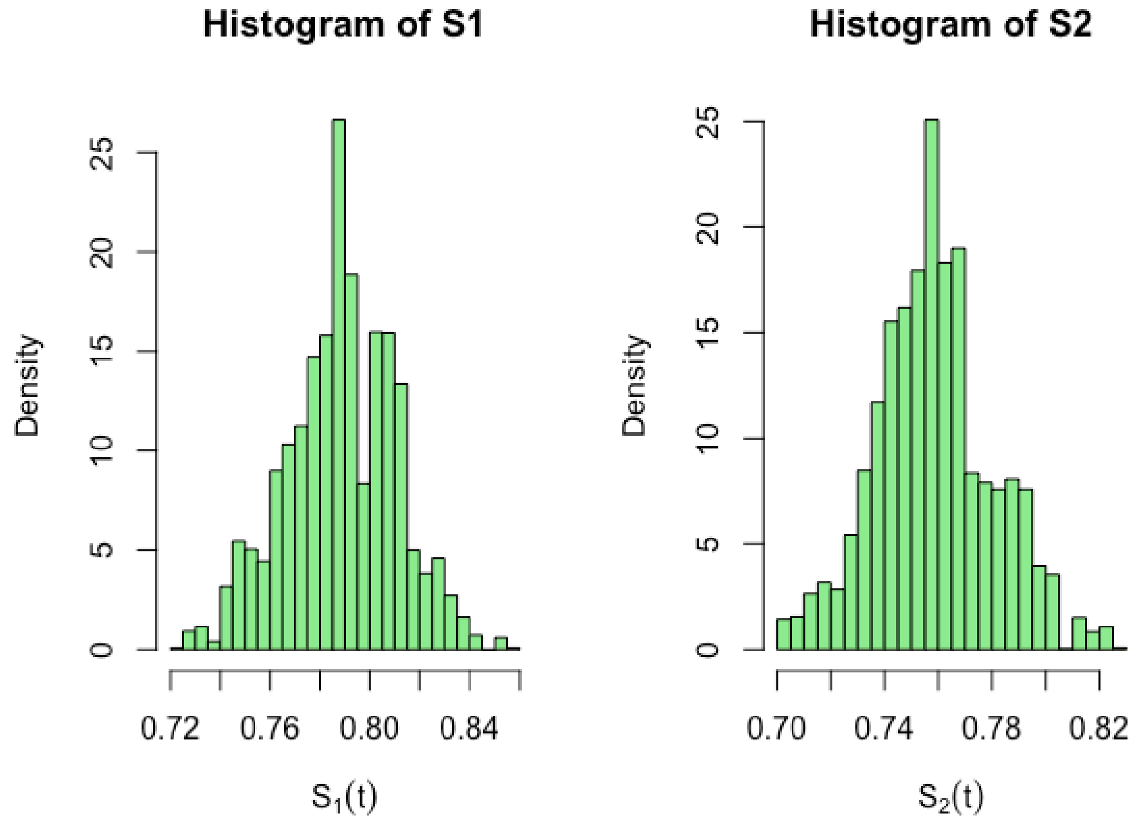


Figure 18: The posterior density for both two survival functions of the risks evaluated at the mean of the available data.

4.3 Survival times of electrical appliances

In this section, we will analyse the survival times of the electrical appliances data set containing the cause of failure and censored data from Lawless (2003). The total number of appliance subjects is 36, which are under an automatic life test. The data consist of two causes of failure; the first cause is failure mode 9, and the second cause includes all other failure modes.

4.3.1 The result of competing risks model

In this study, we used the TPBT model to analyze the survival times of electrical appliances data set by using the maximum likelihood and Bayesian methods. In order to calculate the asymptotic confidence intervals for each parameter, we computed the inverse of the Fisher information matrix to approximate the variance-covariance matrix for the maximum likelihood estimates of the vector of unknown parameters $\theta = (\alpha_1, \lambda_1, \alpha_2, \lambda_2)$. The maximum likelihood estimates and 95% confidence intervals of the four model parameters are shown in Table 30 and Table 31.

Method	$\boldsymbol{\alpha_1}$	$\boldsymbol{\lambda_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\lambda_2}$
MLE	.001625	0.216	0.000598	0.231
BE	1.071e-03	0.2258	0.013370	0.1655

Table 31: Estimates the parameters by maximum likelihood and Bayesian methods.

	Confidence intervals	
Parameters	Lower bound	Upper bound
α_1	0.0010036	0.0022463
λ_1	0.21554682	0.2165431
α_2	0.0003095	0.00088669
λ_2	0.23084136	0.2311586

Table 32: The asymptotic confidence intervals at 95% using maximum likelihood method.

In Bayes Method, we applied the Markov Chain Mont Carlo (MCMC) technique. We suppose all unknown parameters in this model to be independent random variables and follow gamma distribution when all the hyper-parameters are known and equal to 0.001, reflecting non-informative prior. We assumed a small value for all hyper-parameter because of the ”low information” prior that is reasonable since its mean equals one while the variance is 1000.We used the normal approximation method to estimate the unknown parameters.

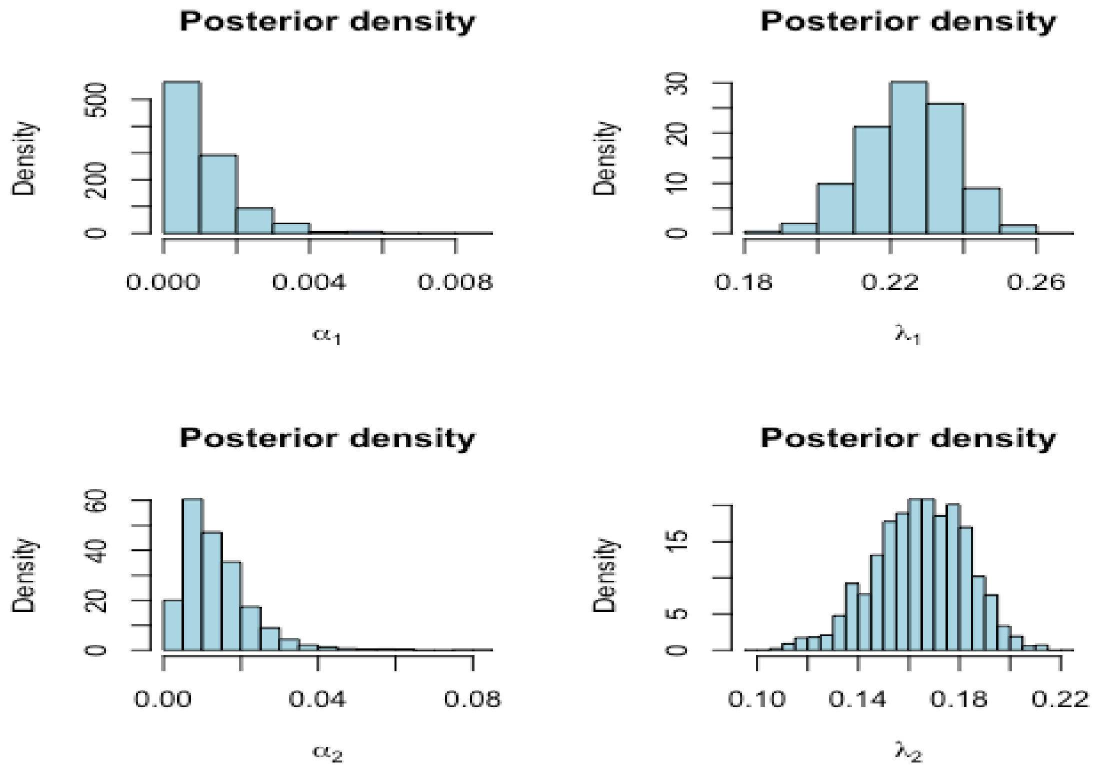
To apply MCMC, the proposal density is a multivariate normal distribution with mean vector zero and the variance-covariance matrix, is obtained from the normal approximation method and a positive value for scale parameter as 0.7. We applied Metropolis random walk algorithm to simulate 10000 samples of draws from the log posterior density distribution. The acceptance rate of the sample draws was equal to 35.22%. Furthermore, after discarding the first 50% of the draws which have from random walk chain method, we can provide more inference on the four unknown parameters by computing Bayes point estimates and the 95% credible intervals of the fours model parameters which shown in Table 30 and Table 31. Hence, we can use the link between the measures of central tendency in Table 33 and the histogram shapes in Figure 17. The measures of central tendency for α_1 and α_2 both of them have means larger than the medians because of that histogram is right-skewed but the measures of central tendency for λ_1 and λ_2 both have means smaller than their medians, because that histogram is left-skewed.

	Probability intervals	
Parameters	Lower bound	Upper bound
α_1	0.0001705914	0.0034735502
λ_1	0.2013185	0.2481062
α_2	0.003107298	0.036062242
λ_2	0.1265825	0.1986200

Table 33: The 95% credible intervals at for the four parameters.

Measures	α_1	λ_1	α_2	λ_2
Min.	2.364e-05	0.1854	0.001278	0.1078
First Quartile	4.846e-04	0.2177	0.007046	0.1539
Median	8.196e-04	0.2261	0.011070	0.1660
Mean	1.071e-03	0.2258	0.013370	0.1655
Third Quartile	1.369e-03	0.2349	0.017580	0.1779
Max.	8.501e-03	0.2663	0.077190	0.2144

Table 34: The measures of central tendency of all model parameters.



Posterior density

Posterior density

Posterior density

Posterior density

Figure 19: Estimated marginal posterior distributions of the model parameters.

Furthermore, all diagnostics test results on the simulations draws, which is instructive to determine if they approximately represent the posterior distribution of interest. Figure 18, the trace plots, shows the simulated draws of $\theta_i = \log(\alpha_1), \log(\lambda_1), \log(\alpha_2), \log(\lambda_2), i = 1, \dots 4$, for an MCMC chain with poor choices of start values and scale factor. We plot the same graph in Figure 19, but after removing the first 5000 draws due to the poor choice of start values and to show better combination of simulated draws than before. Hence, we provided autocorrelation plots of simulated draws of $\alpha_1, \lambda_1, \alpha_2$ and $\lambda_2$for the random walk chain after discarding the first 50% of the draws in Figure 20. Additionally, the autocorrelation plots show the lag decreases to zero very quickly, which means the draws become independent over time. The estimates of the relative risk rates of each cause of failure in the presence of all other causes are shown in Table 34. The trace plots of the draws of $\pi_j, j = 1,2$, and the corresponding marginal posterior density functions are shown in Figure 21.

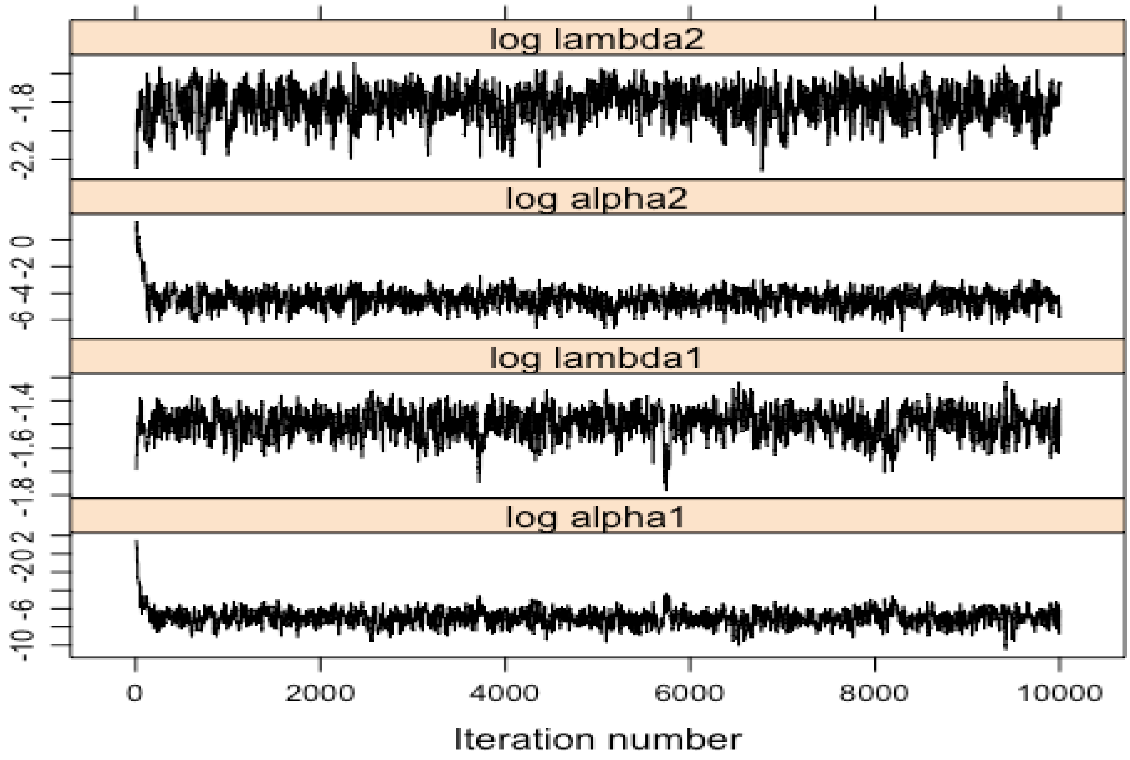


Figure 20: The trace plots of the10000 simulated draws of the transformed parameters.

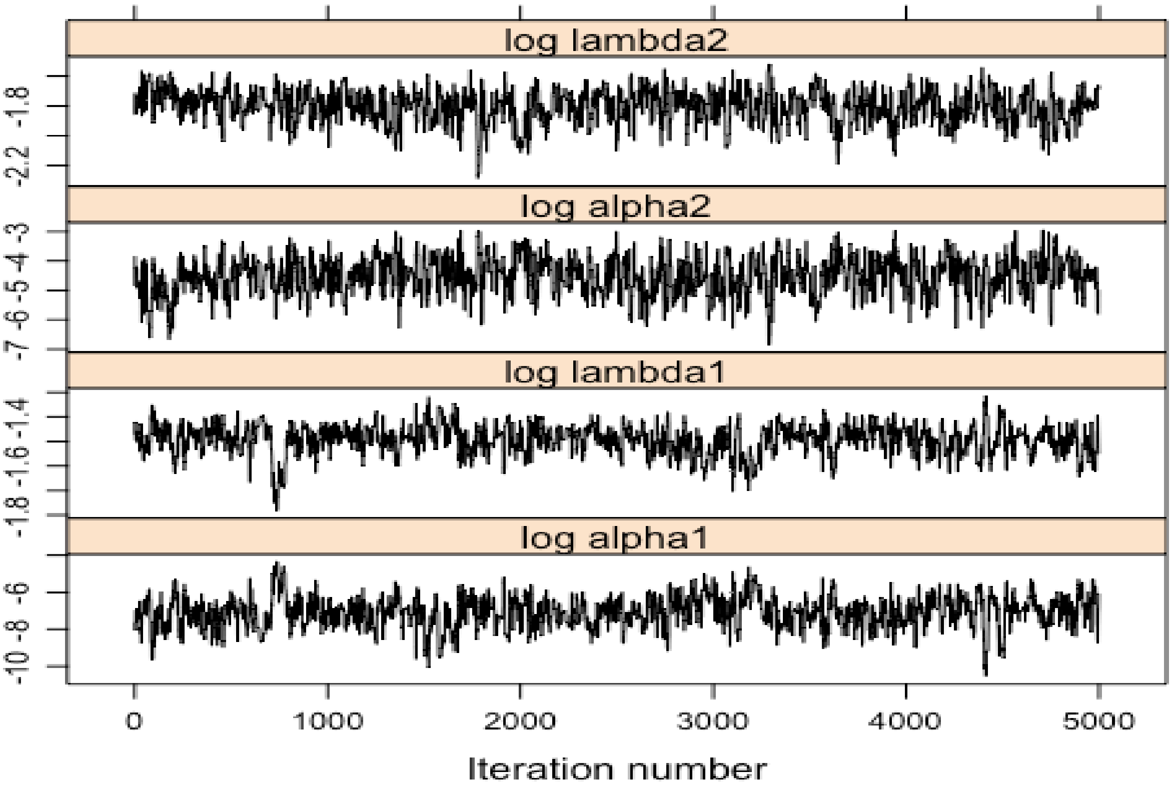


Figure 21: The simulated draws after discarding the early 50% of the draws.

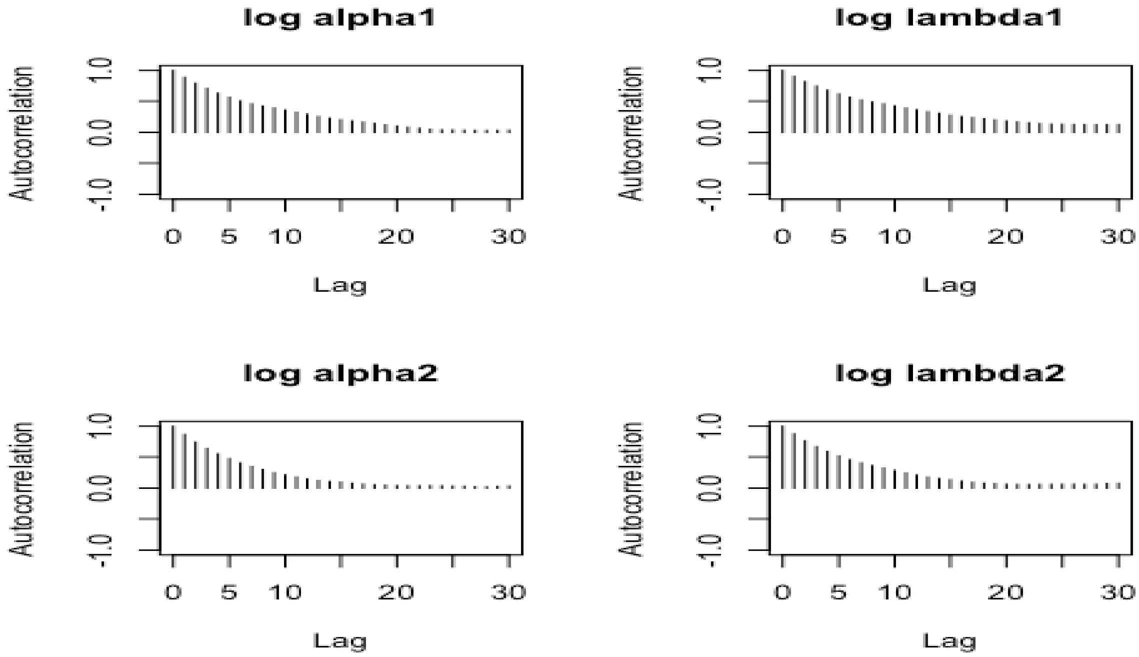


Figure 22: The autocorrelation plots of simulated draws after discarding the early 50% of the draws.

Methods	$\hat{\pi}_1$	$\hat{\pi}_2$
MLE	0.6921	0.3078
BE	0.5378	0.4621

Table 35: The estimation of the relative risk for the two causes.

Measures	$\pi_1(t)$	$\pi_2(t)$
Min.	0.1796	0.3990
First Quartile	0.3407	0.5633
Median	0.3889	0.6111
Mean	0.3888	0.6112
Third Quartile	0.4367	0.6593
Max.	0.6010	0.8204

Table 36: The measures of central tendency for each risk.

	Probability intervals	
The risks	Lower bound	Upper bound
$\hat{\pi}_1$	0.2494453	0.5239376
$\hat{\pi}_2$	0.4760624	0.7505547

Table 37: The 95% credible intervals for each risk.

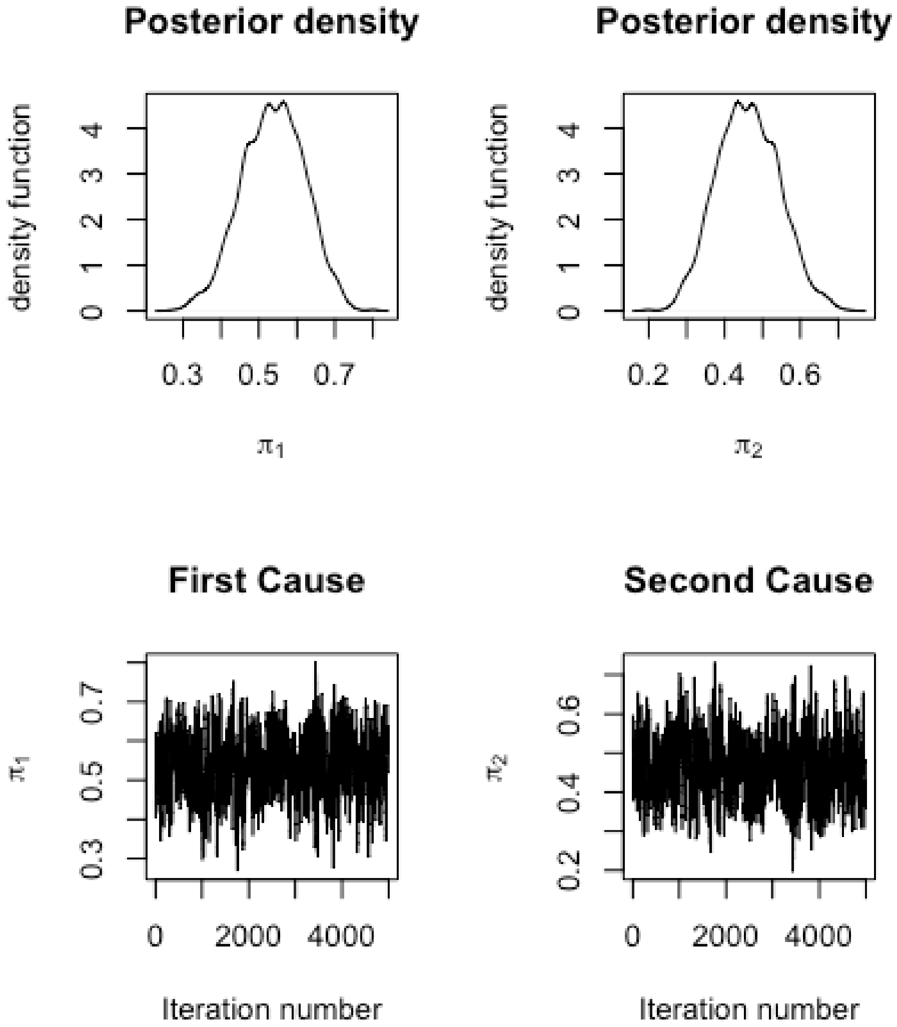


Figure 23:The posterior density and trace plots for each risk $\boldsymbol{\pi_j(t)}$,j=1,2.

Based on the output from the MCMC method, we selected the last 50% of the simulated draws, which come from the posterior density for each parameter. As shown in Table 38 and Table 41, we estimated hazard and survival functions by using the maximum likelihood and MCMC methods and computed the measures of central tendency shown in Table 39 and Table 42. The 95% credible intervals for the hazard and survival functions are shown in Table 40 and Table 43.

Hence, we can use the measures of central tendency in Table 39 and Table 42 to describe the histogram shapes in Figure 24 and Figure 25. The measures of central tendency for hazard functions all have means larger than their medians because that the histogram shape is right-skewed but survival functions all have means smaller than their medians because that histogram shape is left-skewed.

Methods	$h_1(t_0)$	$h_2(t_0)$
MLE	0.0004599476	0.0004796583
BE	0.02144639	0.0219011

Table 38: The estimation hazard function by maximum likelihood and MCMC methods.

Measures	$h_1(t)$	$h_2(t)$
Min.	0.04553	0.03794
First Quartile	0.05991	0.05230
Median	0.06395	0.05635
Mean	0.06402	0.05643
Third Quartile	0.06807	0.06022
Max.	0.08553	0.07614

Table 39: The measures of central tendency for hazard functions evaluated at the mean of the available data

	Probability intervals	
Hazard Functions	Lower bound	Upper bound
$h_1(t)$	0.05257848	0.07694172
$h_2(t)$	0.04542480	0.06868057

Table 40: The 95% credible intervals for hazard functions evaluated at the mean of the available data.

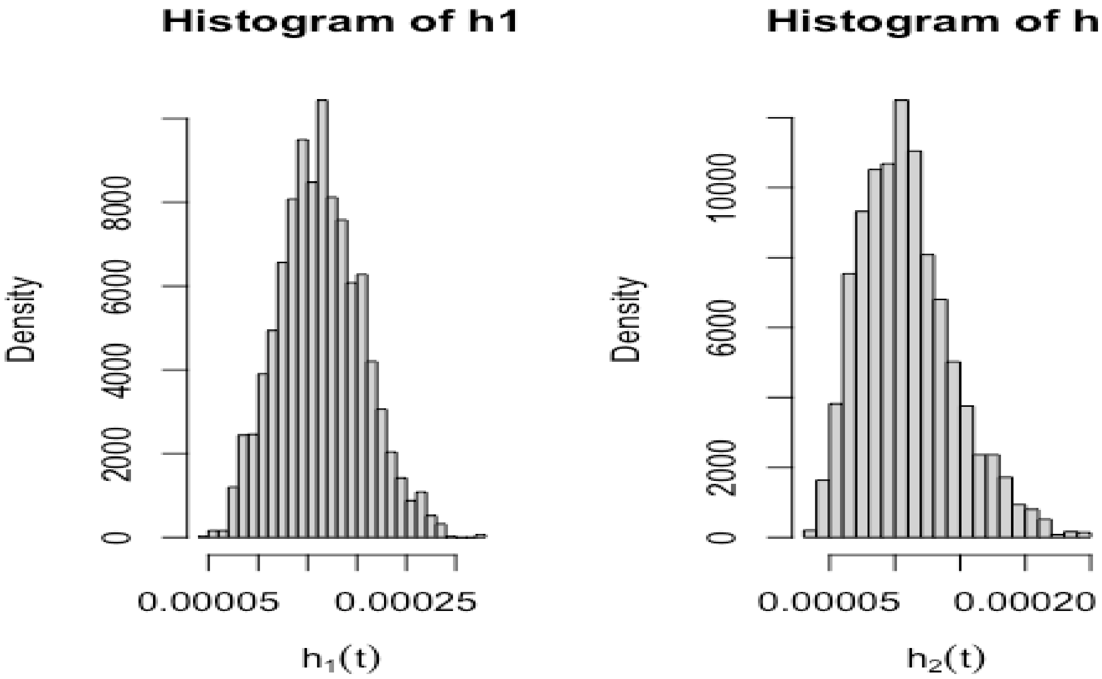


Figure 24: The posterior density for the hazard function of both two risks evaluated at the mean of data.

Methods	$S_1(t)$	$S_2(t)$
MLE	0.5800188	0.6121732
BE	0.7095	0.6274

Table 41: The estimation survival function, evaluated at the mean of the observations, by maximum likelihood and MCMC methods.

Measures	$S_1(t)$	$S_2(t)$
Min.	0.4422	0.3431
First Quartile	0.6650	0.5796
Median	0.7142	0.6303
Mean	0.7095	0.6274
Third Quartile	0.7612	0.6760
Max.	0.8681	0.8288

Table 42: The measures of the central tendency for the survival function of both two risks.

	Probability intervals	
Hazard Functions	Lower bound	Upper bound
$S_1(t)$	0.5599332	0.8183467
$S_2(t)$	0.4831888	0.7598601

Table 43: The 95% credible intervals for the survival function of both two risks.

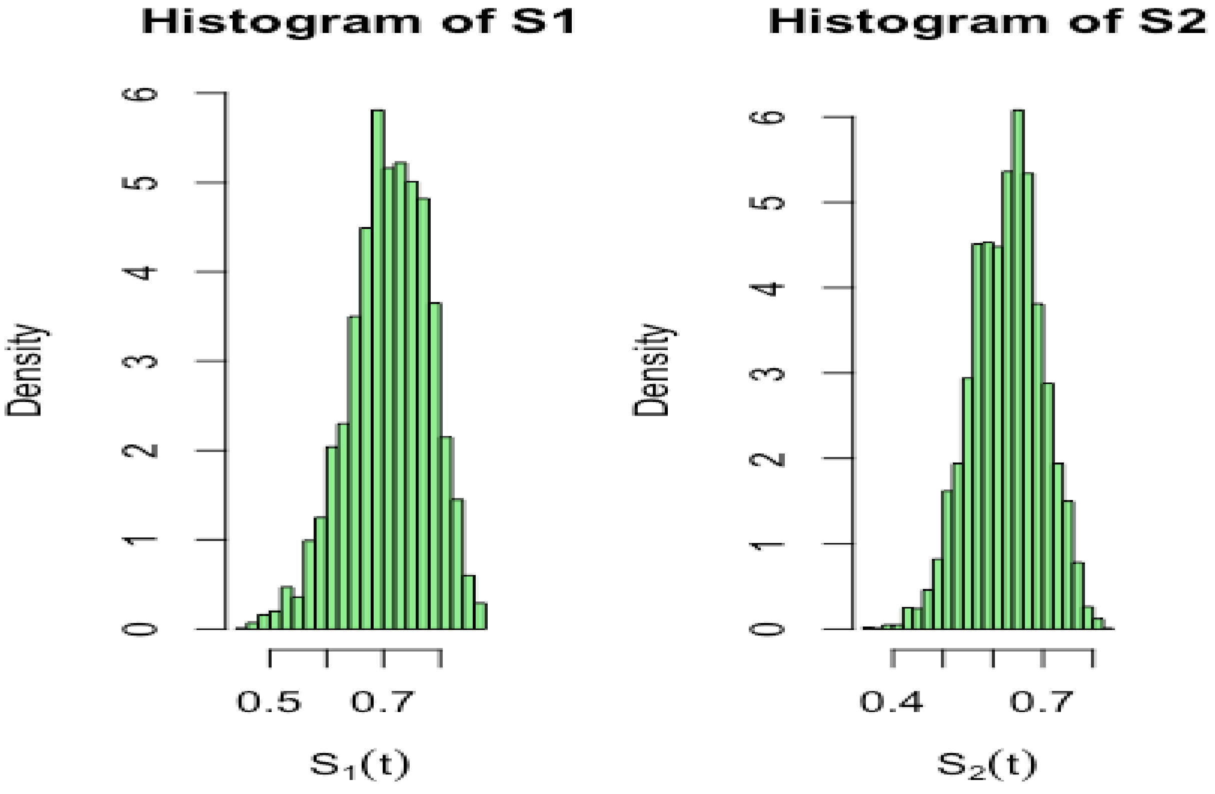


Figure 25:The posterior density for both two survival function, evaluated at the mean of the observations, of the risks.

4.4 Discussion

Putter et al. (2007) analysed the data from the Amsterdam Cohort Studies on HIV infection and AIDS using non-parametric setups. The relative risk rates of each cause of failure in the presence of all other risks are estimated by using the Kaplan–Meier estimate. Additionally, the regression approaches was used to estimate the effect of CCR5 with age at HIV infection on the cause-specific hazard function and the cumulative incidence function.

In this study, we analyzed the AIDSSI data set using parametric setups. The objectives of this study is to estimate the unknown parameters included in competing risks models without and with covariates when the risks follow TPBT distribution with different parameters and the relative risk rates of each cause of failure in the presence of all other causes using MLE and MCMC methods. Also, we estimated some of the reliability measures of the competing risks system.

In case study 1, Table 6, shows the maximum likelihood and Bayes point estimates of the four model parameters Table 7 and Table 8 show asymptotic 95% confidence intervals and credible intervals of the four unknown parameters. Table 10, shows the estimated risks using both two methods. The results from both methods are similar. As a diagnostic test for the MCMC, as shown in Figure 4,5,6, the trace plots show good mix of the sampled draws and the autocorrelation plots show that the lag decreases, which indicates that the draws become approximately independent over time.

In case study 2, Table 19, shows the maximum likelihood and Bayes point estimates of the eight model parameters. Table 20 and Table 21 show asymptotic 95% confidence intervals and credible intervals of the eight unknown parameters. Table 23 and Table 24; show the estimated risks using both two methods. The results from both methods are different because of the effect of CCR5 and age at HIV infection. As a diagnostic test for the MCMC, as shown in Figure 12,13,14 the trace plots show good mix of the sampled draws and the autocorrelation plots show that the lag decreases, which indicates that the draws become approximately independent over time.

Chapter 5: Conclusions

Recently more lifetime distributions have been used to analyse competing risks data. In this study we used the two-parameter bathtub (TPBT) distribution because the hazard rate can be either increasing or a bathtub-shaped, which allows it to be a good fit for several data sets. In competing risks data, it is assumed that the subject is under attack of many risks that compete to destroy it, but only one risk can occur, and all risks are independent. Also, there is a chance that the subject may not received any attack during the study period and in this case, we observe the length of such period (censored time, there is no failure).

The maximum likelihood and Bayes methods were used to estimate the parameters of all risks and some of reliability measures of the system. There was no analytic solution for the likelihood equations of the unknown parameters in case of both the competing risks model and competing risks regression model, therefore numerical methods were used to get the maximum likelihood estimations. In Bayes methods, the Markov Chain Monte Carlo (MCMC) method was applied to obtain the Bayes estimates of the parameters and system's reliability measures, because the posterior distribution of the unknown parameters was not in a convenient form. In this thesis we considered two different real dataset. The first dataset on HIV infection and AIDS from Amsterdam Cohort Studies for men who have sex with men (Geskus et al.; 2000, 2003). The second dataset contains 36 small electrical appliances (Lawless, 2003).

5.1 Future work

Recently bladder cancer has been studied widely due to prevalence new cases and different causes of failure leading to death (Dyrskjøt et al., 2007; Hecker et al., 2013). The competing risks data with high-dimensional covariates in bladder cancer will be analyzed using different regression approaches, such as the cause-specific hazards model, sub-distribution hazards model, and mixture model. Parametric setups will be assumed to study this problem under independent and dependent assumptions. The purpose of analyzing phenotypic data is to determine which a subset of genes, has more significant correlation with time-to-event response (Engler & Li, 2009).

Bibliography

Albert, J. (2009). *Bayesian computation with R*. Springer Science & Business Media.

Ashby, D., & Hutton, J. L. (1996). Bayesian epidemiology. *STATISTICS TEXTBOOKS AND MONOGRAPHS, 151*, 109-140.

Ashby, D., & Smith, A. F. (2000). Evidence-based medicine as Bayesian decision-making. *Statistics in medicine, 19*(23), 3291-3305.

Bakoyannis, G., & Touloumi, G. (2012). Practical methods for competing risks data: A review. *Statistical methods in medical research, 21*(3), 257-272.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer Science & Business Media.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Berger, J. O. and Wolpert, R. (1988), *The Likelihood Principle*, 9, Second Edition, Hayward, California: Institute of Mathematical Statistics, monograph series.

Berkson, J., & Elveback, L. (1960). Competing exponential risks, with particular reference to the study of smoking and lung cancer. *Journal of the American Statistical Association, 55*(291), 415-428.

Bernardo, J. M. and Smith, A. F. M. (1994), *Bayesian Theory*, New York: John Wiley & Sons.

Bernardo, J. M., & Smith, A. F. (2001). Bayesian theory.

Bernoulli D. Essai d'une nouvelle analyse de la mortalit´e caus´ee par la petite V´erole, et des avantages Berrington de Gonzalez, A., Hartge, P., Cerhan, J. R., Flint, A. J., Hannan, L., MacInnis, R. J., ... & Beeson, W. L. (2010). Body-mass index and mortality among 1.46 million white adults. New England Journal of Medicine, 363(23), 2211-2219.

Bertozzi S, Padian NS, Wegbreit J, et al. HIV/AIDS Prevention and Treatment. In: Jamison DT, Breman JG, Measham AR, et al., editors. Disease Control Priorities in Developing Countries. 2nd edition.

Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2006. Chapter 18. Available from: https://www.ncbi.nlm.nih.gov/books/NBK11782/Co-published by Oxford University Press, New York.

Beyersmann J, Schumacher M, Allignol A (2012) Competing Risks and Multistate Models.

Beyersmann, J., Allignol, A., & Schumacher, M. (2011). *Competing risks and multistate models with R.* Springer Science & Business Media.

Bocchetti, D., Giorgio, M., Guida, M., & Pulcini, G. (2009). A competing risk model for the reliability of cylinder liners in marine Diesel engines. *Reliability Engineering & System Safety, 94*(8), 1299-1307.

Box, G. E., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40). John Wiley & Sons.

Breslow, N. (1990). Biostatistics and bayes. Statistical Science, 269-284.

Zhou, C., Raymond, H. F., Ding, X., Lu, R., Xu, J., Wu, G., ... & Xiao, Y. (2013). Anal sex role, circumcision status, and HIV infection among men who have sex with men in Chongqing, China. *Archives of sexual behavior, 42*(7), 1275-1283.

Carlin, B. P. and Louis, T. A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, Second Edition, London: Chapman & Hall.

Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters, 49*(2), 155-161.

Clard J. S. & Gelfand A. E. (2006), Hierarchical Modelling for The Environmental.

Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society. Series B (Methodological),* 411-421.

Crowder, M. J. (2001). *Classical competing risks.* CRC Press.

David, H. A., & Moeschberger, M. L. (1978). *The Theory of Competing Risks: HA David, ML Moeschberger.* C. Griffin.

Deeks, S. G., & Phillips, A. N. (2009). HIV infection, antiretroviral treatment, ageing, and non-AIDS related morbidity. *Bmj, 338*, a3172.

Dignam, J. J., Zhang, Q., & Kocherginsky, M. N. (2012). The use and interpretation of competing risks regression models. *Clinical Cancer Research, clincanres*-2097.

Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics*, 417-431.

Dyrskjøt, L., Zieger, K., Real, F. X., Malats, N., Carrato, A., Hurst, C., ... & Wester, K. (2007). Gene expression signatures predict outcome in non–muscle-invasive bladder carcinoma: a multicenter validation study. *Clinical Cancer Research, 13*(12), 3545-3551.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling, 8*(3), 430-457.

Engler, D., & Li, Y. (2009). Survival analysis with high-dimensional covariates: an application in microarray studies. Statistical applications in genetics and molecular biology, 8(1), 1-22.

Flegal, K. M., Kit, B. K., Orpana, H., & Graubard, B. I. (2013). Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. Jama, 309(1), 71-82.

Geisser, S., Hodges, J., Press, S., & ZeUner, A. (1990). The validity of posterior expansions based on Laplace's method. *Bayesian and likelihood methods in statistics and econometrics, 7,* 473.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). Bayesian data analysis (Vol. 2). Boca Raton, FL: CRC press.

Geskus, R. B. (2000). On the inclusion of prevalent cases in HIV/AIDS natural history studies through a marker- based estimate of time since seroconversion. *Statistics in medicine, 19*(13), 1753-1769.

Gichangi, A., & Vach, W. (2005). The analysis of competing risks data: A guided tour. *Statistics in Medicine, 132*(4), 1-41.

Glickman, M. E., & Van Dyk, D. A. (2007). Basic bayesian methods. *Topics in Biostatistics,* 319-338.

Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. The Annals of statistics, 1141-1154.

Gupta, S., Hassan, S., Bhatt, V. R., Sater, H. A., & Dilawari, A. (2014). Lung cancer trends: smoking, obesity, and sex assessed in the Staten Island University's lung cancer patients. *International journal of general medicine, 7,* 333.

Haller, B., Schmidt, G., & Ulm, K. (2013). Applying competing risks regression models: an overview. *Lifetime data analysis*, 1-26.

Haque, R., Van Den Eeden, S. K., Wallner, L. P., Richert-Boe, K., Kallakury, B., Wang, R., & Weinmann, S. (2014). Association of body mass index and prostate cancer mortality. *Obesity research & clinical practice, 8*(4), e374-e381.

Hecker, N., Stephan, C., Mollenkopf, H. J., Jung, K., Preissner, R., & Meyer, H. A. (2013). A new algorithm for integrated analysis of miRNA-mRNA interactions based on individual classification reveals insights into bladder cancer. *PLoS One, 8*(5), e64543.

Kalbfleisch, J. D., & Prentice, R. L. (2002). Competing risks and multistate models. *The Statistical Analysis of Failure Time Data, Second Edition,* 247-277.

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data* (Vol. 360). John Wiley & Sons.

Buchacz, K., Armon, C., Palella, F. J., Baker, R. K., Tedaldi, E., Durham, M. D., & Brooks, J. T. (2012). CD4 cell counts at HIV diagnosis among HIV outpatient study participants, 2000–2009. *AIDS research and treatment,* 2012.

Klein, J. P. (2006). Modelling competing risks in cancer studies. *Statistics in medicine, 25*(6), 1015-1034.

Klein, J. P., & Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data.* Springer Science & Business Media.

Kundu, D., & Basu, S. (2000). Analysis of incomplete data in presence of competing risks. *Journal of Statistical Planning and Inference, 87*(2), 221-239.

Kundu, D., & Sarhan, A. M. (2006). Analysis of incomplete data in presence of competing risks among several groups. *IEEE Transactions on Reliability, 55*(2), 262-269.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., & Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature, 393*(6686), 648.

Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (Vol. 362). John Wiley & Sons.

Lim, H. J., Zhang, X., Dyck, R., & Osgood, N. (2010). Methods of competing risks analysis of end-stage renal disease and mortality among people with diabetes. *BMC medical research methodology, 10*(1), 97.

Marubini, E., & Valsecchi, M. G. (2004). *Analysing survival data from clinical trials and observational studies* (Vol. 15). John Wiley & Sons.

Mayer, K. H., Wheeler, D. P., Bekker, L. G., Grinsztejn, B., Remien, R. H., Sandfort, T. G., & Beyrer, C. (2013). Overcoming biological, behavioral and structural vulnerabilities: New directions in research to decrease HIV transmission in men who have sex with men. *Journal of acquired immune deficiency syndromes (1999), 63*(0 2), S161.

Miyakawa, M. (1982). Statistical analysis of incomplete data in competing risks model. *Journal of Japanese Soc. Quality Control, 12,* 49-52.

Newcomb ME, Ryan DT, Garofalo R, Mustanski B. The Effects of Sexual Partnerships and Relationship Characteristics on Three Sexual Risk Variables in Young Men Who Have Sex with Men. *Archives of sexual behavior*. 2014;43(1):61-72. doi:10.1007/s10508-013-0207-9.

Noordzij, M., Leffondré, K., van Stralen, K. J., Zoccali, C., Dekker, F. W., & Jager, K. J. (2013). When do we need competing risks methods for survival analysis in nephrology? *Nephrology Dialysis Transplantation, 28*(11), 2670-2677.

O'Hagan, A., & Forster, J. J. (2004). *Kendall's advanced theory of statistics, volume 2B: Bayesian inference* (Vol. 2). Arnold.

Park, C. (2005). Parameter estimation of incomplete data in competing risks using the EM algorithm. *IEEE Transactions on Reliability*, *54*(2), 282-290.

Pepe, M. S. (1991). Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association, 86*(415), 770-778.

Pepe, M. S., & Mori, M. (1993). Kaplan—meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in medicine*, *12*(8), 737-751.

Pintilie, M. (2006). *Competing risks: a practical perspective* (Vol. 58). John Wiley & Sons.

Pintilie, M. (2011). An introduction to competing risks analysis. *Revista Española de Cardiología (English Edition), 64(*7), 599-605.

Press, S. J. (2009*). Subjective and objective Bayesian statistics: principles, models, and applications* (Vol. 590). John Wiley & Sons.

Robert, C. P. (2001), *The Bayesian Choice*, Second Edition, New York: Springer-Verlag.

Sarhan, A. M. (2007). Analysis of incomplete, censored data in competing risks models with generalized exponential distributions. *IEEE Transactions on Reliability, 56*(1), 132-138.

Sarhan, A. M., Hamilton, D. C., & Smith, B. (2010). Statistical analysis of competing risks models. *Reliability Engineering & System Safety, 95*(9), 953-962.

Simon, V., Ho, D. D., & Karim, Q. A. (2006). HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. *The Lancet, 368*(9534), 489-504.

Wasserman, L. (2004). All of statistics: A concise course in statistical inference brief contents. *Simulation, 100*, 461.

Wu, S., Liu, J., Wang, X., Li, M., Gan, Y., & Tang, Y. (2014). Association of obesity and overweight with overall survival in colorectal cancer patients: a meta-analysis of 29 studies. *Cancer Causes & Control, 25*(11), 1489-1502.