PARAMETRIC AND NON-PARAMETRIC ANALYSIS OF COMPETING RISKS MODELS

by

Vishal Sood

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

at

Dalhousie University
Halifax, Nova Scotia
April 2020

Table of Contents

List of Tables **iv**

List of Figures **vi**

Abstract **viii**

List of Abbreviations and Symbol Used **ix**

Acknowledgements **xi**

1 Chapter 1: Introduction **1**

1.1 Background and Motivation 1

1.2 Quantities for Modelling Lifetime Data 3

1.3 Some Useful Lifetime Distributions 6

1.4 Maximum Likelihood Estimation 8

1.5 Bayesian Estimation 10

1.6 Markov Chain Monte Carlo 12

1.7 Outline of Thesis 13

2 Chapter 2: Model Assumptions and Notation **15**

3 Chapter 3: Non-Parametric Estimation **21**

3.1 Kaplan-Meier Estimator 21

3.2 Cumulative Incidence Function 23

4 Chapter 4: Homogeneous Competing Risks Model **24**

4.1 Maximum Likelihood Method 24

4.1.1 Maximum Likelihood for STH-2 Case 26

4.2 Bayesian Estimation 29

4.2.1 Bayesian Estimation for STH-2 Case 31

5 Chapter 5: Competing Risks Regression Model **32**

5.1 Proportional Hazards Model 32

5.2 Estimation using Cox's Cause-Specific Hazard Model 34

5.3 Estimation using Maximum Likelihood 35

5.4 Estimation using Bayesian Methods 36

6 Chapter 6: Data Analysis **39**

6.1 Bone Marrow Transplant Data 39

6.2 Non-Parametric Estimation 41

6.2.1 Kaplan-Meier Estimator 41

6.2.2 Cumulative Incidence Function 44

6.2.3 Comparison of CIF Among the Groups 48

6.3 Parametric Estimation 49

6.3.1 Model Selection using the Behaviour of the Hazard 49

6.3.2 Maximum Likelihood Estimates 52

6.3.3 Bayesian Estimation 60

6.4 Competing Risks with Covariates 72

6.4.1 Cox Regression 72

6.4.2 Maximum Likelihood and Bayesian Estimation 75

7 Chapter 7: Conclusion **82**

Bibliography **85**

List of Tables

1.1 The hazard rate, survival function, and probability density function for each distribution. 7
1.2 The mean of the lifetime distributions. 8

4.1 The hazard rate and cumulative hazard rate of risk j for each distribution. 28

6.1 The survival times of the bone marrow patients. 40
6.2 The number of events for each status by their respective group. . . . 40
6.3 Mean follow-up in days for each status by their respective group. . . . 41
6.4 Gray's Test for testing equality among groups. 48
6.5 The model ranks for each group, where the parameters of the models were estimated using maximum likelihood estimation. 52
6.6 The MLEs estimate for cause 1 and cause 2 for each group. The value of log-likelihood evaluated at the MLEs is given, as well as the AIC and BIC values. 53
6.7 The asymptotic 95% CI for the true model parameters. 54
6.8 The expected failure time in days. 58
6.9 The estimated relative risk for each distribution. 59
6.10 The Kolmogorov-Smirnov (KS) statistics are shown for each risk among groups. Bold values denote the smallest value. 60
6.11 The acceptance rate for the Metropolis-Hastings algorithm using the good prior for all risks model used. 64
6.12 The Bayes estimate and respective variance for each model using a poor choice of hyper-parameters for the prior distribution (all hyper-parameters are set to 0.001). 65
6.13 The Bayes estimate and respective variance for each model using the hyper-parameters from Table 6.14 for the prior distribution (good prior). 66

6.14 The hyperparameters used in the good prior distribution to calculate the Bayes' estimate in Table 6.13 67
6.15 The 95% credible intervals for the parameters. 68
6.16 A description of the covariates (risk factors) for the patients. 73
6.17 The results of Cox's cause-specific hazard model. 74
6.18 The MLEs and Bayes estimates for the parametric regression model where the baseline cause-specific hazard is assumed to follow a Power Lindley distribution. 76
6.19 95% asymptotic confidence intervals for the true parameters and coefficients of the regression model, where the risks are assumed to follow the Power Lindley distribution. 78

List of Figures

1.1 A lifetime of n patients with some being right censored. 1
1.2 Different shapes of the hazard function. 5

2.1 Competing risks model with k fatal causes of failure. 16
2.2 The object can also be interpreted as a series system with k independent but not identical components. T_j can be interpreted as the lifetime of the j^{th} component. 16

6.1 A bone marrow transplant patient experiencing two competing risks. 41
6.2 The 1−KM curves for relapse (top) and death in remission (bottom). 43
6.3 The CIF for patients with ALL (group 1). 45
6.4 The CIF for patients with AML-LR (group 2). 45
6.5 The CIF for patients with AML-HR (group 3). 46
6.6 A comparison between 1−KM and CIF for the three groups, where the top figure is for relapse and the bottom figure is for remission. 47
6.7 The total time on test (TTT) for cause 1 and cause 2 of all three groups. 51
6.8 The CIF (using MLE) for all models (top) and the top three models for ALL. 55
6.9 The CIF (using MLE) for all models (top) and the top three models for AML-LR. 56
6.10 The CIF (using MLE) for all models (top) and the top three models for AML-HR. 57
6.11 The trace plots for the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1). 61
6.12 The autocorrelation plots of the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1). 62

6.13 The marginal posterior densities for α_1 (top left), β_1 (top right), α_2 (bottom left), and β_2 (bottom right), for Power Lindley (group 1). . . 63
6.14 A comparison of the estimated CIF for group 1 (ALL patients) using Bayes', MLE and empirical methods, where the risks follow the Power Lindley distribution. 69
6.15 A comparison of the estimated CIF for group 2 (AML-LR patients) using Bayes', MLE and empirical methods, where the risks follow the Power Lindley distribution. 70
6.16 A comparison of the estimated CIF for group 3 (AML-HR patients) using Bayes', MLE and empirical methods, where the risks follow the Power Lindley distribution. 71
6.17 The trace plots for the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1). 79
6.18 The ACF plots for the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1), with regression. 80
6.19 The marginal densities for the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1), with regression. . . . 81

Abstract

Competing risks refers to the phenomenon where an object or individual is subject to multiple risks that are competing to destroy the object or individual, where the occurrence of one of these risks precludes the occurrence of any other risk. In this thesis, we use some useful lifetime distributions to model the risks, and parametric methods to estimate the unknown parameters of these risk models. Maximum likelihood method is used to estimate the model parameters. As expected, there were no analytic solutions for the maximum likelihood estimators, therefore numerical methods are implemented. Bayesian method is also used to estimate the model parameters. Since the posterior probability density function of the vector of unknown parameters is not in a standard form of a known distribution, MCMC using the Metropolis-Hastings algorithm is performed to obtain the Bayes estimates. Non-parametric techniques are also used to estimate the main characteristics of the competing risks model. Two competing risks models are studied in this thesis: homogeneous and regression models. Data analysis is done on bone marrow transplant patients in which there were two risks: leukemia relapse and death in remission. The cumulative incidence function estimates the probability of a specific risk in the presence of all other risks. We also estimate the cumulative incidence function for every risk at different times using the parametric and non-parametric methods applied in this thesis. Testing on the significance of covariates, patient's age and donor's age, found that at least one of them were significant for patients with acute lymphoblastic leukemia (ALL) and acute myelocytic leukemia low-risk (AML-LR), but not for patients with acute myelocytic leukemia high-risk (AML-HR). A comparison between the homogenous and regression competing risks models, using the bone marrow data, is performed. Further investigation is needed on modelling the risks where each risk is assumed to follow different lifetime distributions.

List of Abbreviations and Symbols Used

- α - Parameter of the lifetime distributions
- β - Parameter of the lifetime distributions
- δ - Indicator variable
- $F_j(\cdot)$ - Cumulative incidence function of risk j
- $f(\cdot)$ - Probability density function
- $f_j(\cdot)$ - Cause-specific probability density function
- $G(\cdot)$ - Cumulative distribution function
- γ - Coefficient of a covariate
- $h(\cdot)$ - Hazard rate function
- $h_j(\cdot)$ - Cause-specific hazard rate function
- $L(\cdot)$ - Likelihood function
- $\mathcal{L}(\cdot)$ - Log-Likelihood function
- $\boldsymbol{\theta}$ - Vector of unknown parameters
- $\hat{\boldsymbol{\theta}}$ - Vector of estimated parameters
- k - Number of competing risks
- n - Number of objects/individuals
- $S(\cdot)$ - Survival function
- $S_j(\cdot)$ - Cause-specific survival function
- T_j - the time of which risk j occurs to the object
- T_{ij} - the lifetime of risk j of object i
- X - the lifetime of an object
- X_i - the lifetime of object i
- AIC - Akaike Information Criterion
- ALL - Acute Lymphoblastic Leukemia

- AML-LR - Acute Myelocytic Leukemia Low Risk
- AML-HR - Acute Myelocytic Leukemia High Risk
- BIC - Bayesian Information Criterion
- BMT - Bone Marrow Transplant
- BT - Bathtub Distribution
- CDF - Cumulative Distribution Function
- CIF - Cumulative Incidence Function
- CSH - Cause Specific Hazard
- Exp - Exponential Distribution
- KM - Kaplan Meier
- KS - Kolmogorov-Smirnov
- MCMC - Markov Chain Monte Carlo
- MLE - Maximum Likelihood Estimation
- MRL - Mean Residual Life
- PDF - Probability Density Function
- PL - Power Lindley Distribution
- STH-1 - One-Parameter Sarhan-Tadj-Hamilton Distribution
- STH-2 - Two-Parameter Sarhan-Tadj-Hamilton Distribution

Acknowledgements

I would like to give my sincerest gratitude to my mentor and supervisor Dr. Ammar Sarhan for all of his guidance and knowledge over the last two years, as well as providing me with his unconditional support and time. I truly can not express how fortunate it was to do my thesis under the supervision of Dr. Sarhan.

I would like to thank my partner Emilee for encouraging me and motivating me through the stressful times that came along with the program, and for the constant support she provided throughout the years. I'd like to thank my sister, Shivani, for all the advice and guidance she gave me over my academic years. Lastly, I would like to give my sincere gratitude to my parents for all the support and love provided throughout my life.

Chapter 1

Introduction

1.1 Background and Motivation

Survival analysis is a branch of statistics that analyses time-to-event data. The time-to-event may be known as a failure time, survival time, or an event time [5, p. 399]. For example, the time-to-event may be time to death, relapse of a disease, or the occurrence of a heart attack. A common attribute of time-to-event data is censored observations. Censoring prevents us from observing the time an event of interest has occurred. A common form of censoring is right censored, which implies the object is event-free until some censored time [11, p. 30]. For example, consider n patients on a life time. Some of these patients being studied may be lost for follow-up, or may have not experienced the event of interest before the researchers stop the test. As a consequence, these patients are right censored and it is unknown whether the patient has experienced the event of interest after the recorded time. In Figure 1.1, we

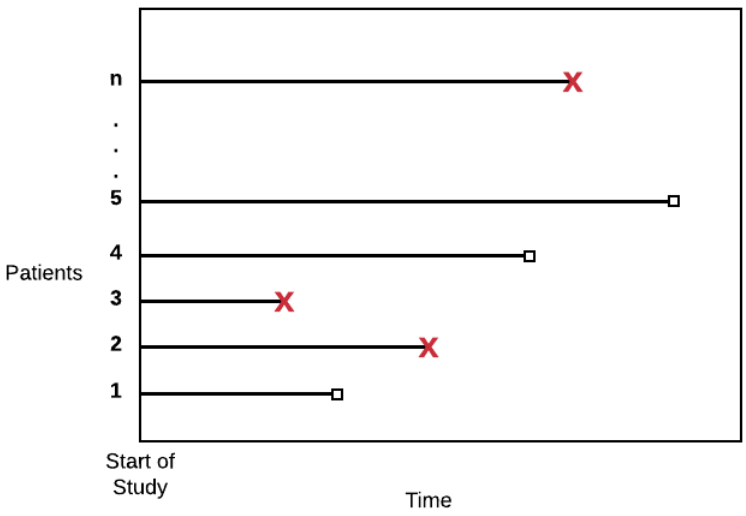


Figure 1.1: A lifetime of n patients with some being right censored.

illustrate this idea. We see that patients 2, 3, and n have experienced the event of interest, however, patients 1, 4, and 5 did not experience the event of interest for reasons listed above and were censored.

Many survival techniques have been implemented and developed to analyse situations where an object or individual may fail or die by one particular cause or risk; however, in reality, objects or individuals are subject to multiple risks that may cause fatality. Competing risks occurs when an object is subject to multiple risks that compete to destroy the object, and the occurrence of one of these risks precludes the occurrence of all other risk [11, p. 39], [6, p. 50]. Common approaches to modelling competing risks consist of looking at one specific cause of failure and treating the observations of all other causes as right-censored observations [3]. However, standard survival analysis techniques, such as the complement of the Kaplan-Meier estimator, does not provide practical probabilities in the presence of competing risks, and require specific inference when dealing with competing risks [3], [6, p. 127].

Competing risks models have many applications to the real life world. Examples branch out to the fields of engineering, economics, finance, sociology, medical sciences and so on. In engineering, one may be interested in examining the presence of multiple failures within a series system, or in medical sciences, a health professional may be interested in the chances of patients dying from cancer with the presence of heart disease acting as a competing risk. To discuss the importance of competing risks, this thesis looks at the concept of competing risks in the field of medical sciences. With the previous example, we looked at patients who may experience death by cancer or death by heart disease. In this case, the medical examiner must determine the survival probability of cancer and heart disease in the presence of one another, and the patients must use this information to determine the best choice of treatment or action for themselves. The examiner may also be interested in what factors may influence the rate of death by cancer or death by heart disease. For the examiner to answer any of these questions, a competing risks model must be used. As we mentioned previously, the typical survival analysis techniques are not suitable to answer these

questions. To fully explore the competing risks model, we present the key quantities used to model the data.

1.2 Quantities for Modelling Lifetime Data

Let the non-negative, continuous, random variable X denote the time-to-event. Here, the event may be fatality, the occurrence of cancer, the failure of a laptop, and so on. The main quantities, or reliability measures, that describe the distribution of X are the hazard rate function, the survival function and the mean residual life. The relationship between the reliability measures will be discussed in this section.

The Survival Function

The survival function, also known as the survivor function and the reliability function in engineering, is the probability of an individual or object surviving past time x, and is the complement of the cumulative distribution function [5, p. 401]. It is given by

$$S(x) = P(X > x) = \int_x^\infty f(t)\, dt = 1 - F(x), \tag{1.1}$$

where $f(x)$ is the probability density function of X, and $F(x)$ is the cumulative distribution function of X. The survival function is a non-increasing and non-negative function, that satisfies

$$\lim_{x\to 0} S(x) = 1 \text{ and } \lim_{x\to\infty} S(x) = 0.$$

The Hazard Function

The hazard function, also known as the force of mortality or the failure rate function, is the instantaneous event rate for an individual who has not yet experienced the event at time x [5, p. 404]. It is worth noting that the hazard function does not represent a density function. It is given by

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x | X > x)}{\Delta x}. \tag{1.2}$$

Using the law of conditional probability, we can rewrite this as

$$\begin{aligned} h(x) &= \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)/P(X > x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{[F(x + \Delta x) - F(x)]/\Delta x}{S(x)} \\ &= \frac{\partial F(x)/\partial x}{S(x)} \\ h(x) &= \frac{f(x)}{S(x)}. \end{aligned} \tag{1.3}$$

Thus, by knowing any one of the survival function, hazard function, cumulative hazard function, or probability density function, the other reliability measures can always be derived [5, p. 405].

If the event of interest is death, then the hazard function represents the instantaneous death rate for individuals who are still alive at time x. The hazard function is non-negative, and its shape may be constant, increasing, decreasing, bathtub, unimodal, or any other shape that may describe the rate of the phenomenon of interest. Some examples of the hazard function taking on these shapes can be seen in Figure 1.2. If the shape of the hazard function is increasing (or decreasing) for an individual or object, then this suggests that the rate of failure for an individual or object is increasing (or decreasing) as it ages. If the shape of the hazard function is bathtub, then this suggests that the rate of failure is decreasing at earlier ages, constant at middle ages, and increasing at later ages. An example of a process that would exhibit a bathtub shape would be a human at birth.

The cumulative hazard function is given as

$$H(x) = \int_0^x h(t)dt \tag{1.4}$$

$$= -\log(S(x)). \tag{1.5}$$

The cumulative hazard function is an increasing, non-negative function that satisfies

$$\lim_{x\to 0} H(x) = 0 \text{ and } \lim_{x\to\infty} H(x) = \infty.$$

Since the hazard function is not a density function, the cumulative hazard function does not represent a probability, but rather describes a measure of risk. For example, assume the event of interest is death. Then, for some time x, the larger the value of $H(x)$, the larger the risk of death by time x.

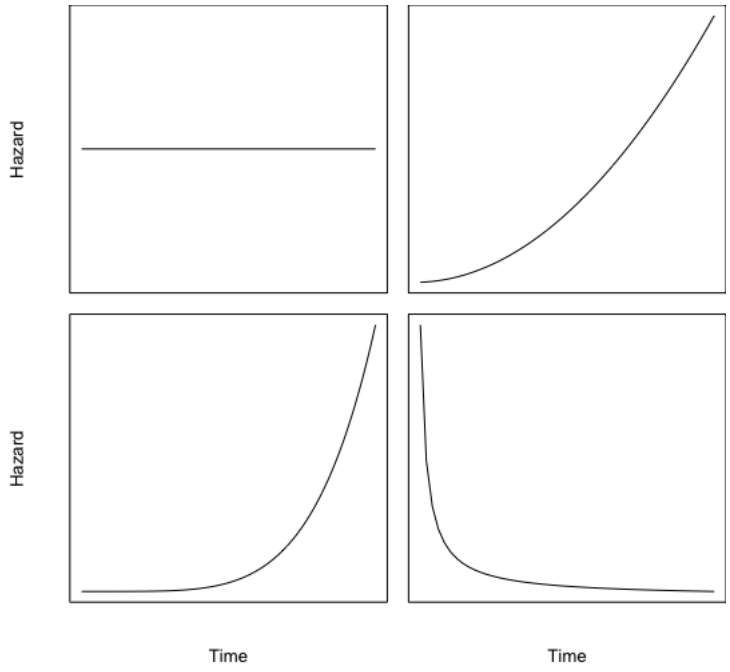


Figure 1.2: Different shapes of the hazard function.

Mean Residual Life

The mean residual life, also known as the mean restricted life [5, p. 406], at time x, is the expected remaining lifetime at time x . It is given by

$$\text{mrl}(x) = E(X - x|X > x) = \frac{\int_x^\infty (t-x)f(t)\,dt}{S(x)} = \frac{\int_x^\infty S(t)\,dt}{S(x)}. \tag{1.6}$$

The expected value of X, or expected failure time, is given by

$$\text{mrl}(0) = \mu = E(X) = \int_0^\infty t f(t)\, dt = \int_0^\infty S(t)\, dt. \tag{1.7}$$

1.3 Some Useful Lifetime Distributions

In this section, we present the properties of some useful lifetime distributions. Quantities from the previous section may be estimated non-parametrically, or using parametric assumptions. For example, the survival function may be estimated using the Kaplan Meier (Chapter 3), or it may be estimated by assuming a parametric form. In the case of the latter, the survival function can take any form of those given by Table 1.1. The hazard function and probability density function can also take on a parametric form. Each distribution, or model, have unique characteristics from one another. By assuming the lifetime data follows a suitable parametric form, then a more accurate and precise analysis may be performed. In this thesis, we examine seven different lifetime distributions that are listed in Table 1.1.

The behaviour of the hazard rate function may vary among the distributions. The exponential distribution is unique as the shape of the hazard rate function can only be constant. For the STH-1 distribution [17], the shape can be increasing, decreasing, and unimodal. For the Lindley distribution, the shape is always increasing. For the STH-2 distribution [17], the shape can be increasing, decreasing, unimodal, decreasing-increasing-decreasing (DID), and increasing-decreasing-increasing (IDI). For the Power Lindley distribution [4], the shape can be increasing, decreasing, and DID. For the Weibull distribution, the shape can be constant, increasing, or decreasing. For the Bathtub distribution, the shape can be bathtub or increasing.

Assuming the lifetime data follows a suitable parametric model, the next goal is to estimate the values of the unknown parameters of the model. The two main methods for the parameter estimation in this thesis are maximum likelihood estimation and Bayesian estimation. These techniques are presented in the next sections.

Dist.	**Hazard Rate** $h(x)$	**Survival Function** $S(x)$	**Probability Density Function** $f(x)$
Exp $x \geq 0$ $\alpha > 0$	α	$e^{-\alpha x}$	$\alpha e^{-\alpha x}$
STH-1 $x \geq 0$ $\alpha > 0$	$\frac{\alpha\left[\alpha + (1+2\alpha x)e^{-\alpha}\right]}{\alpha + (1+\alpha x)e^{-\alpha x}}$	$\frac{1}{1+\alpha}\left[\alpha + (1+\alpha x)e^{-\alpha x}\right]e^{-\alpha x}$	$\frac{\alpha}{1+\alpha}\left[\alpha + (1+2\alpha x)e^{-\alpha x}\right]e^{-\alpha x}$
Lindley $x \geq 0$ $\alpha > 0$	$\frac{\alpha^2(1+x)}{\alpha x, +\alpha + 1}$	$\frac{\alpha x + \alpha + 1}{\alpha + 1}e^{-\alpha x}$	$\frac{\alpha^2(1+x)}{\alpha + 1}e^{-\alpha x}$
STH-2 $x \geq 0$ $\alpha, \beta > 0$	$\frac{\alpha\beta\left[\beta + (1+2\beta x^{\alpha})e^{-\beta x^{\alpha}}\right]}{\beta + (1+\beta x^{\alpha})e^{-\beta x^{\alpha}}}$	$\frac{1}{1+\beta}\left[\beta + (1+\beta x^{\alpha})e^{-\beta x^{\alpha}}\right]e^{-\beta x^{\alpha}}$	$\frac{\alpha\beta x^{\alpha-1}}{1+\beta}\left[\beta + (1+2\beta x^{\alpha})e^{-\beta x^{\alpha}}\right]e^{-\beta x^{\alpha}}$
PL $x \geq 0$ $\alpha, \beta > 0$	$\alpha\beta^2\frac{(1+\alpha)x^{\alpha-1}}{\beta + 1 + \beta x^{\alpha}}$	$\left(1 + \frac{\beta}{1+\beta}x^{\alpha}\right)e^{-\beta x^{\alpha}}$	$\frac{\alpha\beta^2}{\beta+1}x^{\alpha-1}(1+x^{\alpha})e^{-\beta x^{\alpha}}$
Weibull $x \geq 0$ $\alpha, \beta > 0$	$\alpha\beta x^{\beta-1}$	$e^{-\alpha x^{\beta}}$	$\alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}$
Bathtub $x \geq 0$ $\alpha, \beta > 0$	$\alpha\beta x^{\beta-1}e^{x^{\beta}}$	$e^{\alpha(1-e^{x^{\beta}})}$	$\alpha\beta x^{\beta-1}e^{(\alpha(1-e^{x^{\beta}})+x^{\beta})}$

Table 1.1: The hazard rate, survival function, and probability density function for each distribution.

Dist.	**Mean E(X)**
Exp $x \geq 0$ $\alpha > 0$	$\frac{1}{\alpha}$
STH-1 $x \geq 0$ $\alpha > 0$	$\frac{\alpha}{1+\alpha}\left[\frac{1}{\alpha} + \left(\frac{1}{4\alpha} + \frac{1}{4\alpha^2}\right)\right]$
Lindley $x \geq 0$ $\alpha > 0$	$\frac{\alpha+2}{\alpha(\alpha+1)}$
STH-2 $x \geq 0$ $\alpha, \beta > 0$	$\Gamma\left(\frac{1}{\alpha}\right)[2\alpha(1+2^{\frac{1}{\alpha}}\beta)+1][2\alpha^2(1+\beta)(2\beta)^{\frac{1}{\alpha}}]^{-1}$
PL $x \geq 0$ $\alpha, \beta > 0$	$\frac{\Gamma\left(\frac{1}{\alpha}\right)[\alpha(\beta+1)+1]}{\alpha^2\beta^{\frac{1}{\alpha}}(\beta+1)}$
Weibull $x \geq 0$ $\alpha, \beta > 0$	$\alpha^{\frac{-1}{\beta}}\Gamma\left(1+\frac{1}{\beta}\right)$
Bathtub $x \geq 0$ $\alpha, \beta > 0$	–

Table 1.2: The mean of the lifetime distributions.

1.4 Maximum Likelihood Estimation

Maximum likelihood estimation is a technique of estimating the parameters of a probability model by maximizing the likelihood function. Assume that $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ are independent and identically distributed (i.i.d.) from a distribution $f(\boldsymbol{x}; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)$ is a vector of unknown parameters. Then, the likelihood function for $\boldsymbol{\theta}$, given $\boldsymbol{x}$, is

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}). \tag{1.8}$$

The vector $\boldsymbol{\theta}$ is said to be the maximum likelihood estimates (MLEs) if it is the set of parameters that maximizes the likelihood function (1.8) and can be denoted $\hat{\boldsymbol{\theta}}_{\text{MLE}}$. In most cases, the log-likelihood function is used to determine the MLEs as it is often

easier and computationally quicker to work with a sum of functions, rather than a product of functions. The log-likelihood function for $\boldsymbol{\theta}$, given $\boldsymbol{x}$, is

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{i=1}^{n} \log(f(x_i; \boldsymbol{\theta})). \tag{1.9}$$

It can be shown that if $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ maximizes $L(\boldsymbol{\theta}|\boldsymbol{x})$, then it also maximizes $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x})$. In most cases, to determine the MLEs, the following system of equations, known as likelihood equations, are solved:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = 0 \text{ for } i = 1, 2, ..., p. \tag{1.10}$$

Solving for θ_i in the likelihood equations will determine the MLE for the ith parameter. A problem that often arises with maximum likelihood estimation is the complexity of the likelihood equations, and results in no analytic solution being found. To overcome this problem, numerical methods to maximize the likelihood are conducted with the use of statistical software such as R.

Asymptotic Confidence Intervals

A confidence interval can be constructed to determine plausible values of the true but unknown model parameters. To construct an asymptotic confidence interval, the Fisher information matrix is needed to obtain the asymptotic standard error of each parameter. The Fisher information matrix is given as

$$\mathcal{I}_{ij}(\hat{\boldsymbol{\theta}}_{\text{MLE}}) = -E_\theta \Big[\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \Big]_{\theta = \hat{\theta}_{\text{MLE}}} \quad i, j = 1, 2, .., k. \tag{1.11}$$

Then, the estimates of the unknown parameters are asymptotically normal with mean equal to the true parameters, and variance-covariance equal to the inverse of the Fisher information matrix. That is,

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} \sim N(\boldsymbol{\theta}, \mathcal{I}(\hat{\boldsymbol{\theta}}_{\text{MLE}})^{-1}). \tag{1.12}$$

The ijth entry of the inverse matrix provides the covariance between the ith and jth parameters, and the iith entry provides the variance for the ith parameter. Then the $(1-\gamma)100\%$ asymptotic confidence interval of the ith parameter is

$$\hat{\theta}_i \pm Z_{\gamma/2}\sqrt{\mathcal{I}_{ii}^{-1}}, \tag{1.13}$$

where $Z_{\gamma/2}$ is the $\gamma/2$th upper percentile of the standard normal distribution.

1.5 Bayesian Estimation

Bayesian methods, unlike frequentist approaches, view the unknown parameters of a model as random variables, and express all uncertainty (including the uncertainty of the model's unknown parameters) in terms of a posterior probability [14]. In a Bayesian framework, assume we have a data set $\boldsymbol{x}$, and the data set follows a probabilistic model which is described by $\boldsymbol{\theta}$ (a vector of unknown parameters). Our belief about the uncertainty of $\boldsymbol{\theta}$ which is decided before viewing the data, is represented by $p(\boldsymbol{\theta})$, the prior -probability density function. With the use of Bayes' theorem, the beliefs of $\boldsymbol{\theta}$ are updated with the use of the likelihood function, $L(\boldsymbol{x}|\boldsymbol{\theta})$, and provides us with the posterior probability density function of $\boldsymbol{\theta}$, given the data, as

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{\theta}) \cdot L(\boldsymbol{x}|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta}) \cdot L(\boldsymbol{x}|\boldsymbol{\theta}) d\theta}. \tag{1.14}$$

For certain purposes, the posterior probability density function may only need to be defined up to a constant of proportionality, namely a normalizing constant. Therefore, the previous result can be defined as

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{\theta}) \cdot L(\boldsymbol{x}|\boldsymbol{\theta}). \tag{1.15}$$

Therefore, the posterior distribution is proportional to the product of likelihood function and prior probability density function. According to Raftery, the most useful

summaries of the posterior are the posterior mean, posterior mode, and the Bayesian analogue of the frequentist confidence intervals, credible intervals [14].

Loss Function

Another important concept in Bayesian analysis is the loss function. The loss function, $\text{Loss}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}|\boldsymbol{x})$, measures the cost of making an estimate $\hat{\boldsymbol{\theta}}$ if the true value is $\boldsymbol{\theta}$, given $\boldsymbol{x}$. Some examples of the loss function are squared-error loss, binary loss, and absolute loss function. For the thesis, we will look into the squared-error loss function. The squared-error loss function is given as

$$\text{Loss}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}|\boldsymbol{x}) = (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta})^2. \tag{1.16}$$

The estimate $\hat{\boldsymbol{\theta}}$ can be chosen to minimize the posterior expected loss,

$$E(\text{Loss}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}|\boldsymbol{x})) = \int \text{Loss}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}|\boldsymbol{x}) \times p(\boldsymbol{\theta}|\boldsymbol{x}) d\theta \tag{1.17}$$

If $\hat{\boldsymbol{\theta}}$ is the value that minimizes the posterior expected loss over all other values, then it is the Bayes estimate and is denoted $\hat{\boldsymbol{\theta}}_{\text{Bayes}}$. Under squared-error loss, the posterior mean is the Bayes estimate,

$$\hat{\boldsymbol{\theta}}_{\text{Bayes}} = \int \boldsymbol{\theta} \times p(\boldsymbol{\theta}|\boldsymbol{x}) d\theta. \tag{1.18}$$

A problem in Bayesian analysis arises when there is no closed-form expression for the normalizing constant and therefore the posterior distribution. To overcome this barrier, the implementation of Markov Chain Monte Carlo (MCMC) is used to approximate the posterior distribution and draws from this approximated posterior distribution are used to determine the posterior mean, credible intervals, and other characteristics.

1.6 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a technique in which values are randomly generated and drawn from a probability distribution function. A common approach to MCMC is the Metropolis-Hastings algorithm. A special trait of the Markov chain is Markov's property which states that the future state of a process depends only upon the present state. A key ingredient for MCMC is the proposal density. The proposal must be chosen such that it is easy to simulate from, and it mimics the targeted distribution. The algorithm may differ depending on the choice of the proposal distribution. The two main approaches to the Metropolis-Hastings are the random walk chain and the independent chain. In this thesis, we will consider the random walk chain.

Metropolis-Hastings

In Bayesian estimation, the Metropolis-Hastings algorithm is used to simulate draws of $\boldsymbol{\theta}$ which are sequentially drawn and simulated from an approximate posterior distribution, $p(\boldsymbol{\theta}|\boldsymbol{x})$. The Metropolis-Hastings algorithm has the following procedure:

1. Determine the number of steps or draws M.
2. Determine the initial value $\boldsymbol{\theta}_0$.
3. For $m = 1, 2, ..., M$,
   (a) Simulate a candidate value $\boldsymbol{\theta}^*$ from a proposal density $g(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{m-1})$.
   (b) Compute the following ratio

$$R = \frac{p(\boldsymbol{\theta}^*|\boldsymbol{x})g(\boldsymbol{\theta_{m-1}}|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}_{m-1}|\boldsymbol{x})g(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{m-1})}.$$

   (c) Compute the acceptance probability $P = \min\{1, R\}$.

(d) Set

$$\theta_t = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } P \\ \boldsymbol{\theta}_{m-1} & \text{with probability } 1-P \end{cases}$$

If we consider a random walk chain, then the proposal density satisfies

$$g(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{m-1}) = h(\boldsymbol{\theta}^* - \boldsymbol{\theta}_{m-1}),$$

where h is a symmetric density about the origin. Since we are using the random walk chain, then the ratio from step 3(b) simplifies to

$$R = \frac{p(\boldsymbol{\theta}^*|\boldsymbol{x})}{p(\boldsymbol{\theta}_{m-1}|\boldsymbol{x})}$$

Possible choices for the proposal density could be a multivariate Gaussian distribution or a multivariate t-distribution. In this thesis, we will consider a multivariate t-distribution as a proposal density. A multivariate t-distribution has an advantage over the Gaussian distribution in that the tails are heavier on the t-distribution, allowing the chain to more freely explore a broader range of values.

1.7 Outline of Thesis

In this thesis, different parametric and non-parametric methods will be implemented to model the competing risks. Chapter 1 served as relevant background information needed to model the competing risks, as well as discussing the techniques needed to estimate the parameters of the competing risks models. In Chapter 2, a mathematical definition of competing risks model is stated, as well as the notation used throughout the thesis. The assumptions of the competing risks model are discussed and the relations between the reliability measures are also shown in this chapter. In Chapter 3, the non-parametric methods such as the Kaplan-Meier and the estimator of the cumulative incidence function are outlined. In Chapter 4, parameter estimation of

a homogeneous competing risks model is discussed. By a homogeneous model, we discuss the competing risks framework in which no covariates are assumed. Parameter estimation is conducted using the maximum likelihood method, and Bayesian method by implementing a Markov Chain Monte Carlo algorithm, Metropolis-Hastings. The Metropolis-Hastings algorithm is used to simulate draws from a targeted posterior density to obtain the Bayes estimate, as well as other characteristics of the posterior density. In Chapter 5, we incorporate the additional information on the objects by means of a competing risks regression model. The typical approach for regression modelling is the proportional hazards. Parameter estimation is conducted using Cox's partial likelihood function, maximum likelihood estimation and Bayesian estimation. In Chapter 6, a data analysis on bone marrow transplant patients using the ideologies from Chapter 2 to 5 is conducted. The data set considers two risks, leukemia relapse and death in remission. In Chapter 7, conclusion of the results are made and future work for competing risks models are stated.

Chapter 2

Model Assumptions and Notation

Consider an object that can experience multiple causes of failures (risks). The object here can be any organism such as a human, or a cancer cell, or the object could be inanimate such as a computer. Assume that when the risk occurs, it causes fatality or failure to the object. If a human is considered as the object, then the risks could consist of any risks that causes fatality such as cancer or HIV. The object will fail by only one of the k causes of failures, with $k \geq 2$. Let X be the lifetime of the object. Let T_j be the time at which risk j occurs to the object. The lifetime of the object is the time of the first occurring risk to the object. That is, $X = \min\{T_1, ..., T_k\}$. The goal is to model the risks of the objects using parametric and non-parametric techniques. To do this, assume there are n independent and identical objects on a life test. Let X_i be the lifetime of the i^{th} object, and let T_{ij} be the time of which risk j occurs to object i. Then, for object i, the lifetime of the object is the time of the first occurring risk to the object. That is, $X_i = \min\{T_{i1}, ..., T_{ik}\}$. There are three possible cases when testing each object. In the first case, the object is failed due to a known cause of failure. Here, there are two observable quantities being X (the object's life time) and δ (the object's cause of failure). In the second case, the object does not experience a risk during the study period so the object reaches a censoring time. In this case, there is only one observable quantity being X. The notation $\delta = 0$ is used when the object reaches a censoring time. In the last case, the object is failed due to an unknown cause of failure. Here, there is only one observable quantity being X. The notation $\delta = *$ is used when the cause of failure is unknown.

In summary, competing risks data can be described as a bivariate random variable (X, δ), where X is the time-to-event and δ will:

1. Equal an element from the set $\{1, ..., k\}$ if the object has failed due to a known

risk, and thus X will represent the time at which fatality has occurred by an event in the set.

2. Equal $*$ if the object has failed due to an unknown cause, and thus X will represent the time at which fatality has occurred.

3. Equal 0 if the object does not experience a cause of failure during the study period and thus X represents the censoring time observed.

In Figure 2.1 and 2.2, we illustrate two different representation of the competing risks model. Figure 2.1 shows the model as a fatal-shock model with k risks, where fatality is delivered to the object once the shock (risk) hits the object. Figure 2.2 shows the model as a series system of k components, where the system fails if any one of the k components fail.

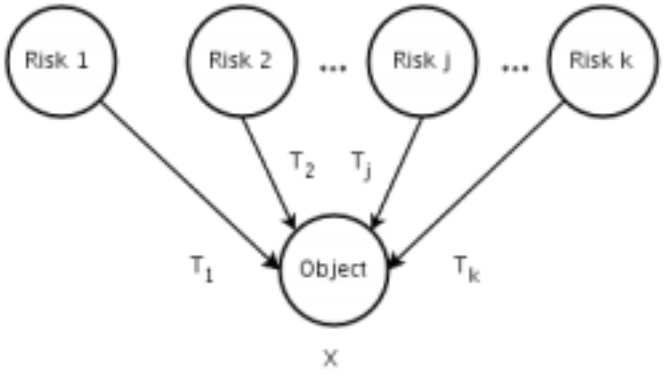


Figure 2.1: Competing risks model with k fatal causes of failure.

Figure 2.2: The object can also be interpreted as a series system with k independent but not identical components. T_j can be interpreted as the lifetime of the j^{th} component.

The following notation will be used in the thesis:

- n: the number of objects on the life test
- k: the number of independent competing risks
- X: the lifetime of an object
- T_j: the time of which risk from cause j occurs to the object
- X_i: the lifetime of object i $(i = 1, ..., n)$
- T_{ij}: the lifetime of cause j $(j = 1, ..., k)$ of object i
- $f(\cdot)$: probability density function of X_i
- $G(\cdot)$: cumulative distribution function (CDF) of X_i
- $S(\cdot)$: survival function of X_i
- $f_j(\cdot)$: probability density function (PDF) of T_{ij}
- $G_j(\cdot)$: cumulative distribution function of T_{ij}
- $S_j(\cdot)$: survival function of T_{ij}
- $h_j(\cdot)$: hazard function of T_{ij}
- mrl$(\cdot)$: mean residual life at a specified time
- δ_i: indicator variable denoting known or unknown cause of failure or censored for object i

In Chapter 1, the reliability measures of X were discussed and defined. In the following, we further discuss the functions used to characterize the competing risks. Using the marginal survival function, $S_j(x)$, the marginal hazard function can be derived when assuming independent competing risks. The marginal hazard function is

$$h_j^M(x) = \lim_{\Delta x \to 0} \frac{P(x < T_j \leq x + \Delta x | T_j > x)}{\Delta x}. \tag{2.1}$$

The cause-specific hazard rate for cause j is given as

$$\begin{aligned} h_j(x) &= \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x, \delta = j | X > x)}{\Delta x} && (2.2) \\ &= \lim_{\Delta x \to 0} \frac{P\left(x < T_j \leq x + \Delta x, \delta = j | T_j > x, j = 1, \ldots, k\right)}{\Delta x}. && (2.3) \end{aligned}$$

The cause-specific hazard function calculates the instantaneous rate at which individuals who have yet to experience any of the causes are experiencing the jth cause of failure. Therefore, the cause-specific hazard is the instantaneous rate of occurrence of experiencing cause j for individuals who are currently event-free [2]. The sum of the k cause-specific hazard gives us the overall hazard rate

$$h(x) = \sum_{j=1}^{k} h_j(x). \tag{2.4}$$

When the competing risks are independent, the marginal hazard and cause-specific hazard function are equivalent to one another. However, this is not the case when the competing risks are dependent [6, p. 51]. Furthermore, we can estimate the cause-specific hazard function regardless of if the risks are dependent or independent, but the marginal hazard function cannot be identified without knowing the dependence structure of the risks [6, p. 51]. Often, one is interested in the probability of a competing risk occurring rather than the hazard rate function. A characteristic function that estimates this probability is the cumulative incidence function (CIF), also known as the sub-distribution function or crude cumulative incidence function. The CIF for cause j is defined as

$$\begin{aligned} F_j(x) &= P(X \leq x, \delta = j) && (2.5)\\ &= \int_0^x h_j(u) \prod_{j=1}^{k} S_j(u)\, du && (2.6)\\ &= \int_0^x h_j(u) \exp\left(- \sum_{j=1}^{k} H_j(u) \right)\, du && (2.7)\\ &= \int_0^x h_j(u) S(u)\, du \quad j = 1, 2, ..., k. && (2.8) \end{aligned}$$

The CIF calculates the probability of failure by time x due to a specific cause in the presence of all other causes. It is a non-decreasing function with $F_j(0) = 0$. The CIF is also not a probability distribution since $F_j(\infty) = P(X \leq \infty, \delta = j) < 1$ [6, p. 52].

The relative risk from cause j, or probability of failure by cause j, is defined as

$$\pi_j = \lim_{x\to\infty} F_j(x) = \int_0^\infty h_j(x)S(x)dx, \quad j = 1, 2, ..., k. \tag{2.9}$$

It is worthwhile to state that the CIF can be estimated without knowing the dependence structure of the risks [6, p. 53], however, Equations (2.6) and (2.7) are formulated with the assumption of independent competing risks. Overall, the thesis will consider independent causes of failures (risks). With the assumption of independent competing risks, we can define the reliability measures as follows. The overall survival function of X is

$$\begin{aligned}
S(x) &= P(X > x) && (2.10)\\
&= \int_x^\infty f(u)du && (2.11)\\
&= P(\min(T_1, ..., T_k) > x) && (2.12)\\
&= P(T_1 > x) \cdot P(T_2 > x) \cdot ... \cdot P(T_k > x) && (2.13)\\
&= \prod_{j=1}^{k} P(T_j > x) && (2.14)\\
&= \prod_{j=1}^{k} S_j(x) && (2.15)\\
&= \exp\left(-\sum_{j=1}^{k} H_j(x)\right), && (2.16)
\end{aligned}$$

where $H_j(x)$, the cumulative hazard rate function for cause j, is

$$H_j(x) = \int_0^x h_j(u)du. \tag{2.17}$$

The overall survival function can be interpreted as the probability of not failing from any cause by time x [12]. The cumulative distribution function of X is

$$\begin{aligned}
G(x) &= P(X \le x) && (2.18)\\
&= \int_0^x f(u)du && (2.19)\\
&= 1 - \exp\left(-\sum_{j=1}^{k} H_j(x)\right). && (2.20)
\end{aligned}$$

The probability density function of X is

$$\begin{aligned} f(x) &= \frac{d}{dx}F(x) && (2.21)\\ &= \sum_{j=1}^{k} f_j(x) \prod_{l=1, l\neq j}^{k} S_l(x) \\ &= \sum_{j=1}^{k} h_j(x) S_j(x) \prod_{l=1, l\neq j}^{k} S_l(x) \\ &= \sum_{j=1}^{k} h_j(x) \prod_{l=1}^{k} S_l(x) && (2.22)\\ &= \sum_{j=1}^{k} h_j(x) S(x). && (2.23) \end{aligned}$$

The mean residual life at time x is

$$\begin{aligned} \text{mrl}(x) &= E(X - x | X > x) && (2.24)\\ &= \frac{\int_x^\infty \prod_{j=1}^{k} S_j(t) dt}{\prod_{j=1}^{k} S_x(t) dx}. && (2.25) \end{aligned}$$

The next goal is to model the competing risks with the help of the quantities above. If we assume no parametric assumptions in the model, we consider the non-parametric estimation techniques as discussed in the following chapter. However, if we assume the risks follow parametric forms, then we consider the parametric estimation as discussed in Chapter 4.

Chapter 3

Non-Parametric Estimation

In this section, the main goals are to model the competing risks using only the observational data. In other words, we assume no parametric form or assumptions. Typically, when competing risks are present, the two main approaches of analysing the data are by firstly, ignoring the competing risks, and secondly, by acknowledging the competing risks [11, p. 5]. The techniques presented in this section are the Kaplan-Meier and the empirical cumulative incidence function. The Kaplan-Meier is an approach that ignores the competing risks, where the cumulative incidence function accounts for the competing risks.

3.1 Kaplan-Meier Estimator

As mentioned in Chapter 1, the survival function is the probability of surviving an event beyond a specified time. A non-parametric estimator of the survival function is the Kaplan-Meier (KM) estimator, also known as the Product-Limit estimator [11, p. 31]. The Kaplan-Meier estimates the probability of objects or individuals who have not experienced an event within a specified time. In other words, it estimates the event-free survival over time. Inversely, we can estimate the proportion of individuals who have experienced an event within a specified time by calculating one minus the event-free survival (1−KM). By inversely using the KM, we estimate the probability of practicing an event. The KM provides accurate probabilities for all values of x within the range of the data; however, not for values beyond the largest observation ($x_{\max}$) in the data. The KM estimator is given by

$$\hat{S}(x) = \begin{cases} 1 & x < x_1 \\ \prod_{x_i \leq x} \left[1 - \frac{d_i}{n_i}\right] & x_1 \leq x, \end{cases}$$

where x_i are distinct event times, d_i denotes the number of events by time x_i, and n_i are the number of subjects at risk by time x_i [5, p. 411]. More specifically, n_i denotes the number of subjects who have not yet experienced the event or are not censored by time x_i. In the case of no censored observation, the KM estimator becomes the empirical survival function, which is the proportion of failure times that are larger than time x. That is,

$$\hat{S}(x) = \frac{\text{Number of } x_i > x}{n}. \tag{3.1}$$

The KM estimator of the cumulative hazard function, $H(x)$ is defined as

$$\hat{H}(x) = -\log(\hat{S}(x)).$$

To model the competing risks using the KM estimator, we must make the following assumptions. Consider the data $(X_1, \delta_1), (X_2, \delta_2), ..., (X_n, \delta_n)$, where $\delta_i = 0$ for censored times, δ_i equals $1, 2, ..., k$ for known causes of failure, and $\delta_i = *$ for unknown causes of failure, for $i = 1, 2, ..., n$. Then, the KM estimator for the jth cause $(j = 1, 2, ..., k)$ is given by treating all X_i in which $\delta_i \neq j$ as censored observations. This simply implies that all observations other than the primary risk of interest are treated as right-censored observations. The probability from this estimator can be intepreted as the probability of dying from one specific cause of failure in a hypothetical world in which the object can not die from any other causes [6, p. 52].

3.2 Cumulative Incidence Function

As mentioned previously, the cumulative incidence function calculates the probability of failure by time x due to a specific risk in the presence of all other risks. To define the empirical CIF, consider the following. Let $x_1 < x_2 < ... < x_m$ be distinct event times. Let n_ℓ be the number of subjects at risk at time x_ℓ, $d_{j\ell}$ be the number of subjects experiencing cause j at time x_ℓ, and y_ℓ be the number of subjects experiencing from any other cause than cause j at time x_ℓ. Then, the estimator for the CIF of cause j is defined as

$$\hat{F}_j(x) = \begin{cases} 0 & x < x_1 \\ \sum_{x_l \leq x} \left[\prod_{j=1}^{l-1} \frac{1-(y_j+d_{jl})}{n_j} \right] \frac{d_{jl}}{n_l} & x_1 \leq x. \end{cases} \tag{3.2}$$

The estimator of the CIF can also be expressed in terms of the Kaplan-Meier estimator:

$$\hat{F}_j(x) = \sum_{x_l \leq x} \frac{d_{jl}}{n_l} \hat{S}(x_{l-1}) \text{ for } x_1 \leq x, \tag{3.3}$$

where $\hat{S}(x)$ is the Kaplan-Meier estimate of the overall survival function [5, p. 415].

Chapter 4

Homogeneous Competing Risks Model

The main goal for this chapter is to estimate the unknown parameters of the distributions modelling the competing risks data. We assume a homogeneous model by considering no covariates. The parameter estimation will consist of using classical techniques such as the maximum likelihood method, as well as Bayesian method by using MCMC, to estimate the unknown parameters. The results of the classical and Bayesian methods will be further discussed using the STH-2 distribution.

4.1 Maximum Likelihood Method

In this section, the unobservable random variables T_{ij} are assumed to be independent and identically distributed over objects $i = 1, 2, ..., n$, and independent but not identically distributed over causes $j = 1, 2, ..., k$. With this assumption, our data have the form $(X_1, \delta_1), (X_2, \delta_2), ..., (X_n, \delta_n)$, where $X_i = \min(T_{i1}, T_{i2}, \cdots, T_{ik})$, $\delta_i = 0$ for censored data, δ_i equals $1, 2, ..., k$ for known causes of failure, and $\delta_i = *$ for unknown causes of failure. The likelihood function given by Equation (1.8) must be adjusted to allow for censored observations. If the observation x_i is a failure time, then its contribution to the likelihood function is given by the probability density function $f(x_i)$. If the observation x_i is censored, then its contribution to the likelihood function is given by the survival function $S(x_i)$. Then, the joint likelihood function over all observations is given as

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} \left\{ \Big[f(x_i; \boldsymbol{\theta}) \Big]^{I(\delta_i \neq 0)} \Big[S(x_i; \boldsymbol{\theta}) \Big]^{I(\delta_i = 0)} \right\}, \tag{4.1}$$

where $\boldsymbol{\theta}$ is a vector of parameters included in the model. The length of this vector is based on which distribution is considered for the risks. If the risks are assumed to

follow a one-parameter distribution, then $\boldsymbol{\theta}$ has length given by (4.2). If the risks are assumed to follow a two-parameter distribution, then $\boldsymbol{\theta}$ has length given by (4.3).

$$\boldsymbol{\theta} = (\alpha_1, ..., \alpha_k) \tag{4.2}$$

$$\boldsymbol{\theta} = (\alpha_1, ..., \alpha_k, \beta_1, ..., \beta_k) \tag{4.3}$$

In general, if the risks follow a p-parameter distribution, then the length of the vector of unknown parameters is $p \times k$.

Using the relationship between the hazard rate, survival function, and probability density function, the likelihood function (4.1) can be rewritten as shown below. The first part of equation 4.1 is rewritten as follows:

$$\begin{aligned}
&\Big[f(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i\neq 0)} \\
= &\prod_{j=1}^{k}\left[f_j(x_i;\boldsymbol{\theta})\prod_{l=1,l\neq j}^{k}S_l(x_i;\boldsymbol{\theta})\right]^{I(\delta_i=j)}\left[\sum_{j=1}^{k}f_j(x_i;\boldsymbol{\theta})\prod_{l=1,l\neq j}^{k}S_l(x_i;\boldsymbol{\theta})\right]^{I(\delta_i=*)} \\
= &\prod_{j=1}^{k}\left[h_j(x_i;\boldsymbol{\theta})\prod_{l=1}^{k}S_l(x_i;\boldsymbol{\theta})\right]^{I(\delta_i=j)}\left[\sum_{j=1}^{k}h_j(x_i;\boldsymbol{\theta})\prod_{l=1}^{k}S_l(x_i;\boldsymbol{\theta})\right]^{I(\delta_i=*)} \\
= &\prod_{j=1}^{k}\Big[h_j(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i=j)}\prod_{j=1}^{k}\Big[S(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i=j)}\Big[\sum_{j=1}^{k}h_j(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i=*)}\Big[S(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i=*)} \\
= &\prod_{j=1}^{k}\left\{\Big[h_j(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i=j)}\Big[\sum_{j=1}^{k}h_j(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i=*)}\right\}\Big[S(x_i;\boldsymbol{\theta})\Big]^{\sum_{j=1}^{k}I(\delta_i=j)+I(\delta_i=*)}.
\end{aligned}$$

When we combine this with the second half of expression 4.1, the exponents on both survival function are grouped together, which results in

$$\sum_{j=1}^{k}I(\delta_i=j)+I(\delta_i=*)+I(\delta_i=0)=\sum_{j=0}^{k}I(\delta_i=j)+I(\delta_i=*)=1.$$

Then, the likelihood function can be written as

$$L(\boldsymbol{\theta}|\boldsymbol{x})=\prod_{i=1}^{n}\left[\prod_{j=1}^{k}\Big[h_j(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i=j)}\Big[\sum_{j=1}^{k}h_j(x_i;\boldsymbol{\theta})\Big]^{I(\delta_i=*)}S(x_i;\boldsymbol{\theta})\right]. \tag{4.4}$$

Therefore the log-likelihood function is

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{i=1}^{n} \left[\sum_{j=1}^{k} \Big[I(\delta_i = j) \log(h_j(x_i;\boldsymbol{\theta})) \Big] + I(\delta_i = *) \log\Big(\sum_{j=1}^{k} h_j(x_i;\boldsymbol{\theta}) \Big) + \log(S(x_i;\boldsymbol{\theta})) \right].$$

Using Equation (2.16), the log-likelihood function can also be expressed as

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{i=1}^{n} \left[\sum_{j=1}^{k} [I(\delta_i = j) \log(h_j(x_i;\boldsymbol{\theta})) - H_j(x_i;\boldsymbol{\theta})] + I(\delta_i = *) \log \left(\sum_{j=1}^{k} h_j(x_i;\boldsymbol{\theta}) \right) \right]. \tag{4.5}$$

which is strictly dependent on $h_j(x_i)$. Equation (4.5) gives the log-likelihood function for a competing risks model in terms of the hazard rate functions of the individual risks. This equation can be used to determine the log-likelihood function for any of the cases with the risks distributions listed in Table 4.1. To determine the MLEs, the likelihood function, or log-likelihood function, should be maximized using methods discussed in Chapter 1. To further discuss this procedure, we examine the STH-2 case.

4.1.1 Maximum Likelihood for STH-2 Case

In this subsection, we assume that the random variable T_{ij} follows the STH-2 distribution with parameters α_j and β_j, for $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$. The survival function and hazard rate function of cause j are respectively as

$$S_j(x) = \frac{\alpha_j \beta_j \left[\beta_j + (1 + 2\beta_j x^{\alpha_j}) \exp(-\beta_j x^{\alpha_j}) \right]}{\beta_j + (1 + \beta_j x^{\alpha_j}) \exp(-\beta_j x^{\alpha_j})}, \tag{4.6}$$

$$h_j(x) = \frac{1}{1 + \beta_j} \left[\beta_j + (1 + \beta_j x^{\alpha_j}) \exp(-\beta_j x^{\alpha_j}) \right] \exp(-\beta_j x^{\alpha_j}), \tag{4.7}$$

where $\alpha_j, \beta_j > 0$ and $x > 0$. Substituting Equations (4.6) and (4.7) into the log-likelihood from Equation (4.5) results in

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{N} \Bigg[\sum_{j=1}^{k} \Big[I(\delta_i = j) \big[&- \log\big(1+\beta_j\big) + \log(\beta_j + (1+\beta_j x^{\alpha_j}) \exp(-\beta_j x^{\alpha_j})\big) \\
&- \beta_j x^{\alpha_j} \Big] + I(\delta_i = *) \Big[\sum_{j=1}^{k} \log\big(\frac{1}{1+\beta_j} [\beta_j + (1+\beta_j x^{\alpha_j}) \exp(-\beta_j x^{\alpha_j})] \exp(-\beta_j x^{\alpha_j})\big) \Big] \\
&+ \log(\alpha_j \beta_j) + \log\big(\beta_j + (1+2\beta_j x^{\alpha_j}) \exp(-\beta_j x^{\alpha_j})\big) \\
&- \log\big(\beta_j + (1+\beta_j x^{\alpha_j}) \exp(-\beta_j x^{\alpha_j})\big) \Big],
\end{aligned}
\tag{4.8}
$$

where $\boldsymbol{\theta} = (\alpha_1, ..., \alpha_k, \beta_1, ..., \beta_k)$. In this case, there is no general analytic solution for the maximum likelihood estimators. Therefore, numerical methods are needed to maximize the log-likelihood function. To perform this task, we implement the Nelder-Mead method in R by using optim. This function is provided the negative log-likelihood to minimize, which equivalently maximizes the log-likelihood. The output of this function will return the MLEs, as well as the hessian matrix. There were also no analytic solutions for the maximum likelihood estimators when the risks followed the other useful lifetime distributions. Therefore, numerical methods are required for all cases to estimate the unknown parameters of the models.

Distribution	$h_j(x_i)$	$H_j(x)$
Exponential	α_j	$\alpha_j x$
STH-1	$\frac{\alpha_j\left[\alpha_j+(1+2\alpha x)e^{-\alpha_j x}\right]}{\alpha_j+(1+\alpha_j x)e^{-\alpha_j x}}$	$\log(\alpha_j+1)-\log(\alpha_j+(1+\alpha_j x)e^{-\alpha_j x})+\alpha_j x$
Lindley	$\frac{\alpha_j^2(1+x)}{\alpha_j x+\alpha_j+1}$	$-\log(\alpha_j x+\alpha_j+1)+\log(\alpha_j+1)+\alpha_j x$
STH-2	$\frac{\alpha_j\beta_j\left[\beta_j+(1+2\beta_j x^{\alpha_j})e^{-\beta_j x^{\alpha_j}}\right]}{\beta_j+(1+\beta_j x^{\alpha_j})e^{-\beta_j x^{\alpha_j}}}$	$\log(1+\beta_j)-\log(\beta_j+(1+\beta_j x^{\alpha_j})e^{-\beta_j x^{-\alpha_j}})+\beta_j x^{-\alpha_j}$
Power Lindley	$\alpha_j\beta_j^2\frac{(1+\alpha_j)x^{\alpha_j-1}}{\beta_j+1+\beta_j x^{\alpha_j}}$	$-\log(\alpha_j x+\alpha_j+1)+\log(\alpha_j+1)+\alpha_j x$
Weibull	$\alpha_j\beta_j x^{\beta_j-1}$	$\alpha_j x^{\beta_j-1}$
Bathtub	$\alpha_j\beta_j x^{\beta_j-1}e^{x^{\beta_j}}$	$-\alpha_j(1-e^{x^{\beta_j}})$

Table 4.1: The hazard rate and cumulative hazard rate of risk j for each distribution.

4.2 Bayesian Estimation

In this section, $\boldsymbol{\theta}$ still denotes a vector containing the parameters of the risk model, where the length is given by (4.2) or (4.3). However, we treat the model parameters as random variables and assume these random variables follow some distribution. Here, we assume each parameter follows a Gamma distribution with a set of non-negative hyperparameters, where all hyperparameters are assumed to be known. For the one-parameter models, we have

$$\alpha_j \sim \text{Gamma}(a_{j1}, a_{j2}), \quad j = 1, 2, \cdots, k. \tag{4.9}$$

For the two-parameter models (TPM), we have

$$\begin{aligned} \alpha_j &\sim \text{Gamma}(a_{j1}, a_{j2}), \quad j = 1, 2, \cdots, k, \\ \beta_j &\sim \text{Gamma}(b_{j1}, b_{j2}), \quad j = 1, 2, \cdots, k. \end{aligned} \tag{4.10}$$

The joint prior probability density function of $\boldsymbol{\theta}$ is given as

$$p(\boldsymbol{\theta}) = \prod_{j=1}^{k} \left[\left(\alpha_j^{a_{j1}-1} \exp\left(\frac{-\alpha_j}{a_{j2}}\right) \right) \left(\beta_j^{b_{j1}-1} \exp\left(\frac{-\beta_j}{b_{j2}}\right) \right)^{I(\text{TPM})} \right], \tag{4.11}$$

where I(TPM) is an indicator variable that equals 1 if the risk model is a two-parameter model, and 0 otherwise. The logarithm of the joint prior distribution of $\boldsymbol{\theta}$ is

$$\log(p(\boldsymbol{\theta})) = \sum_{j=1}^{k} \left[(a_{j1} - 1)\log(\alpha_j) - \frac{\alpha_j}{a_{j2}} + I(\text{TPM}) \left((b_{j1} - 1)\log(\beta_j) - \frac{\beta_j}{b_{j2}} \right) \right] \tag{4.12}$$

The joint posterior probability density function of $\boldsymbol{\theta}$, given the data, up to a normalizing constant, is a product of the likelihood function and joint prior distribution, and

is given by

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = cp(\boldsymbol{\theta}) \cdot L(\boldsymbol{x}|\boldsymbol{\theta}),$$

where c is a constant. Then taking the logarithm of the joint posterior probability density function, the log-posterior is simply proportional to a sum of the log-likelihood function (4.5) and log-prior distribution (4.12). That is,

$$\log(p(\boldsymbol{\theta}|\boldsymbol{x})) \propto \log(p(\boldsymbol{\theta})) + \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}).$$

Then, the log-posterior probability density function of $\boldsymbol{\theta}$, given $\boldsymbol{x}$, up to a constant, is given by

$$\begin{aligned} &= \sum_{i=1}^{n} \left[\sum_{j=1}^{k} [I(\delta_i = j) \log(h_j(x_i; \boldsymbol{\theta})) - H_j(x_i; \boldsymbol{\theta})] + I(\delta_i = *) \log \left(\sum_{j=1}^{k} h_j(x_i; \boldsymbol{\theta}) \right) \right] \\ &\quad + \sum_{j=1}^{k} \left[(a_{j1} - 1) \log(\alpha_j) - \frac{\alpha_j}{a_{j2}} + I(\text{TPM}) \left((b_{j1} - 1) \log(\beta_j) - \frac{\beta_j}{b_{j2}} \right) \right]. \end{aligned} \tag{4.13}$$

Since the posterior probability density function is not in a standard form of a known distribution, numerical methods are required to determine the estimation of the parameters. The Metropolis-Hastings algorithm is used to simulate a random sample from a distribution that converges to the true log-posterior distribution. Under squared-error loss, the Bayes estimate is equal to the posterior mean. Therefore, the Bayes estimate for each parameter can be calculated by determining the mean of the simulated draws. Other characteristics of the model parameters can be determined using the simulated draws as well, such as $(1 - \gamma)\%$ credible intervals. To apply the discussions of this section, we consider the STH-2 case. However, without a loss of generality, the procedure in the subsection may be applied to any of the risks models.

4.2.1 Bayesian Estimation for STH-2 Case

In this subsection, we assume that the random variable T_{ij} follows a STH-2 distribution with parameters α_j and β_j, for $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$. Then, the joint prior and log-prior probability density function, respectively, are given as

$$\begin{aligned} p(\boldsymbol{\theta}) &= \prod_{j=1}^{k} \left[\left(\alpha_j^{a_{j1}-1} \exp\left(\frac{-\alpha_j}{a_{j2}} \right) \right) \left(\beta_j^{b_{j1}-1} \exp\left(\frac{-\beta_j}{b_{j2}} \right) \right) \right], \\ \log(p(\boldsymbol{\theta})) &= \sum_{j=1}^{k} \left[(a_{j1}-1)\log(\alpha_j) - \frac{\alpha_j}{a_{j2}} + \left((b_{j1}-1)\log(\beta_j) - \frac{\beta_j}{b_{j2}} \right) \right]. \end{aligned}$$

The log-posterior is given by adding the log-prior and the log-likelihood (4.8) together. The result for the log-posterior has no closed form expression, and thus requires numerical methods. To do this, we use implement MCMC using the Metropolis-Hastings random walk algorithm. Using this algorithm will require the parameter to be on the real number line. To convert the non-negative parameters of STH-2, a transformation of variable is applied. Let $\boldsymbol{\zeta} = \log(\boldsymbol{\theta}) = (\log(\alpha_1), ..., \log(\alpha_k), \log(\beta_1), ..., \log(\beta_k))$ and is of length $2k$. The posterior of the transformed posterior will then be multiplied by the determinant of Jacobian matrix. The determinant of the Jacobian is

$$|J| = \begin{pmatrix} \exp(\zeta_1) & 0 & \cdots & 0 \\ 0 & \exp(\zeta_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \exp(\zeta_{2k}) \end{pmatrix} = \exp\left(\sum_{u=1}^{2k} \zeta_u \right) \tag{4.14}$$

Then, the log-posterior of the transformed vector of unknown parameter, is given as

$$\log p(\boldsymbol{\zeta}|\boldsymbol{x}) = \log p(\boldsymbol{\theta} = \exp(\boldsymbol{\zeta})|\boldsymbol{x}) + \sum_{u=1}^{2k} \zeta_u. \tag{4.15}$$

The draws from the Metropolis-Hastings algorithm can be converted back to non-negative values by using the inverse of the transformation, $\boldsymbol{\theta} = \exp(\boldsymbol{\zeta})$.

Chapter 5

Competing Risks Regression Model

In the previous chapter, we discussed a competing risks model with only the use of the time-to-event data (homogeneous model). In this chapter, we incorporate the additional information on the objects by presenting a model that quantifies the relationship between the time-to-event and a set of covariates by introducing a competing risks regression model. The traditional approach for regression modelling to analyse time-to-event data with covariates is the proportional hazards model. When including covariates, the observations for the ith object would be $(X_i, \delta_i, \boldsymbol{Z}_i)$, where X_i and δ_i are still defined as previously, and $\boldsymbol{Z}_i$ is the covariate, or similarly risk factor, for object i. For simplicity, we use D_i to represent the available data set that consists of $(X_i, \delta_i, \boldsymbol{Z_i})$, for $i = 1, 2, \cdots, n$. The main goals of this chapter are to estimate the parameters and other characteristics of the regression model by Cox's partial likelihood function, maximum likelihood estimation and Bayesian estimation. The vector of the model parameters $\boldsymbol{\theta}$ will contain the original vector of unknown parameters, as well as the regression coefficients $\boldsymbol{\gamma}$.

5.1 Proportional Hazards Model

Before examining the competing risks regression model, we present the proportional hazards model that is commonly used in survival analysis. Let $h(x; \mathbf{Z})$ represent the hazard rate at time x with covariate vector $\mathbf{Z}$. The proportional hazard model is

defined as

$$\begin{aligned} h(x;\mathbf{Z}) &= h_0(x)\exp(\boldsymbol{\gamma}'\mathbf{Z}) && (5.1)\\ &= h_0(x)\exp\Big(\sum_{q=1}^{p}\gamma_q Z_q\Big), && (5.2)\end{aligned}$$

where $h_0(x)$ is a baseline hazard function that may take any parametric form, for example, a form from Table 1.1, and $\boldsymbol{\gamma}$ is the vector of coefficients for the covariates. The baseline hazard function is also known as the underlying hazard function or as the hazard function for a standard subject, which is a subject with $\boldsymbol{\gamma} = 0$ [5, p. 428]. The baseline hazard function may also be left unspecified by implementing Cox's proportional hazard model. The portion $\exp(\boldsymbol{\gamma}'\mathbf{Z})$ is also known as the relative hazard function and is often the primary interest as it describes the relative effects of the covariates [5, p. 428]. The model is stated as a proportional hazard model because when comparing two individuals with covariates Z_a and Z_b, the ratio of their hazard rates is

$$\frac{h(x;\boldsymbol{Z_a})}{h(x;\boldsymbol{Z_b})} = \frac{h_0(x)\exp\Big(\sum_{k=1}^{p}\gamma_k Z_{ak}\Big)}{h_0(x)\exp\Big(\sum_{k=1}^{p}\gamma_k Z_{bk}\Big)} = \exp\Big(\sum_{k=1}^{p}\gamma_k(Z_{ak}-Z_{bk})\Big)$$

and is independent of time x. This implies that the hazard for any objects must be proportional and cannot cross. Assuming a parametric form for the hazard, the proportional hazard model may also be expressed in terms of the cumulative hazard function and survival function, respectively, as

$$\begin{aligned} H(x;\boldsymbol{Z}) &= H_0(x)\exp(\boldsymbol{\gamma}'\boldsymbol{Z}), && (5.3)\\ S(x;\boldsymbol{Z}) &= \exp[-H_0(x)\exp(\boldsymbol{\gamma}'\boldsymbol{Z})]\\ &= \exp[-H_0(x)]^{\exp(\boldsymbol{\gamma}'\boldsymbol{Z})}\\ &= [S_0(x)]^{\exp(\boldsymbol{\gamma}'\boldsymbol{Z})}, && (5.4)\end{aligned}$$

where $H_0(x)$ is a baseline cumulative hazard function and $S_0(x)$ is a baseline survival function. The proportional hazard model assumes a linear relationship between the covariates and log baseline hazard. That is,

$$\log(h(x; \boldsymbol{Z})) = \log(h_0(x)) + \boldsymbol{\gamma}' \boldsymbol{Z}. \tag{5.5}$$

In the presence of competing risks, an adjustment to the proportional hazard model is needed to model the cause-specific hazard function. In the next sections, we present the adapted model and discuss parameter estimation by methods of Cox's partial likelihood function, maximum likelihood estimation, and Bayesian estimation.

5.2 Estimation using Cox's Cause-Specific Hazard Model

The cause-specific hazard model using Cox's method is given as

$$h_j(x; \boldsymbol{Z}) = h_{0j}(x) \exp(\boldsymbol{\gamma}_j' \boldsymbol{Z}), \tag{5.6}$$

where $h_{0j}(x)$ is an unspecified baseline cause-specific hazard function. This model is known as a semi-parametric model as the baseline cause-specific hazard is treated non-parametrically, and the parametric assumption is made on the covariate effect on the hazard rate [6, p. 244]. More specifically, the model assumes that the log cause-specific hazard is linearly related to the covariates. If one is interested in the inference about about $\boldsymbol{\gamma}$ rather than the shape of $h_{0j}(x)$, then the Cox model is ideal as it allows us to ignore $h_{0j}(x)$ and treats it as a nuisance function [5, p. 475]. The Cox model is also preferable when the true cause-specific hazard model is complex [5, p. 475].

The coefficients of cause j, $\boldsymbol{\gamma}_j$, are estimated by using a partial likelihood function for $\boldsymbol{\gamma_j}$. Since the cause-specific hazard function is left unspecified, no parameter estimates for the underlying hazard are made using this method. Let $x_{j1} < x_{j2} < ... < x_{jr}$ denote the r times of cause j failures, for $j = 1, 2, ..., k$. Let Z_{ji} denote the

risk factors for individuals that fail at time x_{ji}. Then, the partial likelihood function is given as

$$L(\gamma_1, ..., \gamma_k) = \prod_{j=1}^{k} \prod_{i=1}^{r} \frac{\exp(\gamma_j^{'} \boldsymbol{Z_{ji}})}{\sum_{R_{ji}} \exp(\gamma_j^{'} \boldsymbol{Z_{ji}})}, \tag{5.7}$$

where R_{ji} is the risk set (individuals still at risk for cause j at time x_{ji}) [13].

5.3 Estimation using Maximum Likelihood

The proportional hazard model has been adapted to model the cause-specific hazard function. The model for the cause-specific hazard for cause j is defined as

$$\begin{aligned} h_j(x; \mathbf{Z}) &= h_{j0}(x) \exp(\boldsymbol{\gamma}_j^{'} \mathbf{Z}) & (5.8) \\ &= h_{j0}(x) \exp\Big(\sum_{q=1}^{p} \gamma_{qj} Z_q \Big) & (5.9) \end{aligned}$$

where $h_{j0}(x)$ is the baseline cause-specific hazard for cause j and $\boldsymbol{\gamma}_j$ is the vector of coefficients for the covariates of the jth cause. Assuming a parametric form for the baseline cause-specific hazard, then we can define the cumulative hazard, survival function, and probability density function of cause j as

$$\begin{aligned} H_j(x; \mathbf{Z}) &= H_{0j}(x) \exp(\boldsymbol{\gamma}_j^{'} \mathbf{Z}), & (5.10) \\ S_j(x; \mathbf{Z}) &= \exp\{-H_j(x; \mathbf{Z})\}, & (5.11) \\ f_j(x; \mathbf{Z}) &= h_{j0}(x) \exp(\boldsymbol{\gamma}_j^{'} \mathbf{Z}) \exp\{-H_j(x; \mathbf{Z}). & (5.12) \end{aligned}$$

To determine the likelihood function, we adjust Equation (4.1) to account for the covariates. The likelihood function, including the coefficients of the covariates, is

$$L(\boldsymbol{\theta}|\boldsymbol{D}) = \prod_{i=1}^{N} \left\{ \Big[f(x_i; \boldsymbol{Z}_i) \Big]^{I(\delta_i \neq 0)} \Big[S(x_i; \boldsymbol{Z}_i) \Big]^{I(\delta_i = 0)} \right\}, \tag{5.13}$$

where $\boldsymbol{\theta}$ contains the unknown parameters of the risks model and the unknown coefficients of the regression (covariates). Combining Equations (5.10), (5.11), and (5.12) into Equation (5.13), the likelihood can be written as

$$L(\boldsymbol{\theta}|\boldsymbol{D}) = \prod_{i=1}^{N} \left[\prod_{j=1}^{k} \Big[h_j(x_i; \boldsymbol{Z}_i) \Big]^{I(\delta_i = j)} \Big[\sum_{j=1}^{k} h_j(x_i; \boldsymbol{Z}_i) \Big]^{I(\delta_i = *)} S(x_i; \boldsymbol{Z}_i) \right] \tag{5.14}$$

Then, using the relation between the survival function and cumulative hazard function, the log-likelihood function is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{D}) = \sum_{i=1}^{N} \Bigg[\sum_{j=1}^{k} \Big[I(\delta_i = j) \Big(\log \big(h_{0j}(x_i) \big) + \boldsymbol{\gamma}_j' \mathbf{Z}_i \Big) - H_{0j}(x_i) \exp(\boldsymbol{\gamma}_j' \mathbf{Z}_i) \Big] \\ + I(\delta_i = *) \Big(\log \Big(\sum_{j=1}^{k} h_{j0}(x_i) \exp(\boldsymbol{\gamma}_j' \mathbf{Z}_i) \Big) \Big) \Bigg]. \end{aligned} \tag{5.15}$$

The estimates are typically found for the parameters of the cause-specific hazard and for the coefficients $\boldsymbol{\gamma}_j$ by maximizing the likelihood function, or similarly, the log-likelihood function. However, there is no analytic solution for the maximum likelihood estimators. Therefore, numerical methods should be used.

5.4 Estimation using Bayesian Methods

In this section, let $\boldsymbol{\theta}$ be a vector of the unknown parameters of the risks model, as well as the unknown coefficients for the covariates. Let p denote the total number of covariates. For the one-parameter risks models, we assume

$$\begin{aligned} \alpha_j &\sim \text{Gamma}(a_{j1}, a_{j2}), \quad j = 1, 2, ..., k, \\ \gamma_{qj} &\sim N(\mu_{qj}, \sigma_{qj}^2), \qquad\quad j = 1, 2, ..., k, \ q = 1, 2, ..., p. \end{aligned} \tag{5.16}$$

For the two-parameter risks models, we assume

$$\begin{aligned} \alpha_j &\sim \text{Gamma}(a_{j1}, a_{j2}), \quad j = 1, 2, ..., k, \\ \beta_j &\sim \text{Gamma}(b_{j1}, b_{j2}), \quad j = 1, 2, ..., k, \\ \gamma_{qj} &\sim N(\mu_{qj}, \sigma^2_{qj}), \qquad j = 1, 2, ..., k,\ q = 1, 2, ..., p, \end{aligned} \tag{5.17}$$

where all the hyperparameters $a_{j1}, a_{j2}, b_{j1}, b_{j2}, \mu_{qj}$, and σ^2_{qj} are assumed to be known. Then the joint prior probability density function of $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta}) = \prod_{j=1}^{k} \left[\alpha_j^{a_{j1}-1} \exp\left(\frac{-\alpha_j}{a_{j2}}\right) \left(\beta_j^{b_{j1}-1} \exp\left(\frac{-\beta_j}{b_{j2}}\right)\right)^{I(\text{TPM})} \exp\left(-\sum_{q=1}^{p} \frac{(\gamma_{qj} - \mu_{qj})}{2\sigma^2_{qj}}\right)\right],$$

where $I(\text{TPM})$ is an indicator variable that equals 1 if it is a two-parameter risks model, and 0 otherwise. Taking the logarithm of this equation, we find the log-prior probability density function as

$$\begin{aligned} \log(p(\boldsymbol{\theta})) = \sum_{j=1}^{k} \Bigg[& (a_{j1} - 1)\log(\alpha_j) - \frac{\alpha_j}{a_{j2}} + I(\text{TPM})\left((b_{j1} - 1)\log(\beta_j) - \frac{\beta_j}{b_{j2}}\right) \\ & - \sum_{q=1}^{p} \frac{(\gamma_{qj} - \mu_{qj})}{2\sigma^2_{qj}} \Bigg] \end{aligned} \tag{5.18}$$

The log-posterior, up to a normalizing constant, is then determined by adding the log-likelihood function (Equation (5.15)) and log-prior probability density function (Equation (5.18)) together. As we seen in Chapter 4, a transformation of variable was applied to convert the non-negative parameters to be real valued. The same transformation is applied in this section, however, the transformation is only applied to the parameters of the baseline cause-specific hazard and not the coefficients of the regression. Therefore, the log-posterior probability density function of the transformed vector of model parameters is given as

$$\log p(\boldsymbol{\theta}, \boldsymbol{\zeta}|\boldsymbol{D}) \propto \log p(\boldsymbol{\theta} = \exp(\boldsymbol{\zeta}), \boldsymbol{\gamma}|\boldsymbol{x}) + \sum_{u=1}^{2k} \zeta_u. \tag{5.19}$$

Here, θ_i for $i = 1, 2, ..., 2k$, are transformed using $\zeta_i = \log(\theta_i)$. However, θ_i for $i = 2k + 1, ..., 2k + p$, will remain the same. That is, the transformed vector of unknown parameters $\boldsymbol{\zeta}$ satisfies:

$$\boldsymbol{\zeta} = \begin{cases} u_i(\theta_i) = \log(\theta_i) & i = 1, 2, ..., 2k \\ u_i(\theta_i) = \theta_i & i = 2k + 1, ..., 2k + p \end{cases}$$

where $\theta_i = u_i^{-1}(\zeta_i)$.

Chapter 6

Data Analysis

In this chapter, we illustrate the competing risks model by analysing a real life data set of bone marrow transplant patients. The analysis will implement the techniques discussed in the previous chapters.

6.1 Bone Marrow Transplant Data

With the analysis, we will find answers to the following questions:

- What is the probability of surviving a specific risk in the presence of other risks at a given time?
- Are the failure probabilities equal among the group?
- If we assume a parametric form, then what is the best fit model for the competing risks data?
- Do the covariates have a significant effect on the cause-specific hazard?

However, before investigating the questions above, we must explore the data set. The data set consists of 137 bone marrow transplant patients collected from four different hospitals [6]. The data set has two independent causes of failures, or risks, which are the patient experiences leukemia relapse ($\delta = 1$) or the patient dies in remission ($\delta = 2$). If the failure time is censored, it is denoted by $\delta = 0$. This data set did not contain any unknown causes of failures ($\delta = *$). In this situation, if the patient relapses, then they are no longer in remission, and if the patient dies in remission, then they can no longer relapse since they have died. Thus, the occurrence of one of these risks precludes the occurrence of the other risk. The data is sorted

into three groups. The groups consists of acute lymphoblastic leukemia (ALL), acute myelocytic leukemia low risk (AML-LR) and acute myelocytic leukemia high risk (AML-HR). The groups are all forms of leukemia, which is the cancer of the body's bone marrow. Furthermore, ALL starts in cells that become lymphocytes where AML starts in early myeloid cells. Some of the notation for the following data set are $k = 2$, $N = 137$, and $g = 3$. The notation n_l is used to indicate the number of observations in the l^{th} group. Thus, $n_1 = 38$, $n_2 = 54$, and $n_3 = 45$.

Group 1										
Risk 1	54	74	104	109	110	122	129	192	230	383
	609	662								
Risk 2	1	86	107	122	172	194	276	332	418	466
	487	526								
Censored	226	530	996	1111	1167	1182	1199	1330	1377	1433
	1462	1496	1602	2081						
Group 2										
Risk 1	211	219	272	381	421	486	748			
Risk 2	10	35	48	53	79	80	105	288	390	414
	481	641	704	1063	1074	2204				
Censored	248	606	847	848	860	932	957	1030	1258	1324
	1363	1384	1447	1470	1527	1535	1562	1674	1709	1799
	1829	1843	1850	1857	1870	2218	2246	2409	2506	2569
Group 3										
Risk 1	32	47	47	48	64	76	84	93	100	113
	115	120	157	242	268	273	390	422	456	467
	625									
Risk 2	2	16	63	74	80	105	162	164	183	318
	363	677								
Censored	845	1136	1238	1345	1631	2024	2133	2140	2252	2430
	2640									

Table 6.1: The survival times of the bone marrow patients.

Group	**Censored**	**Cause 1**	**Cause 2**	Total
ALL	14	12	12	38
AML-LR	31	7	16	54
AML-HR	11	21	13	45
Total	56	40	41	137

Table 6.2: The number of events for each status by their respective group.

Group	Censored	Risk 1	Risk 2
ALL	1228.00	253.58	265.58
AML-LR	1520.81	391.14	479.31
AML-HR	1801.27	201.86	182.69

Table 6.3: Mean follow-up in days for each status by their respective group.

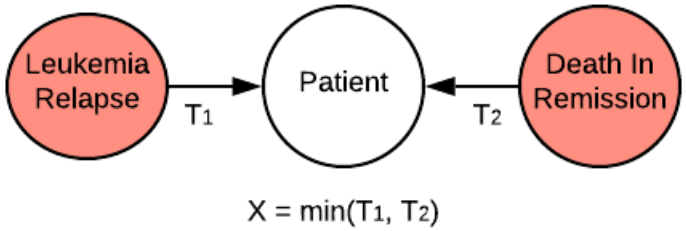


Figure 6.1: A bone marrow transplant patient experiencing two competing risks.

6.2 Non-Parametric Estimation

In this section, we model the competing risks data using only the observational data. The following will be done by using the Kaplan-Meier method, as well as using the empirical cumulative incidence function (CIF). These techniques will provide us insight on the failure probability of each cause over a range of time.

6.2.1 Kaplan-Meier Estimator

The KM for the BMT data shows the proportion of individuals who have yet to experience one of the two causes of death. Inversely using the KM (1−KM), we can calculate the proportion of patients who have experienced one of the risks over time. To use the KM estimator to model the BMT data, we must treat the observations of one of the two causes of death as censored observations. For example, if relapse is the primary concern, then we find the proportion of patients relapsing by treating remission as censored observations.

In Figure 6.2, the 1−KM curves in the top plot indicates that patients with AML-HR have the highest probability of experiencing relapse, where patients with

AML-LR have the lowest probability of experiencing relapse. Therefore, we find that patients with AML-LR have the best odds to not experience relapse during the entire duration of the study. The $1-$KM curves for remission in the bottom plot are more tightly packed compared to the relapse curves. We find patients with AML-LR have the lowest probability of experiencing remission for most of the duration. The probability for remission after 500 days is approximately equal for patients with ALL and AML-HR.

Overall, these probabilities are estimated by assuming the observations of the secondary risk are right-censored. If we consider relapse as the primary interest, then the complement of the KM provides the probability of relapse in a hypothetical world where it is impossible to die in remission, and rarely useful in a clinical setting [6, p. 127]. To account for the competing risk without treating it as censored observations, we examine the estimator for the cumulative incidence function (CIF).

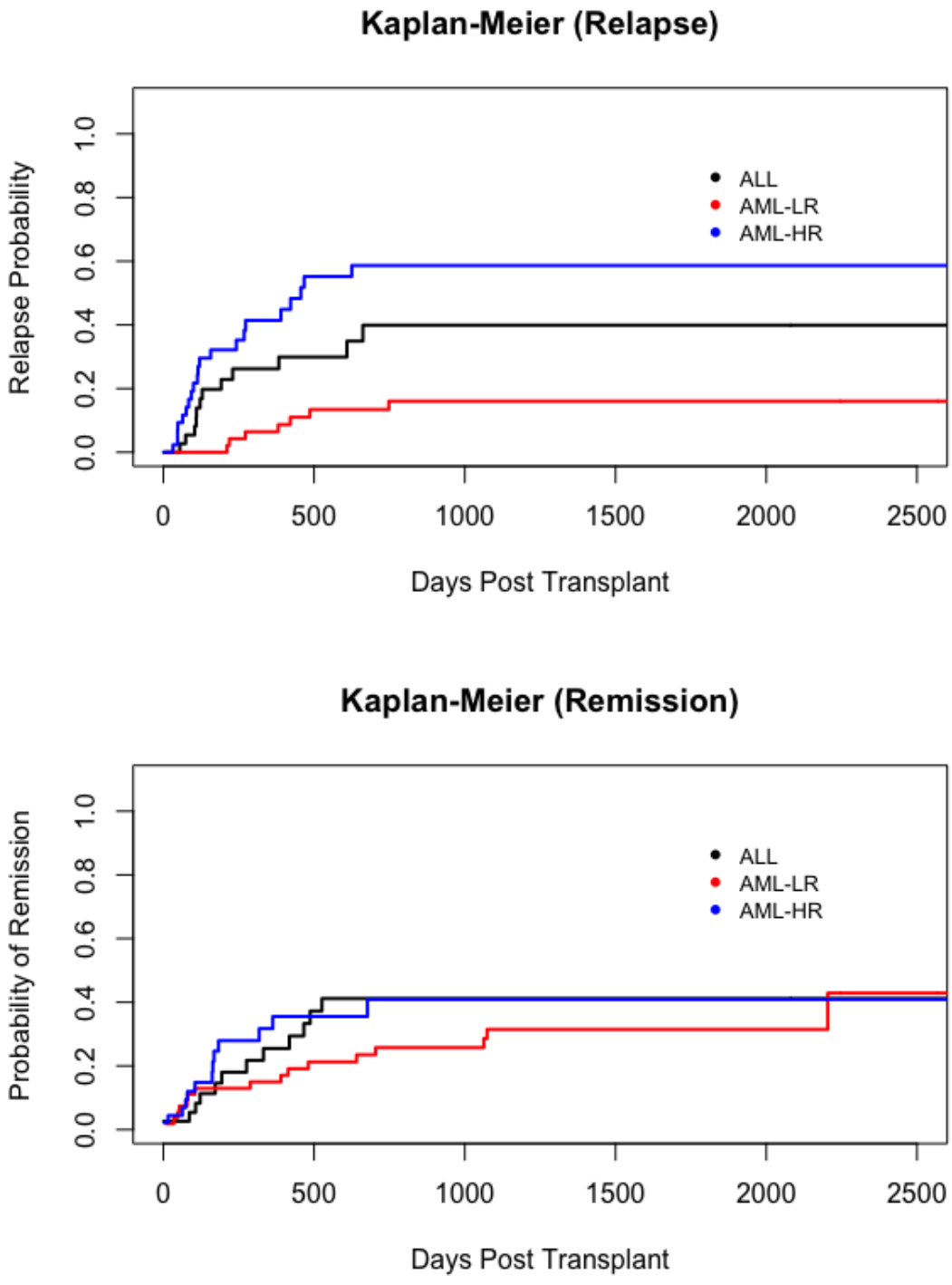


Figure 6.2: The 1−KM curves for relapse (top) and death in remission (bottom).

6.2.2 Cumulative Incidence Function

At any time x, the CIF calculates the probability of risk j in the presence of all other risks. In this technique, the secondary risks are not considered to be censored. Therefore, the CIF is the more appropriate choice compared to the KM method. In Figure 6.3, 6.4, and 6.5, stacked probabilities are shown, in which probabilities of being event-free, relapsing, and remission can be determined. For example, the height from the base to the first curve is the probability of relapsing, the height between the first and second curve is the probability of death in remission, and the height between the second curve and horizontal line at 1 is the probability of being event-free. To demonstrate this idea, we determine the probability of a specific risk and being event-free at 500 days for each group. In Figure 6.3, the probability of relapsing at 500 days after transplant is approximately 0.2654, the probability of dying in remission is approximately 0.2952, and the probability of being event-free is approximately 0.4394. In Figure 6.4, the probability of relapsing at 500 days after transplant is approximately 0.0938, the probability of dying in remission is approximately 0.2054, and the probability of being event-free is approximately 0.7008. In Figure 6.5, the probability of relapsing at 500 days after transplant is approximately 0.4444, the probability of dying in remission is approximately 0.2667, and the probability of being event-free is approximately 0.2889. Thus, death in remission and relapsing are approximately likely at 500 days post transplant for ALL patients. However, death in remission is a higher risk than relapsing for patients with AML-LR, where relapsing is a higher risk than death in remission for patients with AML-HR. Patients with AML-LR also have the greatest survival probability after transplant compared to patients with ALL and AML-HR.

In the previous section, we estimated the probabilities for each risk using the KM method. To explore the difference between estimation in the two methods, we plot the KM curves with the CIF curves in Figure 6.6. As expected, the figure shows higher failure probabilities when using the complement of the KM versus the CIF.

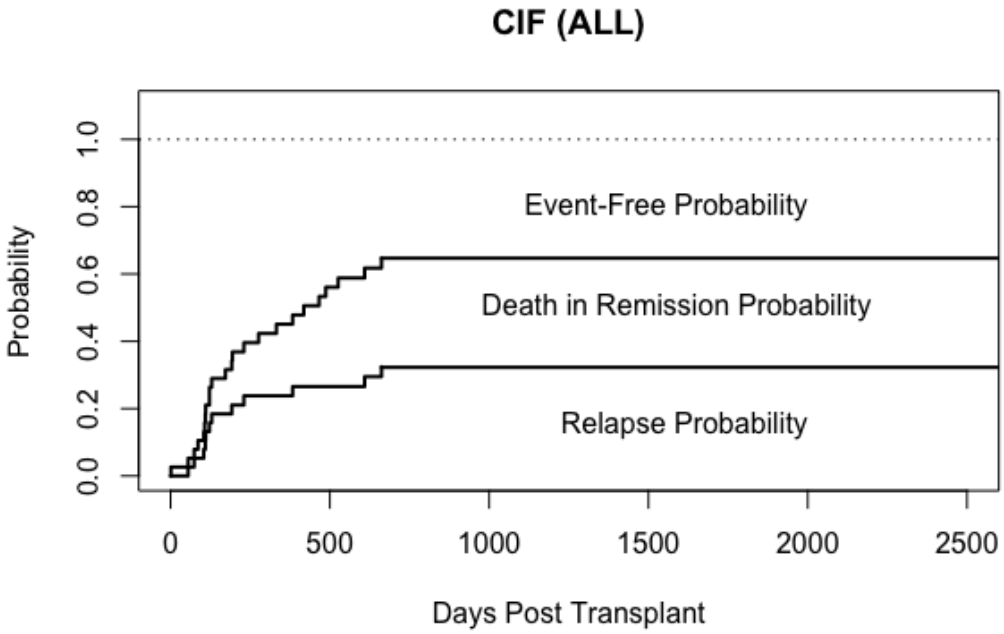


Figure 6.3: The CIF for patients with ALL (group 1).

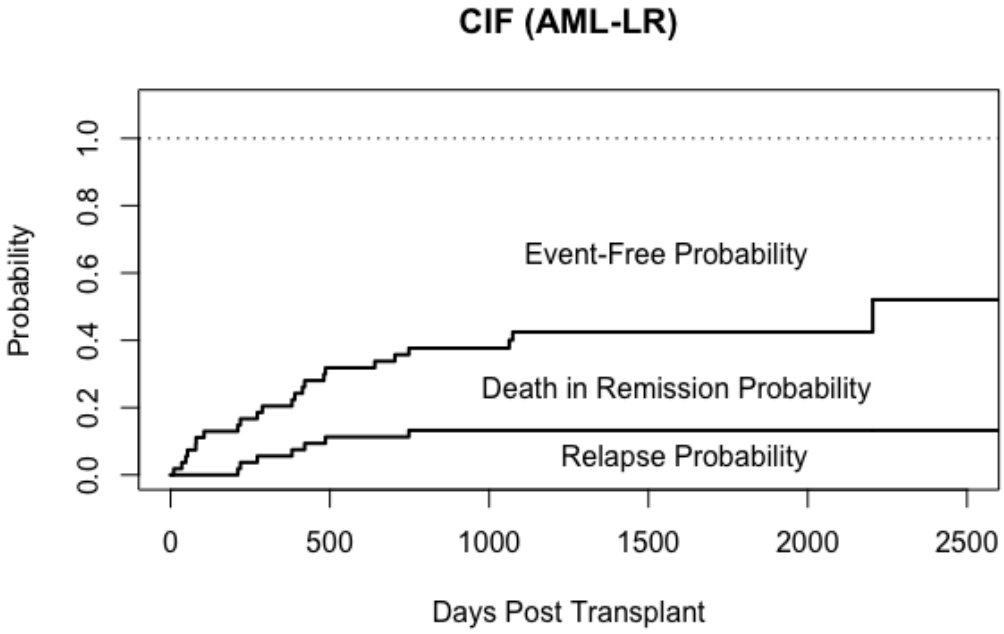


Figure 6.4: The CIF for patients with AML-LR (group 2).

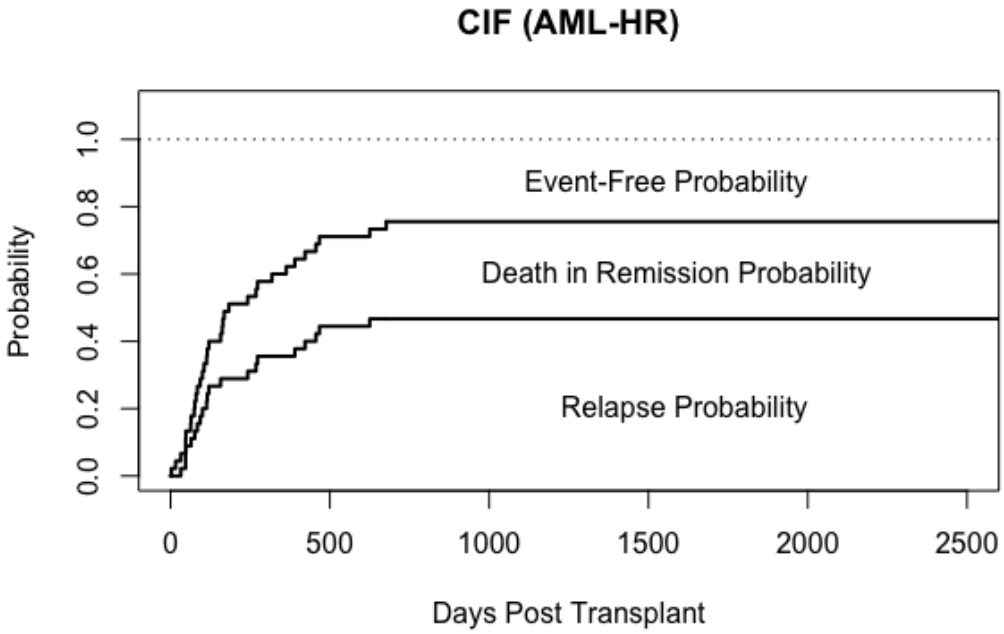


Figure 6.5: The CIF for patients with AML-HR (group 3).

In Figure 6.6, and as we mentioned in the previous section, we observe a difference in relapse probability among the groups, and almost equal death in remission probability among the groups. We can see that the results from the CIF are consistent with this idea. To formally investigate whether the probability of each cause are equal among the groups, a hypothesis test using the CIF is performed.

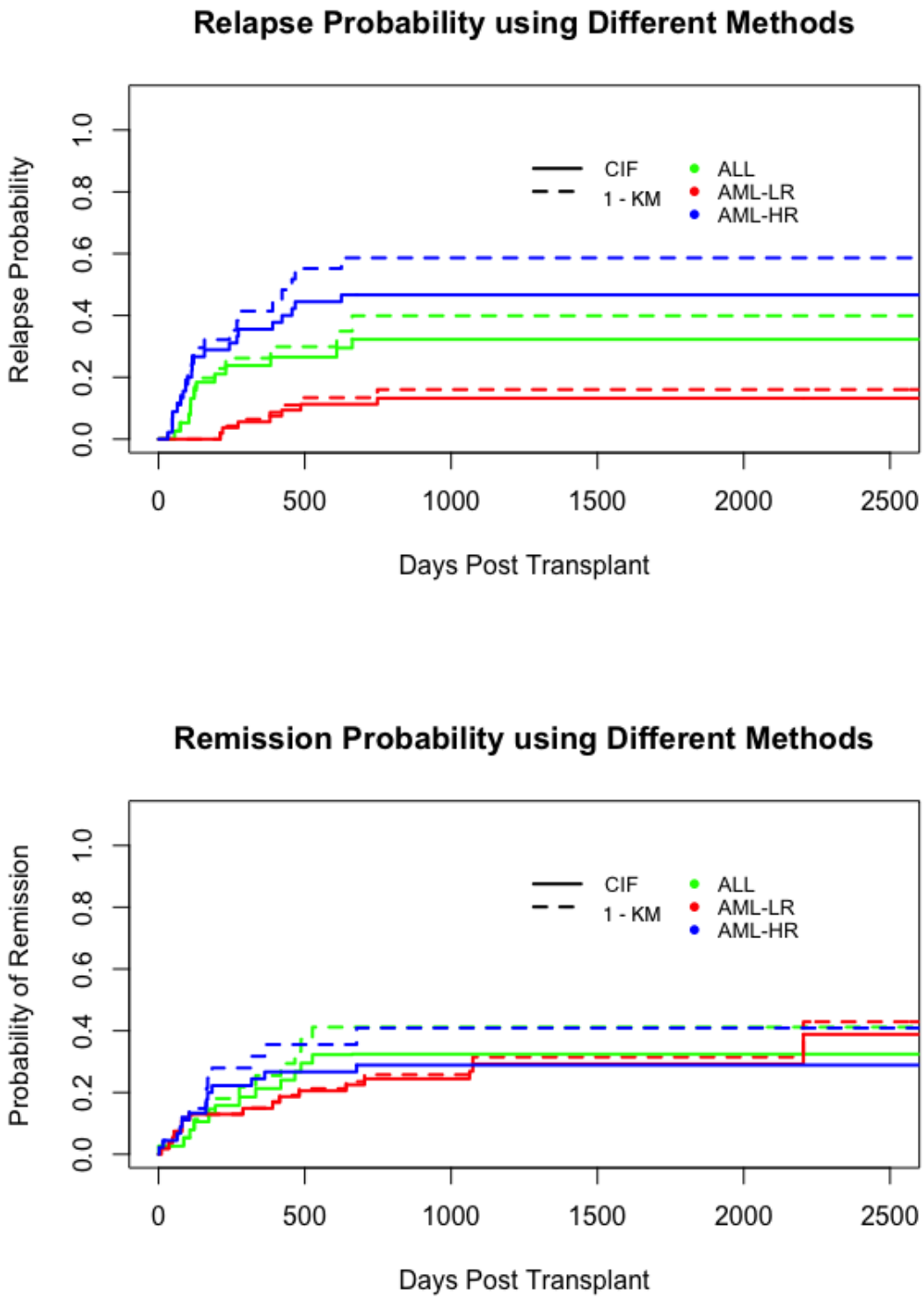


Figure 6.6: A comparison between 1−KM and CIF for the three groups, where the top figure is for relapse and the bottom figure is for remission.

6.2.3 Comparison of CIF Among the Groups

To determine whether there is a difference between the probability of a risk occurring among the groups, we conduct a hypothesis test using the CIF. Let F_j^g represent the CIF of risk j for group g, for $j = 1, 2$ and $g = 1, 2, 3$. Thus, we have the following hypotheses:

- $H_0 : F_1^1 = F_1^2 = F_1^3$ $\quad H_a$: At least one F_1^g differs for risk 1 (relapse)
- $H_0 : F_2^1 = F_2^2 = F_2^3$ $\quad H_a$: At least one F_2^g differs for risk 2 (death in remission)

To test this, we implement Gray's Test in R, and present the results in Table 6.4.

	Test Statistic	**DF**	**P-Value**
Risk 1	14.716	2	0.0006
Risk 2	0.122	2	0.9411

Table 6.4: Gray's Test for testing equality among groups.

The results from Gray's Test shows statistically significant results for risk 1, and can reject the null hypothesis that the CIF are equal for the groups at any significance level greater than 0.0006, and conclude in favour for the alternative that there is a difference among the CIF for relapse. However, the test did not show significant results for cause 2 and thus we fail to reject the null hypothesis and conclude that we cannot reject the claim that the CIF are equal among the groups. Thus, Gray's test is consistent with Figure 6.6 as we observe the curves separated in the relapse plot, and tightly together in the remission plot.

6.3 Parametric Estimation

In this section, we use maximum likelihood methods and Bayesian methods to estimate the unknown parameters for the competing risks model using the BMT data. The MLEs and Bayes estimates are then used to calculate some of the reliability measures.

6.3.1 Model Selection using the Behaviour of the Hazard

To help with our model selection, we first analyse the behaviour of the hazard function for each group to determine which distribution can take on this behaviour. To do this, we plot the total time on test (TTT) for each group and risk (Figure 6.7). In this figure, we characterize each plot by abbreviating the increasing-decreasing pattern in the plot. For example, IDI would indicate an increasing-decreasing-increasing hazard function. The top-left plot would be characterized as IDI, the top-right and middle-left plot is strictly increasing, the middle-right is strictly decreasing, the bottom-left is IDI, and the bottom-right is DID. Based on these characteristics, the following distributions may be appropriate to model the causes for each group:

1. Group 1 (ALL)
    (a) Lindley (Remission)
    (b) STH-1 (Remission)
    (c) Weibull (Remission)
    (d) Bathtub (Remission)
    (e) Power Lindley (Remission)
    (f) STH-2 (Relapse and Remission)
2. Group 2 (AML-LR)
    (a) Lindley (Relapse)

(b) Bathtub (Relapse)

(c) STH-1 (Relapse and Remission)

(d) STH-2 (Relapse and Remission)

(e) Power Lindley (Relapse and Remission)

(f) Weibull (Relapse and Remission)

3. Group 3 (AML-HR)

(a) STH-2 (Relapse)

(b) Power Lindley (Remission)

The only model that did not appear to be appropriate for any causes in any group was the exponential distribution as it can only take on a constant hazard rate.

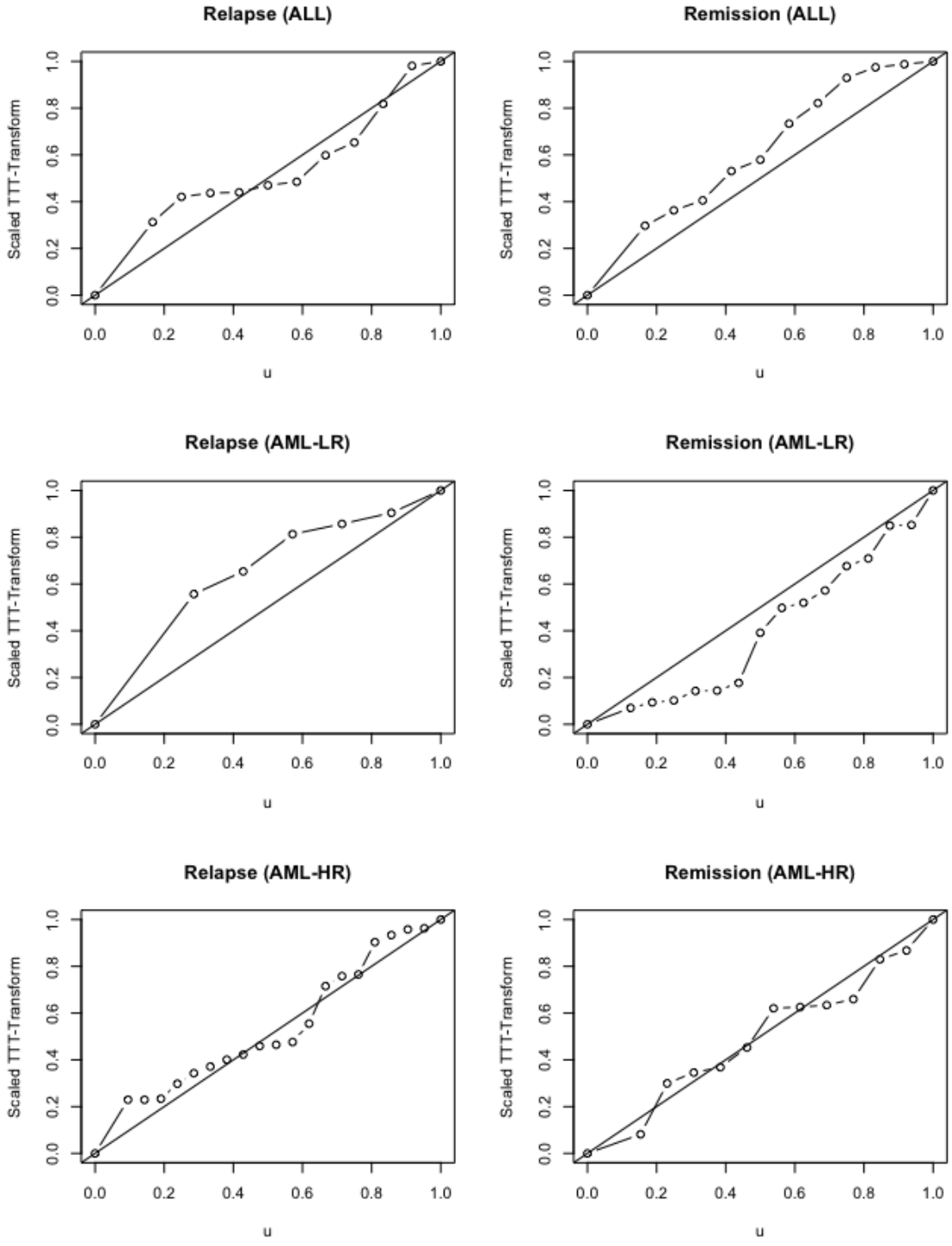


Figure 6.7: The total time on test (TTT) for cause 1 and cause 2 of all three groups.

6.3.2 Maximum Likelihood Estimates

In Table 6.6, we present the MLEs for the lifetime distributions. To help with our model selection, we calculate the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for each model, defined as

$$\text{AIC} = -2\mathcal{L} + 2p \text{ and BIC} = -2\mathcal{L} + p\ln(n),$$

where $\mathcal{L}$ is the value of the log-likelihood evaluated at the MLEs, p is the number of parameters, and n is the number of observations. When comparing AIC values, the model with the smallest AIC value is the model that minimizes the information lost, and thus the best choice candidate model to represent the true model. For BIC values, we also look for the smallest value when comparing models. For group 1 (ALL), the AIC and BIC values support that the Power Lindley is the best choice model. For group 2 (AML-LR), the AIC and BIC values support that the Weibull is the best choice model. For group 3 (AML-HR), the AIC and BIC values support that the Power Lindley is the best choice model. However, since the difference between the AIC and BIC in the top three models are small in each group, any of the three model could be deemed well suited to model the data. In Figure 6.8, 6.9, and 6.10, the CIF for every model is shown for the respective group in the top plot, with the three best suited model in the bottom plot. The CIF was constructed using the invariant property of maximum likelihood . The top three models are shown in Table 6.5. Both group 1 and 3 had the same models in the same ranking order, where group 2 had a slightly different order with one different model.

Model	**Group 1**	**Group 2**	**Group 3**
1^{st}	Power Lindley	Weibull	Power Lindley
2^{nd}	Weibull	STH-2	Weibull
3^{rd}	STH-2	Power Lindley	STH-2

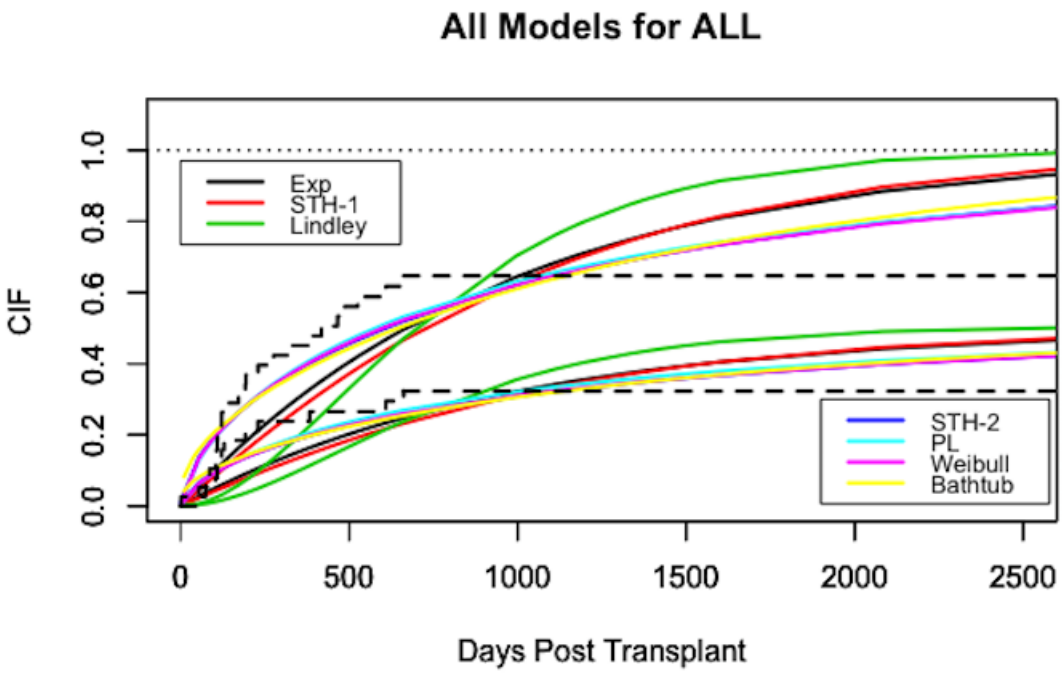
Table 6.5: The model ranks for each group, where the parameters of the models were estimated using maximum likelihood estimation.

Model	MLE		$\mathcal{L}$	AIC	BIC
	Cause 1	Cause 2			
Group 1					
Exp	$\alpha_1 = 5.180 \times 10^{-4}$	$\alpha_2 = 5.182 \times 10^{-4}$	-205.565	415.129	417.486
Lindley	$\alpha_1 = 1.553 \times 10^{-3}$	$\alpha_2 = 1.533 \times 10^{-3}$	-223.323	450.660	453.002
STH-1	$\alpha_1 = 4.233 \times 10^{-4}$	$\alpha_2 = 4.269 \times 10^{-4}$	-207.395	418.789	421.146
PL	$\alpha_1 = 0.4467$	$\alpha_2 = 0.4285$	-201.859	411.719	416.430
	$\beta_1 = 0.0602$	$\beta_2 = 0.0659$			
STH-2	$\alpha_1 = 0.6171$	$\alpha_2 = 0.5933$	-202.519	413.038	417.750
	$\beta_1 = 5.850 \times 10^{-3}$	$\beta_2 = 6.800 \times 10^{-3}$			
Weibull	$\alpha_1 = 4.609 \times 10^{-3}$	$\alpha_2 = 5.689 \times 10^{-3}$	-202.162	412.323	417.036
	$\beta_1 = 0.6748$	$\beta_2 = 0.6435$			
Bathtub	$\alpha_1 = 0.0119$	$\alpha_2 = 0.0111$	-204.689	417.379	422.090
	$\beta_1 = 0.1894$	$\beta_2 = 0.1925$			
Group 2					
Exp	$\alpha_1 = 1.216 \times 10^{-4}$	$\alpha_2 = 2.780 \times 10^{-4}$	-217.107	438.215	440.485
Lindley	$\alpha_1 = 5.178 \times 10^{-4}$	$\alpha_2 = 8.827 \times 10^{-4}$	-236.162	476.324	478.595
STH-1	$\alpha_1 = 1.101 \times 10^{-4}$	$\alpha_2 = 2.343 \times 10^{-4}$	-218.208	440.416	442.687
PL	$\alpha_1 = 0.3170$	$\alpha_2 = 0.4190$	-214.240	436.480	441.022
	$\beta_1 = 0.0633$	$\beta_2 = 0.0531$			
STH-2	$\alpha_1 = 0.8330$	$\alpha_2 = 0.5067$	-213.644	435.287	439.830
	$\beta_1 = 3.705 \times 10^{-4}$	$\beta_2 = 8.373 \times 10^{-3}$			
Weibull	$\alpha_1 = 1.469 \times 10^{-3}$	$\alpha_2 = 4.119 \times 10^{-3}$	-213.473	434.946	439.488
	$\beta_1 = 0.6532$	$\beta_2 = 0.6243$			
Bathtub	$\alpha_1 = 2.837 \times 10^{-3}$	$\alpha_2 = 9.947 \times -3$	-215.392	438.785	443.326
	$\beta_1 = 0.1939$	$\beta_2 = 0.1785$			
Group 3					
Exp	$\alpha_1 = 7.942 \times 10^{-4}$	$\alpha_2 = 4.920 \times 10^{-4}$	-282.915	569.830	572.883
Lindley	$\alpha_1 = 2.020 \times 10^{-3}$	$\alpha_2 = 1.435 \times 10^{-3}$	-319.017	642.039	645.087
STH-1	$\alpha_1 = 5.937 \times 10^{-4}$	$\alpha_2 = 3.838 \times 10^{-4}$	-287.692	579.383	582.437
PL	$\alpha_1 = 0.4288$	$\alpha_2 = 0.3650$	-271.890	547.780	557.885
	$\beta_1 = 0.0912$	$\beta_2 = 0.0986$			
STH-2	$\alpha_1 = 0.5501$	$\alpha_2 = 0.4896$	-273.157	550.319	560.419
	$\beta_1 = 0.0137$	$\beta_2 = 0.0135$			
Weibull	$\alpha_1 = 0.0120$	$\alpha_2 = 0.0122$	-272.363	548.725	558.831
	$\beta_1 = 0.6041$	$\beta_2 = 0.5283$			
Bathtub	$\alpha_1 = 0.0252$	$\alpha_2 = 0.0186$	-277.583	559.167	569.271
	$\beta_1 = 0.1770$	$\beta_2 = 0.1691$			

Table 6.6: The MLEs estimate for cause 1 and cause 2 for each group. The value of log-likelihood evaluated at the MLEs is given, as well as the AIC and BIC values.

Model	Par.	95% CI	Par.	95% CI
Group 1				
Exp	α_1	$[4.643 \times 10^{-4}, 5.772 \times 10^{-4}]$	α_2	$[4.645 \times 10^{-4}, 5.774 \times 10^{-4}]$
Lindley	α_1	$[1.462 \times 10^{-3}, 1.650 \times 10^{-3}]$	α_2	$[1.422 \times 10^{-3}, 1.651 \times 10^{-3}]$
STH-1	α_1	$[3.819 \times 10^{-3}, 4.675 \times 10^{-4}]$	α_2	$[3.585 \times 10^{-3}, 4.713 \times 10^{-4}]$
PL	α_1	$[0.4153, 0.4806]$	α_2	$[0.4009, 0.4581]$
	β_1	$[0.0489, 0.0732]$	β_2	$[0.0550, 0.0791]$
STH-2	α_1	$[0.5817, 0.6547]$	α_2	$[0.5584, 0.6346]$
	β_1	$[4.572 \times 10^{-3}, 7.486 \times 10^{-3}]$	β_2	$[5.255 \times 10^{-3}, 8.804 \times 10^{-3}]$
Weibull	α_1	$[3.698 \times 10^{-3}, 5.745 \times 10^{-3}]$	α_2	$[4.430 \times 10^{-3}, 7.298 \times 10^{-3}]$
	β_1	$[0.6416, 0.7097]$	β_2	$[0.6068, 0.6824]$
Bathtub	α_1	$[9.751 \times 10^{-3}, 0.0146]$	α_2	$[9.007 \times 10^{-3}, 0.0136]$
	β_1	$[0.1821, 0.1970]$	β_2	$[0.1852, 0.2000]$
Group 2				
Exp	α_1	$[1.023 \times 10^{-4}, 1.448 \times 10^{-4}]$	α_2	$[2.540 \times 10^{-4}, 3.036 \times 10^{-4}]$
Lindley	α_1	$[4.835 \times 10^{-4}, 5.533 \times 10^{-4}]$	α_2	$[8.432 \times 10^{-4}, 9.227 \times 10^{-4}]$
STH-1	α_1	$[9.420 \times 10^{-5}, 1.295 \times 10^{-4}]$	α_2	$[2.161 \times 10^{-4}, 2.547 \times 10^{-4}]$
PL	α_1	$[0.2904, 0.3461]$	α_2	$[0.3986, 0.4405]$
	β_1	$[0.0524, 0.0764]$	β_2	$[0.0460, 0.0614]$
STH-2	α_1	$[0.8018, 0.8654]$	α_2	$[0.4807, 0.5341]$
	β_1	$[2.999 \times 10^{-4}, 4.576 \times 10^{-4}]$	β_2	$[5.459 \times 10^{-3}, 0.0113]$
Weibull	α_1	$[8.280 \times 10^{-4}, 2.609 \times 10^{-3}]$	α_2	$[3.356 \times 10^{-3}, 5.055 \times 10^{-3}]$
	β_1	$[0.6271, 0.6804]$	β_2	$[0.5982, 0.6515]$
Bathtub	α_1	$[2.469 \times 10^{-3}, 3.258 \times 10^{-3}]$	α_2	$[8.524 \times 10^{-3}, 0.0116]$
	β_1	$[0.1889, 0.1991]$	β_2	$[0.1732, 0.1841]$
Group 3				
Exp	α_1	$[7.393 \times 10^{-4}, 8.530 \times 10^{-4}]$	α_2	$[4.446 \times 10^{-4}, 5.413 \times 10^{-4}]$
Lindley	α_1	$[1.934 \times 10^{-3}, 2.109 \times 10^{-3}]$	α_2	$[1.358 \times 10^{-3}, 1.514 \times 10^{-3}]$
STH-1	α_1	$[5.547 \times 10^{-4}, 6.344 \times 10^{-4}]$	α_2	$[3.499 \times 10^{-4}, 4.201 \times 10^{-4}]$
PL	α_1	$[0.4105, 0.4482]$	α_2	$[0.3445, 0.3866]$
	β_1	$[0.0809, 0.1028]$	β_2	$[0.0866, 0.1124]$
STH-2	α_1	$[0.5270, 0.5742]$	α_2	$[0.4594, 0.5218]$
	β_1	$[0.0116, 0.0160]$	β_2	$[0.0109, 0.0168]$
Weibull	α_1	$[9.811 \times 10^{-3}, 0.0146]$	α_2	$[9.659 \times 10^{-3}, 0.0153]$
	β_1	$[0.5758, 0.6338]$	β_2	$[0.4956, 0.5632]$
Bathtub	α_1	$[0.02220.0286]$	α_2	$[0.0159, 0.0218]$
	β_1	$[0.1722, 0.1819]$	β_2	$[0.1628, 0.1755]$

Table 6.7: The asymptotic 95% CI for the true model parameters.



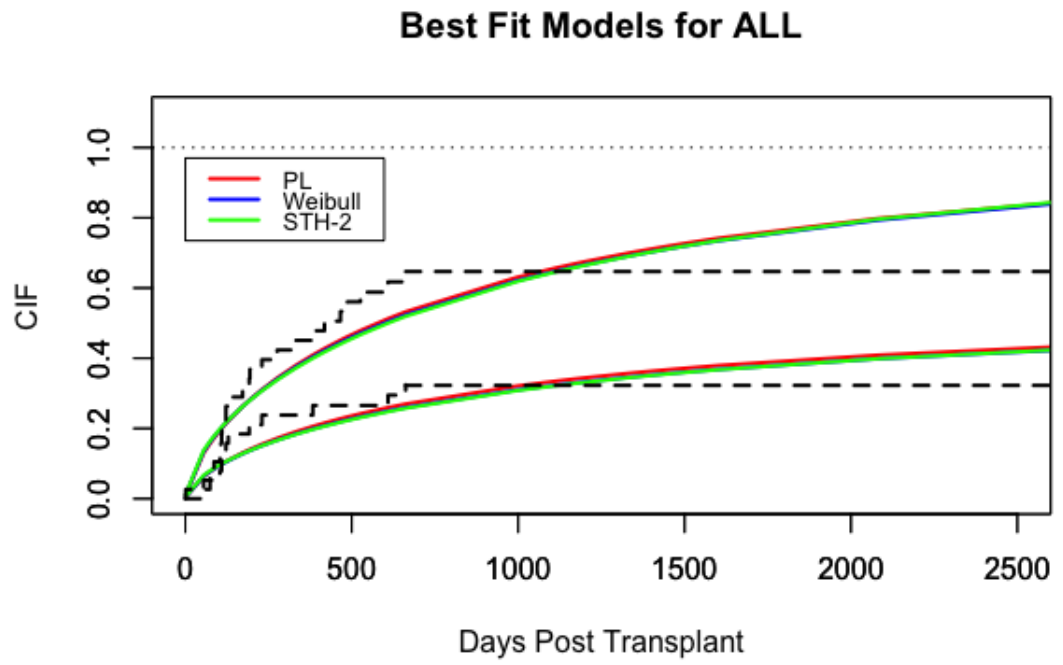


Figure 6.8: The CIF (using MLE) for all models (top) and the top three models for ALL.

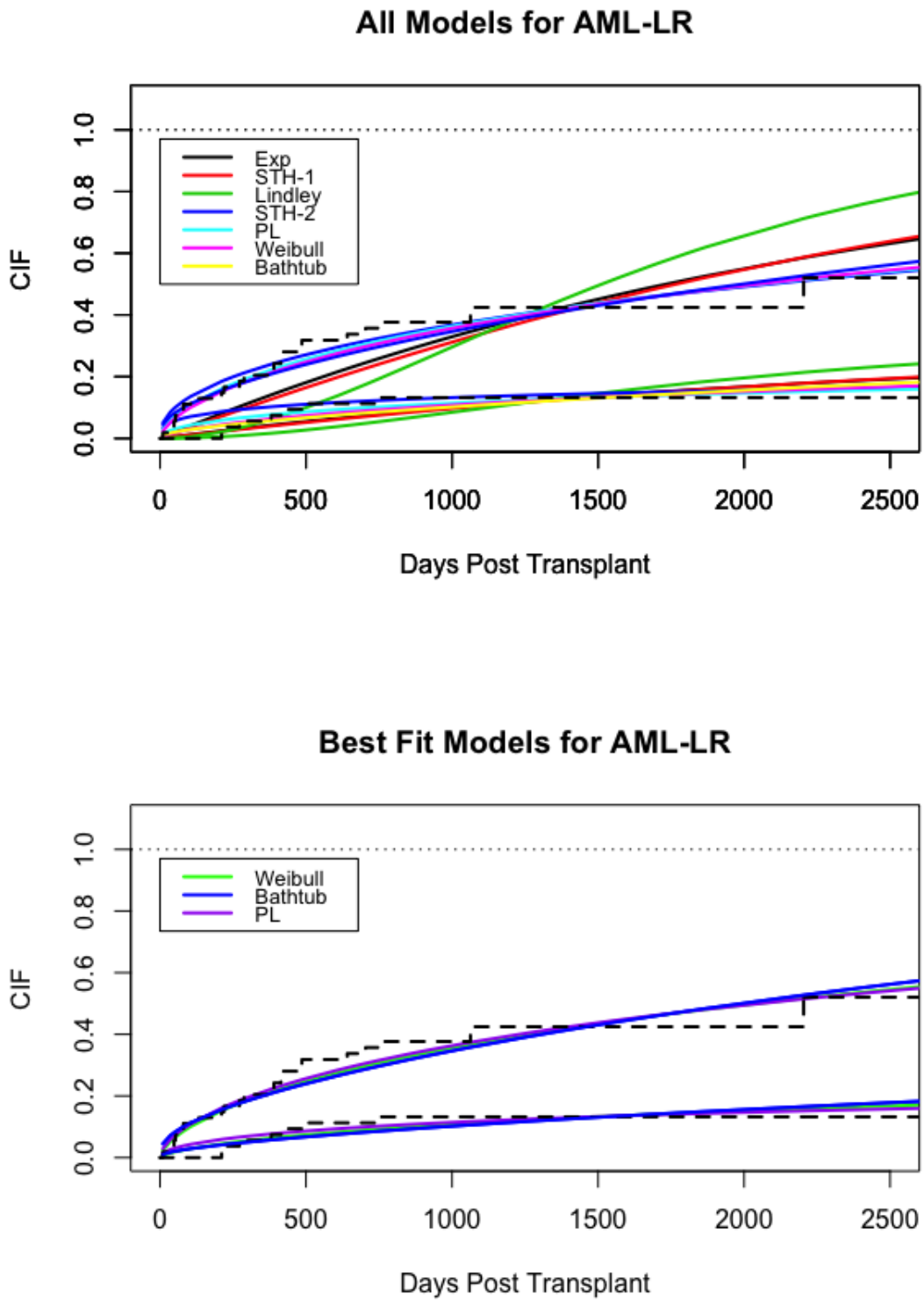


Figure 6.9: The CIF (using MLE) for all models (top) and the top three models for AML-LR.

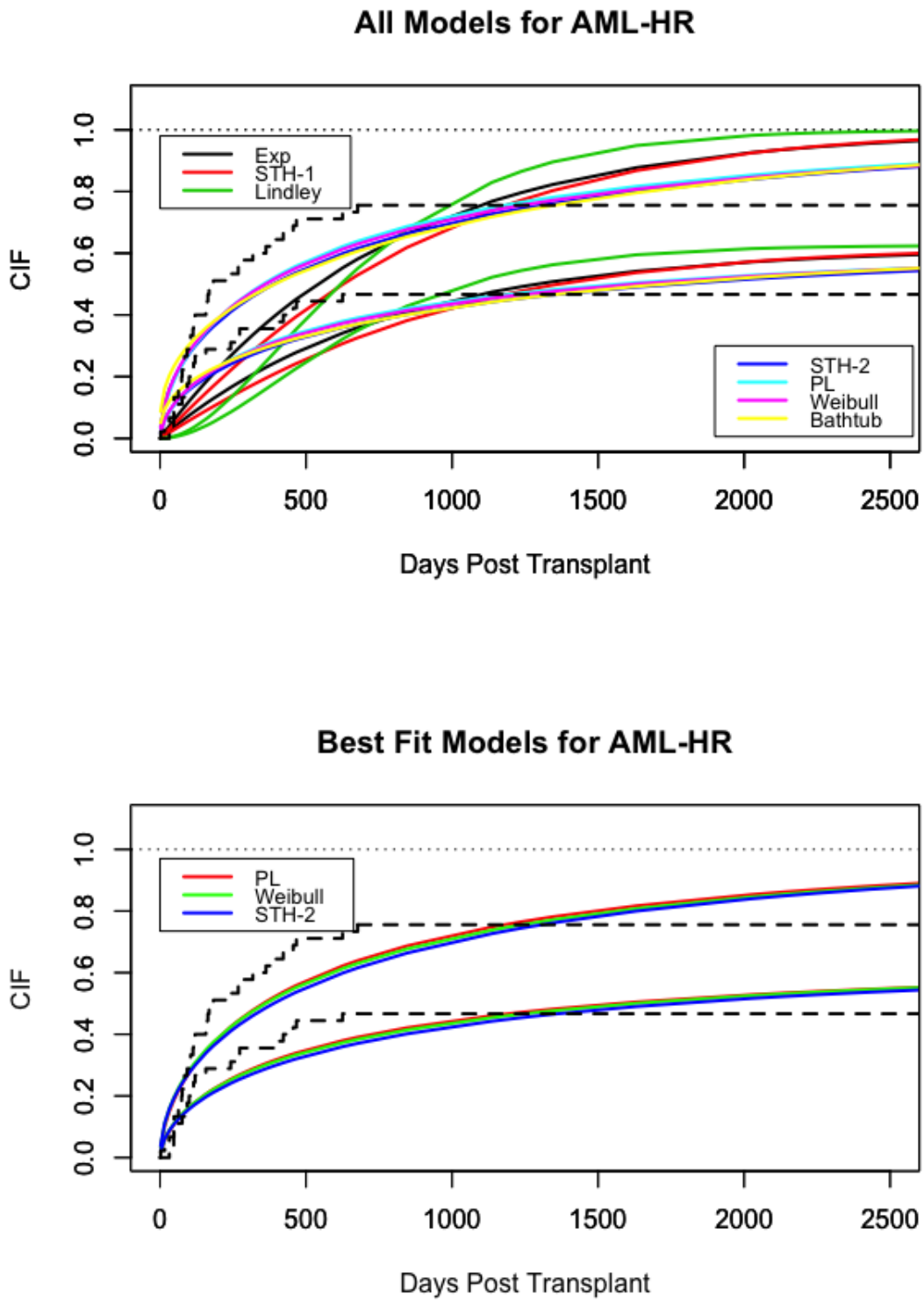


Figure 6.10: The CIF (using MLE) for all models (top) and the top three models for AML-HR.

The expected failure time at time 0 for each risk among the groups is given in Table 6.8. The expected failure time at time 0 gives the mean time of a specific risk occurring to a patient. For example, consider a patient with ALL. Based on the risks model following a Power Lindley distribution, then the average time for leukemia relapse and death in remission at time 0 are approximately 4219 days and 5049 days, respectively. As expected, patients with AML-LR have a much larger expected failure time for leukemia relapse than patients in the other disease groups. Based on the relative risk for relapse in group 2 (Table 6.9), these patients had the lowest relapse probability compared to patients with ALL and ALL-HR, and therefore would be expected to experience relapse at a much later time.

Model	$E(X)$	
Group 1	**Risk 1**	**Risk 2**
Exp	1930.56	1929.72
Lindley	1286.99	1303.93
STH-1	591.55	586.64
PL	4218.61	5048.75
STH-2	3564.09	3839.26
Weibull	3805.72	4261.47
Bathtub	-	-
Group 2	**Risk 1**	**Risk 2**
Exp	8220.87	3597.51
Lindley	3861.40	2264.86
STH-1	2272.32	1068.08
PL	175328.50	10570.19
STH-2	10091.8	12509.43
Weibull	29508.60	9476.36
Bathtub	-	-
Group 3	**Risk 1**	**Risk 2**
Exp	1259.12	2032.42
Lindley	989.11	1393.09
STH-1	422.10	652.42
PL	2311.42	8697.98
STH-2	2286.46	6785.79
Weibull	2264.70	7650.69
Bathtub	-	-

Table 6.8: The expected failure time in days.

Dist.	**Group 1**		**Group 2**		**Group 3**	
	π_1	π_2	π_1	π_2	π_1	π_2
Exp	0.500	0.500	0.304	0.696	0.617	0.383
STH-1	0.497	0.502	0.300	0.700	0.620	0.380
Lindley	0.505	0.495	0.309	0.691	0.626	0.374
STH-2	0.504	0.496	0.215	0.785	0.641	0.359
PL	0.515	0.484	0.257	0.743	0.624	0.376
Weibull	0.507	0.493	0.316	0.684	0.630	0.370
Bathtub	0.494	0.506	0.356	0.644	0.627	0.373

Table 6.9: The estimated relative risk for each distribution.

The relative risk for each risk in the three groups using each distribution is given in Table 6.9. In terms of the BMT data, the relative risk calculates the probability of relapsing or death in remission if the individual's age could hypothetically approach infinity. For group 1, both risk 1 and risk 2 have approximately the same probability. In group 2, the probability of risk 2 is much greater than the probability of risk 1. For group 3, the probability of risk 1 is much greater than the probability of risk 2.

The Kolmogorov-Smirnov (KS) statistics are shown in Table 6.10. The KS statistics are used to quantify the distance between two curves (parametric and non-parametric). For a particular model, we calculate the distance between the CIF using the parameter estimates from the maximum likelihood method with the empirical CIF for a set of time points, and the maximum distance is recorded. The bold value represents the minimum value for the respective group and cause, or simply indicates the best fit model among the list of models used. For group 1, we find PL had the smallest KS compared to all other distributions for risk 1, where Weibull had the smallest KS for risk 2. In group 2, Weibull had the smallest KS for risk 1 and Bathtub had the smallest KS for risk 2. For group 3, PL had the smallest KS for risk 1, and PL and Weibull both had the same smallest KS for risk 2. The results from the KS statistics shows some support of the conclusions drawn from the AIC/BIC scores. These results may also indicate that each cause may need to be modelled using different distributions. For example, modelling ALL patients where risk 1 follows PL distribution and risk 2 follows Weibull distribution.

Model	**Group 1**		**Group 2**		**Group 3**	
	Risk 1	Risk 2	Risk 1	Risk 2	Risk 1	Risk 2
Exp	0.1320	0.1125	0.05906	0.1010	0.1784	0.1420
STH-1	0.1460	0.1279	0.06239	0.1052	0.2044	0.1560
Lindley	0.1884	0.1479	0.0868	0.1378	0.2518	0.1938
STH-2	0.0873	0.0868	0.1170	0.0766	0.1239	0.0706
PL	**0.0828**	0.0858	0.0787	0.0620	**0.1072**	**0.0686**
Weibull	0.0871	**0.0847**	**0.0400**	0.0582	0.1110	**0.0686**
Bathtub	0.0832	0.0975	0.0468	**0.0457**	0.1208	0.0750

Table 6.10: The Kolmogorov-Smirnov (KS) statistics are shown for each risk among groups. Bold values denote the smallest value.

6.3.3 Bayesian Estimation

To determine the Bayes estimate, we used an MCMC algorithm (Metropolis-Hastings) to simulate the log-posterior distribution to obtain draws from it. The algorithm is used twice. First, a Gamma prior is used with all hyper-parameters set equal to 0.001, in which we refer to this as a poor prior. Using the mean and variance of the draws, the hyper-parameters a and b are solved for from each parameter and used as the hyper-parameters for the prior (Table 6.14) in a second iteration of MCMC, in which we refer to this as a good prior. The Bayes estimate using a poor prior and good prior, along with the variance, can be seen in Table 6.12 and 6.13, respectively. The Bayes estimate in both iterations are consistent with one another, however, as expected, the variance of the second iteration is much smaller. The Bayes estimates for each distribution are also consistent with the results of the maximum likelihood estimates. For example, the estimates for modelling group 1 with the Power Lindley distribution using maximum likelihood estimation and Bayesian estimation, respectively, were

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = (0.4467, 0.0602, 0.4285, 0.0659)$$

$$\hat{\boldsymbol{\theta}}_{\text{Bayes}} = (0.4986, 0.0433, 0.4726, 0.0504)$$

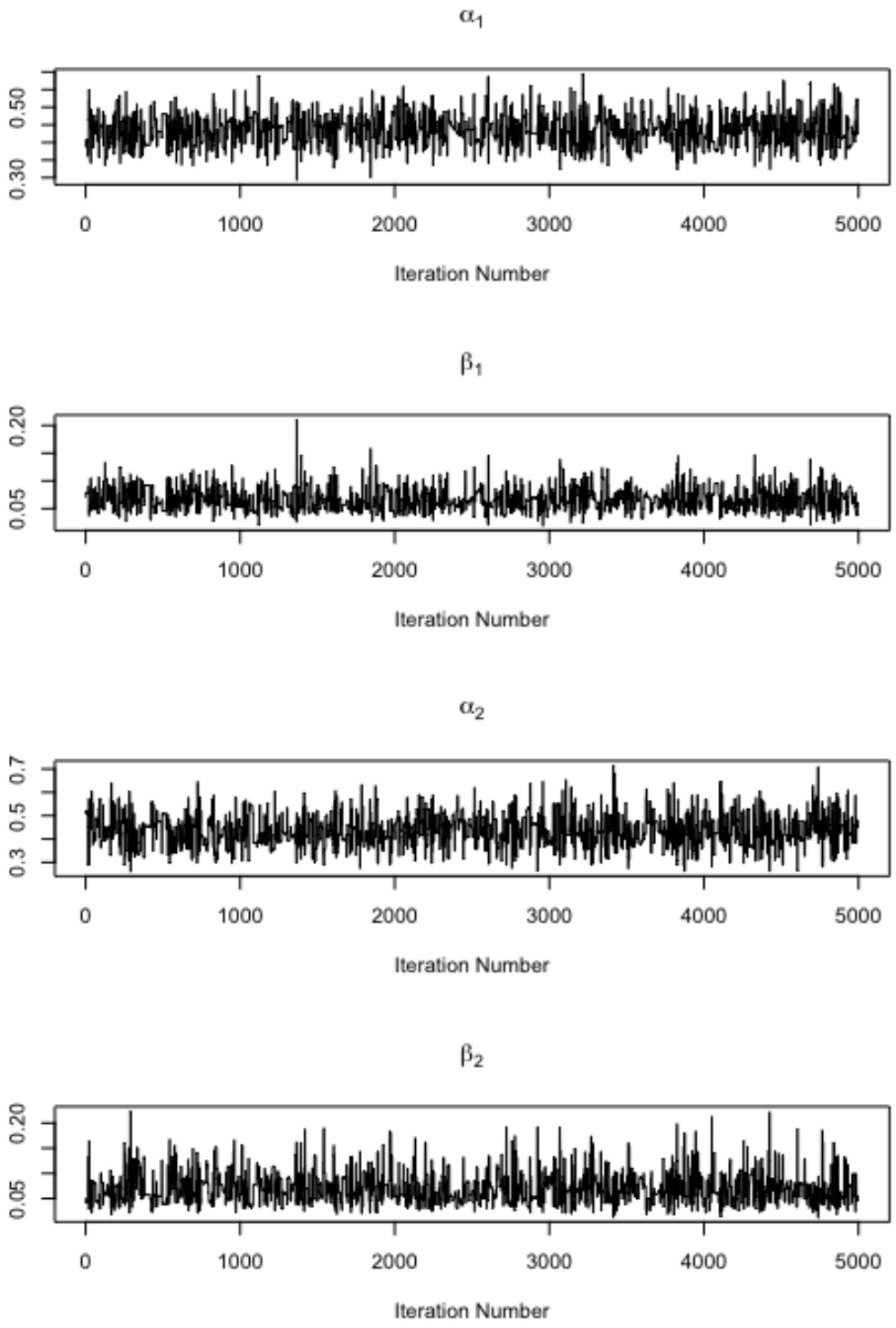


Figure 6.11: The trace plots for the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1).

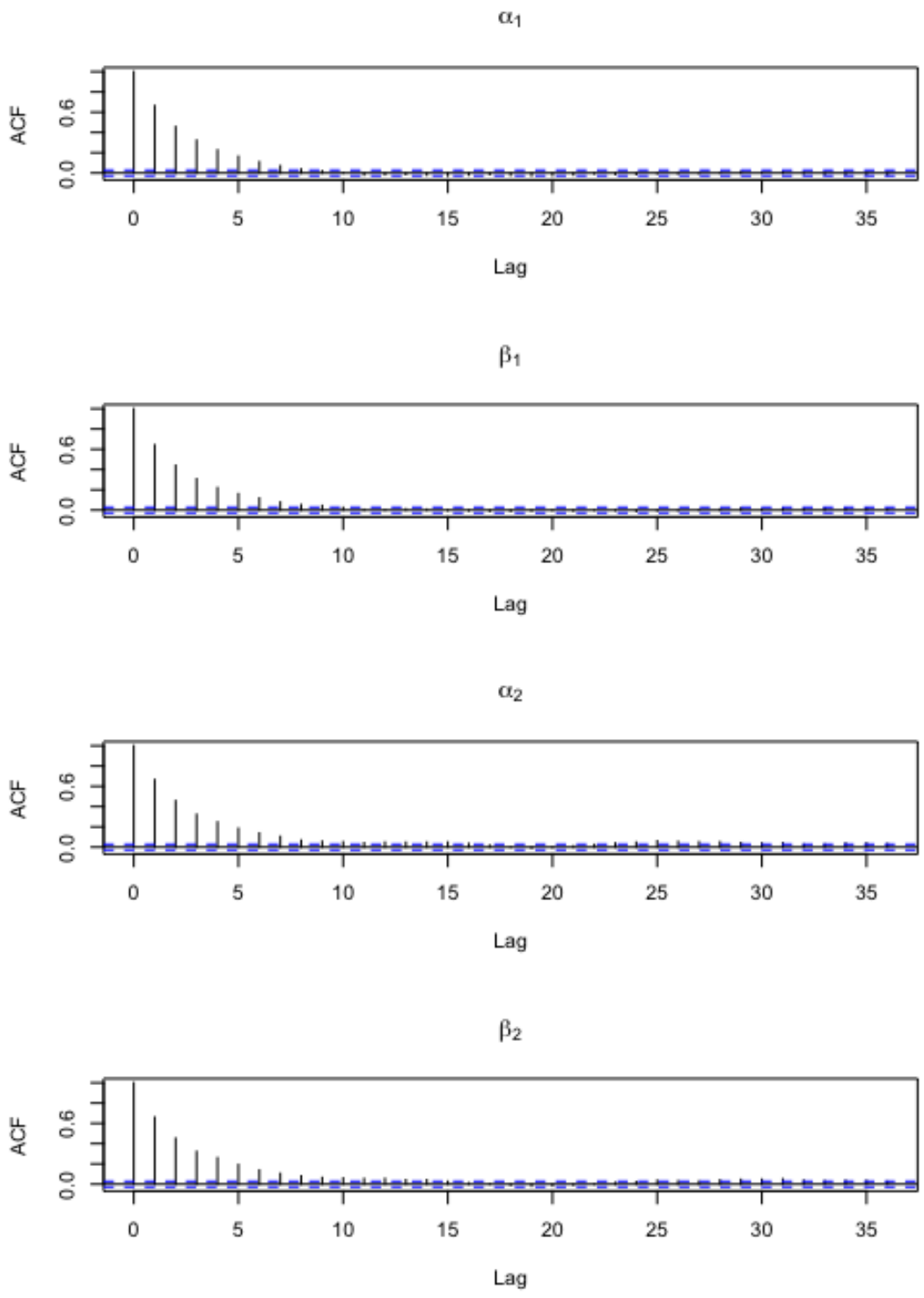


Figure 6.12: The autocorrelation plots of the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1).

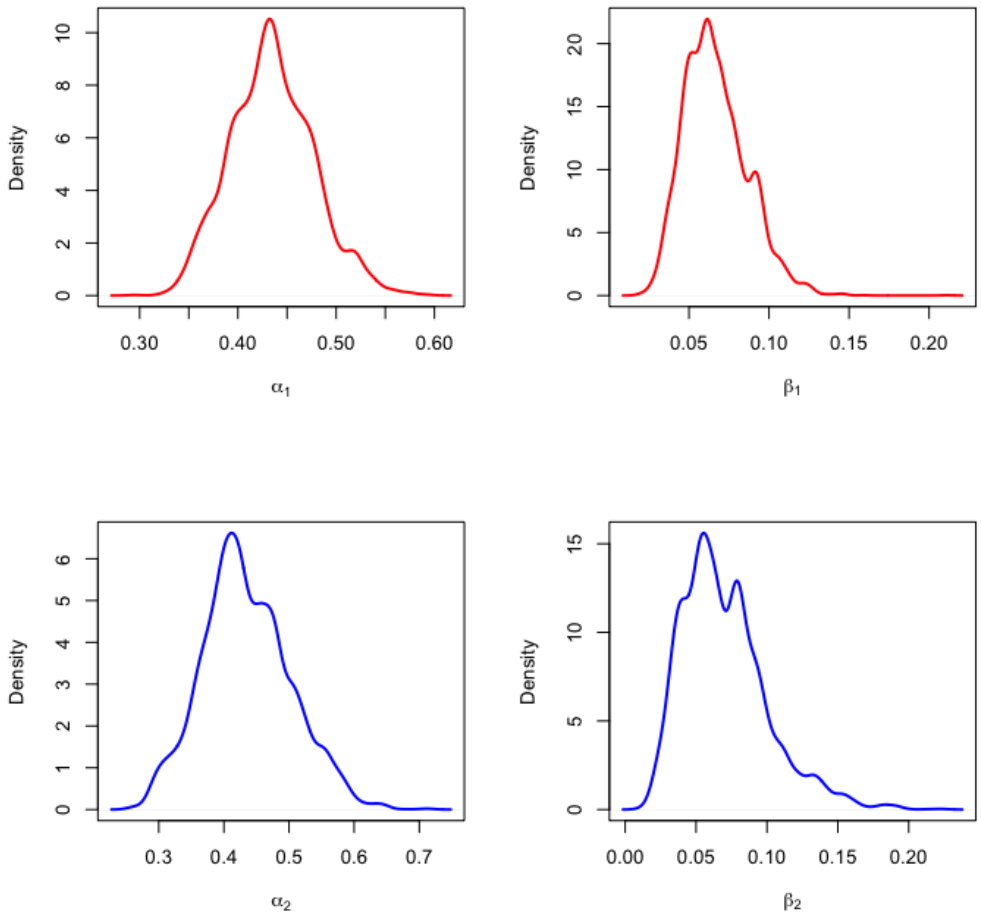


Figure 6.13: The marginal posterior densities for α_1 (top left), β_1 (top right), α_2 (bottom left), and β_2 (bottom right), for Power Lindley (group 1).

To determine the performance of the Metropolis-Hastings algorithm, diagnostic plots for the PL case (group 1) are shown in Figure 6.11 and 6.12. The trace plot appears to show randomness in the simulation and does not produce any visible patterns. The ACF plot shows correlation in the simulation process, however, goes to 0 after a few lags. In Figure 6.13, the marginal density for each parameter is shown. The red curves are the parameters for cause 1, where the blue curves are the parameters for cause 2. Diagnostic and marginal density plots are available for groups of all other models upon request. Furthermore, the acceptance rates for the

algorithm for all models are given in Table 6.11, which indicates good evidence about the sampling method used in the MCMC algorithm.

	Acceptance Rate		
Model	**Group 1**	**Group 2**	**Group 3**
Exponential	43.37	44.42	43.65
STH-1	43.36	44.46	43.46
Lindley	43.93	44.43	43.77
STH-2	23.95	24.28	24.97
Power Lindley	24.96	25.05	26.83
Weibull	24.04	23.26	21.59
Bathtub	24.91	24.01	26.17

Table 6.11: The acceptance rate for the Metropolis-Hastings algorithm using the good prior for all risks model used.

To contrast with the confidence intervals used in the maximum likelihood section, 95% credible intervals are calculated for the parameters (Table 6.15). However, the credible intervals are not interpreted in the same way as the confidence intervals are. Given a 95% credible interval, we can say that there is a 95% probability that an unobserved parameter will fall within this interval. For example, consider the 95% credible for β_1 in the Weibull model for group 2 is $[0.5644, 0.7436]$. Then, for some unobserved value of β_1, we can say that there is a 95% probability that this value will fall within the interval. By definition, the credible interval is given by determining the interval such that the integral of the posterior, over the limits of this interval, is 0.95. Since we do not have a closed form for the posterior, we calculate the credible intervals by using the simulated draws from the Metropolis-Hastings algorithm.

We also calculate the CIF using Bayesian methods to contrast with the CIF using maximum likelihood methods. Figure 6.14, 6.15, and 6.16 show the empirical CIF, the estimated CIF using maximum likelihood estimation, and the estimated CIF using Bayesian estimation together. From these figures, the estimated CIF using maximum likelihood and Bayesian estimation are consistent with one another.

Model	Bayes' Estimate		Variance	
	Cause 1	Cause 2	Cause 1	Cause 2
Group 1				
Exp	$\alpha_1 = 4.744 \times 10^{-4}$	$\alpha_2 = 4.775 \times 10^{-4}$	1.255×10^{-8}	1.233×10^{-8}
Lindley	$\alpha_1 = 1.508 \times 10^{-3}$	$\alpha_2 = 1.483 \times 10^{-3}$	1.777×10^{-8}	1.912×10^{-8}
STH-1	$\alpha_1 = 3.930 \times 10^{-4}$	$\alpha_2 = 3.978 \times 10^{-4}$	7.732×10^{-9}	7.473×10^{-9}
PL	$\alpha_1 = 0.5050$	$\alpha_2 = 0.4787$	5.611×10^{-3}	5.455×10^{-3}
	$\beta_1 = 4.513 \times 10^{-2}$	$\beta_2 = 5.611 \times 10^{-2}$	5.916×10^{-4}	6.618×10^{-4}
STH-2	$\alpha_1 = 0.5968$	$\alpha_2 = 0.6269$	1.083×10^{-2}	1.147×10^{-2}
	$\beta_1 = 8.685 \times 10^{-3}$	$\beta_2 = 7.255 \times 10^{-3}$	4.354×10^{-5}	3.507×10^{-5}
Weibull	$\alpha_1 = 4.554 \times 10^{-3}$	$\alpha_2 = 5.742 \times 10^{-3}$	4.546×10^{-7}	6.693×10^{-7}
	$\beta_1 = 0.6735$	$\beta_2 = 0.6386$	1.989×10^{-3}	1.762×10^{-3}
Bathtub	$\alpha_1 = 9.460 \times 10^{-3}$	$\alpha_2 = 8.477 \times 10^{-3}$	2.854×10^{-5}	2.478×10^{-5}
	$\beta_1 = 0.2001$	$\beta_2 = 0.2043$	3.248×10^{-4}	3.370×10^{-4}
Group 2				
Exp	$\alpha_1 = 1.066 \times 10^{-4}$	$\alpha_2 = 2.607 \times 10^{-4}$	1.154×10^{-9}	2.803×10^{-9}
Lindley	$\alpha_1 = 4.096 \times 10^{-4}$	$\alpha_2 = 8.660 \times 10^{-4}$	3.125×10^{-9}	4.912×10^{-9}
STH-1	$\alpha_1 = 9.842 \times 10^{-5}$	$\alpha_2 = 2.227 \times 10^{-4}$	8.898×10^{-10}	1.771×10^{-9}
PL	$\alpha_1 = 0.6486$	$\alpha_2 = 0.4089$	1.443×10^{-2}	3.371×10^{-3}
	$\beta_1 = 8.808 \times 10^{-3}$	$\beta_2 = 6.006 \times 10^{-2}$	6.118×10^{-5}	6.010×10^{-4}
STH-2	$\alpha_1 = 0.7164$	$\alpha_2 = 0.5652$	2.464×10^{-2}	8.444×10^{-3}
	$\beta_1 = 1.584 \times 10^{-3}$	$\beta_2 = 6.920 \times 10^{-3}$	4.415×10^{-6}	2.497×10^{-5}
Weibull	$\alpha_1 = 1.439 \times 10^{-3}$	$\alpha_2 = 4.175 \times 10^{-3}$	4.138×10^{-8}	3.386×10^{-7}
	$\beta_1 = 0.6567$	$\beta_2 = 0.6204$	2.103×10^{-3}	1.310×10^{-3}
Bathtub	$\alpha_1 = 1.353 \times 10^{-3}$	$\alpha_2 = 8.331 \times 10^{-3}$	1.569×10^{-6}	1.908×10^{-5}
	$\beta_1 = 0.2198$	$\beta_2 = 0.1868$	4.292×10^{-4}	2.524×10^{-4}
Group 3				
Exp	$\alpha_1 = 7.621 \times 10^{-4}$	$\alpha_2 = 4.591 \times 10^{-4}$	1.728×10^{-8}	1.085×10^{-8}
Lindley	$\alpha_1 = 1.983 \times 10^{-3}$	$\alpha_2 = 1.396 \times 10^{-3}$	5.047×10^{-8}	4.011×10^{-8}
STH-1	$\alpha_1 = 5.689 \times 10^{-4}$	$\alpha_2 = 3.616 \times 10^{-4}$	9.031×10^{-9}	5.864×10^{-9}
PL	$\alpha_1 = 0.4483$	$\alpha_2 = 0.3876$	2.411×10^{-3}	2.976×10^{-3}
	$\beta_1 = 8.429 \times 10^{-2}$	$\beta_2 = 8.899 \times 10^{-2}$	7.107×10^{-4}	1.018×10^{-3}
STH-2	$\alpha_1 = 0.5302$	$\alpha_2 = 0.4989$	4.442×10^{-3}	6.838×10^{-3}
	$\beta_1 = 1.827 \times 10^{-2}$	$\beta_2 = 1.515 \times 10^{-2}$	7.857×10^{-5}	7.764×10^{-5}
Weibull	$\alpha_1 = 1.186 \times 10^{-2}$	$\alpha_2 = 1.223 \times 10^{-2}$	2.855×10^{-6}	3.622×10^{-6}
	$\beta_1 = 0.6043$	$\beta_2 = 0.5273$	1.155×10^{-3}	1.320×10^{-3}
Bathtub	$\alpha_1 = 2.278 \times 10^{-2}$	$\alpha_2 = 1.614 \times 10^{-2}$	7.318×10^{-5}	5.157×10^{-5}
	$\beta_1 = 0.1812$	$\beta_2 = 0.1750$	1.741×10^{-4}	2.890×10^{-4}

Table 6.12: The Bayes estimate and respective variance for each model using a poor choice of hyper-parameters for the prior distribution (all hyper-parameters are set to 0.001).

Model	Bayes Estimate		Variance	
	Cause 1	Cause 2	Cause 1	Cause 2
Group 1				
Exp	$\alpha_1 = 4.792 \times 10^{-4}$	$\alpha_2 = 4.767 \times 10^{-4}$	4.769×10^{-9}	4.998×10^{-9}
Lindley	$\alpha_1 = 1.506 \times 10^{-3}$	$\alpha_2 = 1.488 \times 10^{-3}$	2.013×10^{-9}	1.894×10^{-9}
STH-1	$\alpha_1 = 3.941 \times 10^{-4}$	$\alpha_2 = 3.995 \times 10^{-4}$	2.907×10^{-9}	2.957×10^{-9}
PL	$\alpha_1 = 0.4986$	$\alpha_2 = 0.4726$	1.578×10^{-3}	1.331×10^{-3}
	$\beta_1 = 4.325 \times 10^{-2}$	$\beta_2 = 5.042 \times 10^{-2}$	1.295×10^{-3}	1.492×10^{-4}
STH-2	$\alpha_1 = 0.6012$	$\alpha_2 = 0.6173$	5.357×10^{-3}	1.222×10^{-2}
	$\beta_1 = 7.328 \times 10^{-3}$	$\beta_2 = 7.929 \times 10^{-3}$	1.478×10^{-5}	4.333×10^{-5}
Weibull	$\alpha_1 = 4.544 \times 10^{-3}$	$\alpha_2 = 5.741 \times 10^{-3}$	1.896×10^{-7}	6.246×10^{-7}
	$\beta_1 = 0.6742$	$\beta_2 = 0.6391$	1.519×10^{-3}	1.651×10^{-3}
Bathtub	$\alpha_1 = 8.826 \times 10^{-3}$	$\alpha_2 = 7.994 \times 10^{-3}$	9.075×10^{-6}	7.882×10^{-6}
	$\beta_1 = 0.2000$	$\beta_2 = 0.2032$	1.187×10^{-4}	1.162×10^{-4}
Group 2				
Exp	$\alpha_1 = 1.067 \times 10^{-4}$	$\alpha_2 = 2.616 \times 10^{-4}$	4.692×10^{-10}	1.115×10^{-9}
Lindley	$\alpha_1 = 4.927 \times 10^{-4}$	$\alpha_2 = 8.665 \times 10^{-4}$	3.142×10^{-9}	4.465×10^{-9}
STH-1	$\alpha_1 = 9.846 \times 10^{-5}$	$\alpha_2 = 2.227 \times 10^{-4}$	3.511×10^{-10}	6.810×10^{-10}
PL	$\alpha_1 = 0.6352$	$\alpha_2 = 0.4036$	3.886×10^{-3}	8.655×10^{-4}
	$\beta_1 = 7.461 \times 10^{-3}$	$\beta_2 = 5.861 \times 10^{-2}$	1.232×10^{-5}	1.440×10^{-4}
STH-2	$\alpha_1 = 0.7104$	$\alpha_2 = 0.5723$	1.300×10^{-2}	9.047×10^{-3}
	$\beta_1 = 1.304 \times 10^{-3}$	$\beta_2 = 6.705 \times 10^{-3}$	1.839×10^{-6}	2.462×10^{-5}
Weibull	$\alpha_1 = 1.454 \times 10^{-3}$	$\alpha_2 = 4.155 \times 10^{-3}$	1.968×10^{-8}	3.022×10^{-7}
	$\beta_1 = 0.6556$	$\beta_2 = 0.6185$	1.852×10^{-3}	1.157×10^{-3}
Bathtub	$\alpha_1 = 1.258 \times 10^{-3}$	$\alpha_2 = 7.736 \times 10^{-3}$	4.553×10^{-7}	5.132×10^{-6}
	$\beta_1 = 0.2179$	$\beta_1 = 0.1865$	2.012×10^{-4}	9.569×10^{-5}
Group 3				
Exp	$\alpha_1 = 7.612 \times 10^{-4}$	$\alpha_2 = 4.599 \times 10^{-4}$	7.150×10^{-9}	4.409×10^{-9}
Lindley	$\alpha_1 = 1.986 \times 10^{-3}$	$\alpha_2 = 1.398 \times 10^{-3}$	1.968×10^{-8}	1.438×10^{-8}
STH-1	$\alpha_1 = 5.705 \times 10^{-4}$	$\alpha_2 = 3.617 \times 10^{-4}$	3.367×10^{-9}	2.125×10^{-9}
PL	$\alpha_1 = 0.4986$	$\alpha_2 = 0.4726$	1.578×10^{-3}	1.331×10^{-3}
	$\beta_1 = 4.325 \times 10^{-2}$	$\beta_2 = 5.042 \times 10^{-2}$	1.295×10^{-4}	1.492×10^{-4}
STH-2	$\alpha_1 = 0.5296$	$\alpha_2 = 0.4939$	2.390×10^{-3}	6.808×10^{-3}
	$\beta_1 = 1.710 \times 10^{-2}$	$\beta_2 = 1.533 \times 10^{-2}$	3.833×10^{-5}	7.720×10^{-5}
Weibull	$\alpha_1 = 1.179 \times 10^{-2}$	$\alpha_2 = 1.226 \times 10^{-2}$	1.133×10^{-6}	2.761×10^{-6}
	$\beta_1 = 0.6033$	$\beta_2 = 0.5258$	8.910×10^{-4}	1.269×10^{-3}
Bathtub	$\alpha_1 = 2.189 \times 10^{-2}$	$\alpha_2 = 1.510 \times 10^{-2}$	2.132×10^{-5}	1.721×10^{-5}
	$\beta_1 = 0.1815$	$\beta_2 = 0.1764$	5.488×10^{-5}	9.583×10^{-5}

Table 6.13: The Bayes estimate and respective variance for each model using the hyper-parameters from Table 6.14 for the prior distribution (good prior).

Model	**Hyperparameters**			
	Cause 1		**Cause 2**	
Group 1	α_1		α_2	
Exp	$a_1 = 18$	$b_1 = 37796$	$a_2 = 19$	$b_2 = 38726$
Lindley	$a_1 = 48$	$b_1 = 31748$	$a_2 = 47$	$b_2 = 31503$
STH-1	$a_1 = 20$	$b_1 = 50832$	$a_2 = 21$	$b_2 = 52233$
Group 1	α_1	β_1	α_2	β_1
PL	$a_{11} = 45$	$b_{11} = 4$	$a_{21} = 42$	$b_{21} = 4$
	$a_{12} = 90$	$b_{12} = 87$	$a_{22} = 88$	$b_{22} = 80$
STH-2	$a_{11} = 32$	$b_{11} = 2$	$a_{21} = 34$	$b_{21} = 2$
	$a_{12} = 55$	$b_{12} = 199$	$a_{22} = 55$	$b_{22} = 207$
Weibull	$a_{11} = 50$	$b_{11} = 48$	$a_{21} = 258$	$b_{21} = 249$
	$a_{12} = 11061$	$b_{12} = 8320$	$a_{22} = 383$	$b_{22} = 390$
Bathtub	$a_{11} = 3$	$b_{11} = 123$	$a_{21} = 3$	$b_{21} = 124$
	$a_{12} = 331$	$b_{12} = 616$	$a_{22} = 342$	$b_{22} = 606$
Group 2	α_1		α_2	
Exp	$a_1 = 10$	$b_1 = 92330$	$a_2 = 24$	$b_2 = 92978$
Lindley	$a_1 = 30$	$b_1 = 61483$	$a_2 = 66$	$b_2 = 76097$
STH-1	$a_1 = 11$	$b_1 = 110605$	$a_2 = 28$	$b_2 = 125781$
Group 2	α_1	β_1	α_2	β_2
PL	$a_{11} = 29$	$b_{11} = 1$	$a_{21} = 50$	$b_{21} = 6$
	$a_{12} = 45$	$b_{12} = 144$	$a_{22} = 121$	$b_{22} = 100$
STH-2	$a_{11} = 21$	$b_{11} = 1$	$a_{21} = 38$	$b_{21} = 2$
	$a_{12} = 29$	$b_{12} = 359$	$a_{22} = 67$	$b_{22} = 277$
Weibull	$a_{11} = 47$	$b_{11} = 191$	$a_{21} = 50$	$b_{21} = 310$
	$a_{12} = 32531$	$b_{12} = 293$	$a_{22} = 12044$	$b_{22} = 502$
Bathtub	$a_{11} = 1$	$b_{11} = 113$	$a_{21} = 4$	$b_{21} = 138$
	$a_{12} = 863$	$b_{12} = 512$	$a_{22} = 437$	$b_{22} = 740$
Group 3	α_1		α_2	
Exp	$a_1 = 34$	$b_1 = 44116$	$a_2 = 19$	$b_2 = 42308$
Lindley	$a_1 = 78$	$b_1 = 32924$	$a_2 = 49$	$b_2 = 34800$
STH-1	$a_1 = 36$	$b_1 = 62989$	$a_2 = 21$	$b_2 = 53233$
Group 3	α_1	β_1	α_2	β_2
PL	$a_{11} = 83$	$b_{11} = 10$	$a_{21} = 50$	$b_{21} = 8$
	$a_{12} = 186$	$b_{12} = 119$	$a_{22} = 130$	$b_{22} = 87$
STH-2	$a_{11} = 63$	$b_{11} = 4$	$a_{21} = 36$	$b_{21} = 3$
	$a_{12} = 119$	$b_{12} = 232$	$a_{22} = 73$	$b_{22} = 195$
Weibull	$a_{11} = 44$	$b_{11} = 299$	$a_{21} = 46$	$b_{21} = 205$
	$a_{12} = 3733$	$b_{12} = 496$	$a_{22} = 3765$	$b_{22} = 390$
Bathtub	$a_{11} = 8$	$b_{11} = 200$	$a_{21} = 5$	$b_{21} = 120$
	$a_{12} = 351$	$b_{12} = 1099$	$a_{22} = 326$	$b_{22} = 678$

Table 6.14: The hyperparameters used in the good prior distribution to calculate the Bayes' estimate in Table 6.13

Model	Par.	95% Cred. Int.	Par.	95% Cred. Int.
Group 1				
Exp	α_1	$[3.484 \times 10^{-4}, 6.206 \times 10^{-4}]$	α_2	$[3.486 \times 10^{-4}, 6.355 \times 10^{-4}]$
Lindley	α_1	$[1.259 \times 10^{-3}, 1.803 \times 10^{-3}]$	α_2	$[1.240 \times 10^{-3}, 1.761 \times 10^{-3}]$
STH-1	α_1	$[2.446 \times 10^{-4}, 5.834 \times 10^{-4}]$	α_2	$[2.472 \times 10^{-4}, 5.770 \times 10^{-4}]$
PL	α_1	$[0.3594, 0.5309]$	α_2	$[0.3156, 0.5828]$
	β_1	$[0.0344, 0.1117]$	β_2	$[0.0252, 0.1444]$
STH-2	α_1	$[0.4746, 0.7457]$	α_2	$[0.4354, 0.8525]$
	β_1	$[2.527 \times 10^{-3}, 0.0166]$	β_2	$[1.270 \times 10^{-3}, 0.0227]$
Weibull	α_1	$[3.780 \times 10^{-3}, 5.504 \times 10^{-3}]$	α_2	$[4.205 \times 10^{-3}, 7.456 \times 10^{-3}]$
	β_1	$[0.5941, 0.7463]$	β_2	$[0.5649, 0.7198]$
Bathtub	α_1	$[3.749 \times 10^{-3}, 0.0155]$	α_2	$[3.925 \times 10^{-3}, 0.0278]$
	β_1	$[0.1746, 0.2290]$	β_2	$[0.1563, 0.2264]$
Group 2				
Exp	α_1	$[6.826 \times 10^{-5}, 1.525 \times 10^{-4}]$	α_2	$[2.002 \times 10^{-4}, 3.329 \times 10^{-4}]$
Lindley	α_1	$[3.869 \times 10^{-4}, 6.082 \times 10^{-4}]$	α_2	$[7.370 \times 10^{-4}, 1.013 \times 10^{-3}]$
STH-1	α_1	$[4.810 \times 10^{-5}, 1.637 \times 10^{-4}]$	α_2	$[1.514 \times 10^{-4}, 3.176 \times 10^{-4}]$
PL	α_1	$[0.3488, 0.5659]$	α_2	$[0.2814, 0.4963]$
	β_1	$[0.0116, 0.0568]$	β_2	$[0.0322, 0.1393]$
STH-2	α_1	$[0.5342, 0.9276]$	α_2	$[0.4147, 0.7600]$
	β_1	$[1.806 \times 10^{-4}, 4.143 \times 10^{-3}]$	β_2	$[1.400 \times 10^{-3}, 0.0181]$
Weibull	α_1	$[1.186 \times 10^{-3}, 1.723 \times 10^{-3}]$	α_2	$[3.154 \times 10^{-3}, 5.717 \times 10^{-3}]$
	β_1	$[0.5644, 0.7436]$	β_2	$[0.5432, 0.6926]$
Bathtub	α_1	$[3.334 \times 10^{-4}, 2.917 \times 10^{-3}]$	α_2	$[4.012 \times 10^{-3}, 0.0233]$
	β_1	$[0.1909, 0.2493]$	β_2	$[0.1480, 0.2076]$
Group 3				
Exp	α_1	$[6.069 \times 10^{-5}, 9.421 \times 10^{-4}]$	α_2	$[3.413 \times 10^{-5}, 5.935 \times 10^{-4}]$
Lindley	α_1	$[1.711 \times 10^{-3}, 2.283 \times 10^{-3}]$	α_2	$[1.161 \times 10^{-3}, 1.644 \times 10^{-3}]$
STH-1	α_1	$[4.082 \times 10^{-4}, 7.696 \times 10^{-4}]$	α_2	$[2.239 \times 10^{-4}, 5.271 \times 10^{-4}]$
PL	α_1	$[0.3632, 0.4805]$	α_2	$[0.2661, 0.4728]$
	β_1	$[0.0638, 0.1450]$	β_2	$[0.0517, 0.1886]$
STH-2	α_1	$[0.4522, 0.6126]$	α_2	$[0.3493, 0.6613]$
	β_1	$[8.614 \times 10^{-3}, 0.0307]$	β_2	$[4.300 \times 10^{-3}, 0.0397]$
Weibull	α_1	$[9.790 \times 10^{-3}, 0.0139]$	α_2	$[9.552 \times 10^{-3}, 0.0157]$
	β_1	$[0.5415, 0.6661]$	β_2	$[0.4552, 0.5982]$
Bathtub	α_1	$[0.0142, 0.0320]$	α_2	$[8.213 \times 10^{-3}, 0.0403]$
	β_1	$[0.1642, 0.1988]$	β_2	$[0.1381, 0.1981]$

Table 6.15: The 95% credible intervals for the parameters.

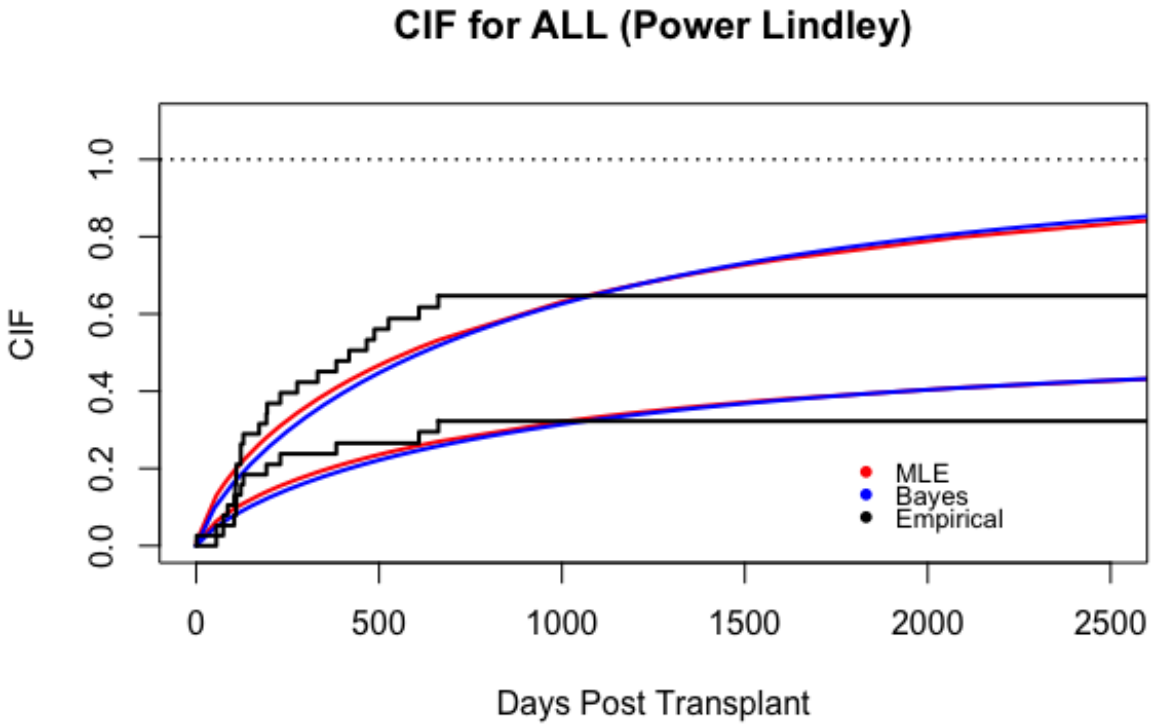


Figure 6.14: A comparison of the estimated CIF for group 1 (ALL patients) using Bayes', MLE and empirical methods, where the risks follow the Power Lindley distribution.

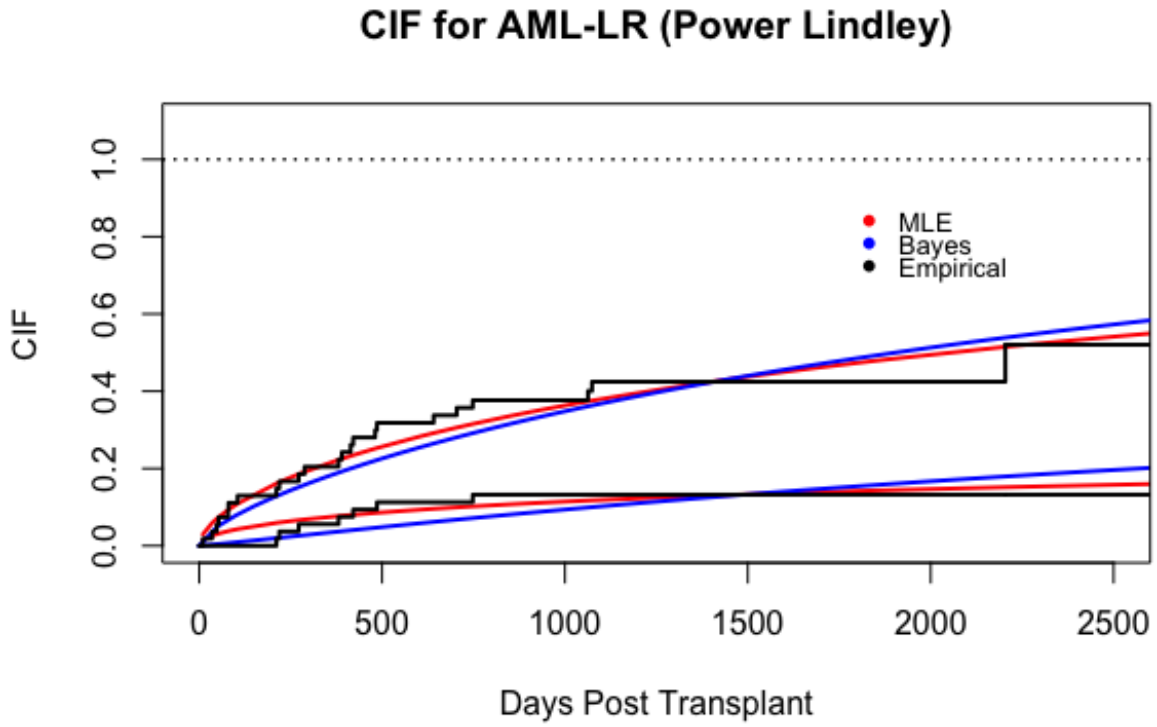


Figure 6.15: A comparison of the estimated CIF for group 2 (AML-LR patients) using Bayes', MLE and empirical methods, where the risks follow the Power Lindley distribution.

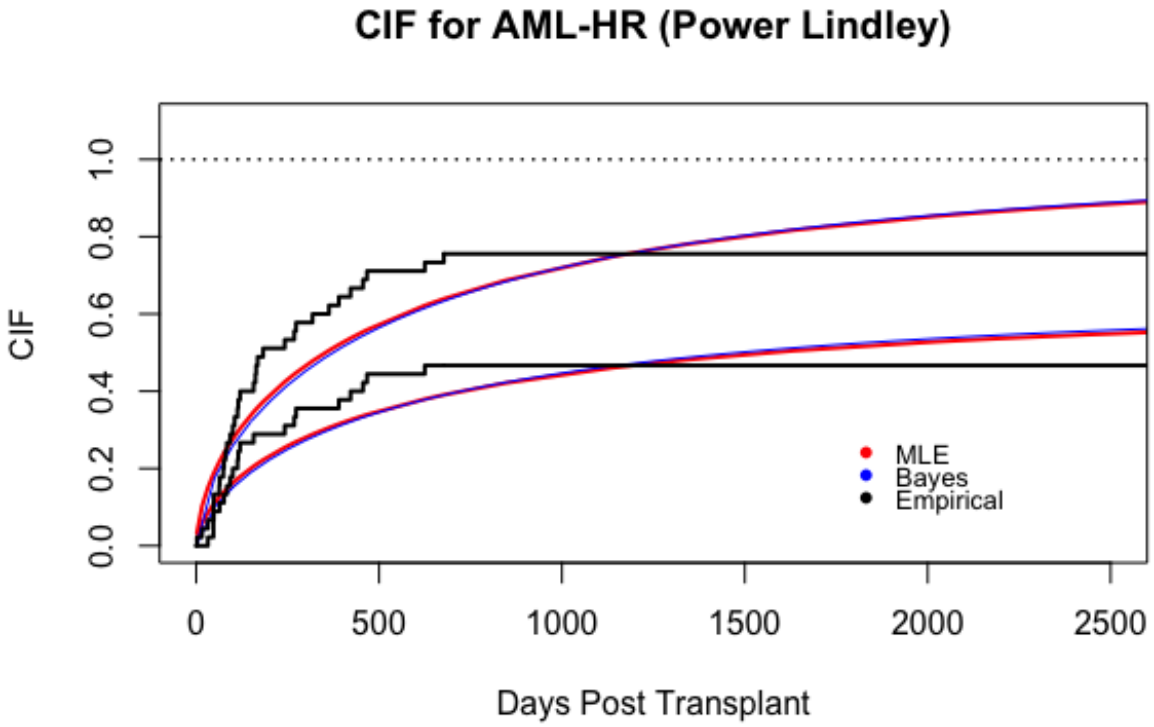


Figure 6.16: A comparison of the estimated CIF for group 3 (AML-HR patients) using Bayes', MLE and empirical methods, where the risks follow the Power Lindley distribution.

6.4 Competing Risks with Covariates

The BMT data consist of 10 covariates. The effects on relapse and remission from covariates such as patient's and donor's age, sex, CMV status, waiting time for transplant, FAB indicator, location of hospital, and MTX indicator may be analysed depending on the goals of the researcher. The summary for each variable can be seen in Table 6.16. Table 6.16 provides count data for each indicator for categorical variables, and the mean for quantitative variables.

For the purpose of the thesis and the questions we wish to explore, an analysis solely based on the patient's age and donor's age will be conducted. This simplification of covariates will also reduce the complexity of estimating 2×10 ($k \times p$) coefficients of the covariates, which will ultimately affect the accuracy and computational time of the algorithms used for estimation. However, other researchers may still have different research questions and may want to study the effects using additional covariates or a set of different covariates, which we will consider in future work.

6.4.1 Cox Regression

The results of the cause-specific hazard regression are presented in this section. To perform the regression, we construct two regression models for each group in which the competing risks are treated as censored observations. For example, the regression model for relapse considered observations from remission as censored, and the regression model for remission considered observations from relapse as censored.

For each group, we only consider the patient's age (Z_1) and donor's age (Z_2) as the risk factors for the patients. The results of the cause-specific cox regression for all three groups are seen in Table 6.17. For group 1, the patient's age is found to be significant for the relapse hazard model, where the donor's age is found to be significant for the remission hazard model at a 10% significance level.

Variable	Description	Data Summary
Z_1	Age of Patient (years)	Group 1: 24.42 Group 2: 29.41 Group 3: 30.44
Z_2	Age of Donor (years)	Group 1: 26.79 Group 2: 28.07 Group 3: 29.93
Z_3	Patient Sex	Group 1: M (26), F (12) Group 2: M (30), F (24) Group 3: M (24), F (21)
Z_4	Donor Sex	Group 1: M (26), F (12) Group 2: M (34), F (20) Group 3: M (28), F (17)
Z_5	Patient CMV Status	Group 1: CMV + (15), CMV − (23) Group 2: CMV + (26), CMV − (28) Group 3: CMV + (27), CMV − (18)
Z_6	Donor CMV Status	Group 1: CMV + (17), CMV − (21) Group 2: CMV + (22), CMV − (34) Group 3: CMV + (19), CMV − (26)
Z_7	Waiting Time for Transplants (days)	Group 1: 477.18 Group 2: 138.06 Group 3: 268.87
Z_8	FAB Indicator: 1 - FAB Grade 4 or 5 and AML 0 - Otherwise	Group 1: 1 (0), 0 (45) Group 2: 1 (18), 0 (36) Group 3: 1 (27), 0 (18)
Z_9	Location of Hospital: 1 - Ohio State, 2 - Alferd 3 - St. Vincent, 4 - Hahnemann	Group 1: 1 (21), 2 (8), 3 (9), 4 (0) Group 2: 1 (27), 2 (5), 3 (7), 4 (15) Group 3: 1 (28), 2 (4), 3 (7), 4 (6)
Z_{10}	MTX Used as a Graft-Versus-Host-Prophylactic	Group 1: Yes (17), No (21) Group 2: Yes (12), No (42) Group 3: Yes (11), No (34)

Table 6.16: A description of the covariates (risk factors) for the patients.

Covariates	**Cause 1**			**Cause 2**		
	Coef.	exp(Coef.)	P-Value	Coef.	exp(Coef.)	P-Value
Group 1						
Z_1	0.1155	1.1224	0.029	0.0939	1.0984	0.1309
Z_2	-0.0361	0.9645	0.451	-0.0871	0.9166	0.0606
Group 2						
Z_1	-0.0226	0.9777	0.735	-0.0297	0.9707	0.496
Z_2	-0.0448	0.9562	0.522	0.1391	1.1493	0.010
Group 3						
Z_1	-0.0390	0.9618	0.310	-0.0540	0.9475	0.289
Z_2	0.0332	1.0338	0.358	0.0298	1.0302	0.527

Table 6.17: The results of Cox's cause-specific hazard model.

For group 2, no covariates are found to be significant for the relapse model, however, the donor's age was found significant for the remission model at a 1% significance level. For group 3, both models showed no significant covariates. To discuss the interpretation of the results, we discuss the meaning of the regression coefficient. A coefficient greater than 0 is known as a bad prognostic factor, and is associated to an increase of the hazard for the underlying risk. A coefficient less than 0 is known as a good prognostic factor, and is associated to a reduction of the hazard for the underlying risk. For example, in group 1, the relapse model found that the patient's age was significant. Since the coefficient was greater than 0, then we find the that patients' age is a bad prognostic factor. Assuming the donor's age is kept constant, then an additional year in the patient's age is associated to an increase of the hazard for relapse by a factor of 1.224. In the remission model for the same group, the donor's age was found significant. The coefficient is less than 0, and is seen as a good prognostic factor. Assuming that the patient's age is kept constant, then an additional year in the donor's age is associated to a reduction of the hazard for remission by a factor of 0.9166.

In this section, we left the baseline cause-specific hazard unspecified and assumed no parametric form. In the next section, we assume a parametric form for the baseline cause-specific hazard and estimate the parameters of this hazard, as well as the coefficients of the covariates, using maximum likelihood and Bayesian methods.

6.4.2 Maximum Likelihood and Bayesian Estimation

In this section, we estimate the parameters of a fully parametric cause-specific hazard model with covariates. Based on the results of our homogeneous competing risks model, we found that the Power Lindley model was the best fit model for group 1 and 3, and was the second best fit model for group 2. For this reason, we will only consider the estimation of the full model parameters for the Power Lindley risks model. For the covariates, we will estimate the coefficients of Z_1 and Z_2 for demonstration purposes regardless of it being found to significant in Cox's regression. The MLEs, the value of the log-likelihood evaluated at the MLEs, and the Bayes estimates can be found in Table 6.18. The estimates of the parameters and the estimates of the coefficients from the regression parametric model are consistent with the MLEs and Bayes estimates found from our homogeneous competing risks model, and with the coefficient estimates found from the Cox model. The trace plots and ACF plots for the simulated draws for Group 1 are shown in Figure 6.17 and 6.18, as diagnostic tests. From these figures, we can see that there is a good mix in the sample and the draws become more independently quickly. The acceptance rates for the Metropolis-Hastings algorithm are 30.06% (group 1), 29.52% (group 2) and 30.81% (group 3).

MLE		Bayes Estimate	
Cause 1	**Cause 2**	**Cause 1**	**Cause 2**
Group 1 $\mathcal{L} = -196.531$		**Group 1**	
$\alpha_1 = 0.4595$	$\alpha_2 = 0.4832$	$\alpha_1 = 0.4489$	$\alpha_2 = 0.4892$
$\beta_1 = 0.0178$	$\beta_2 = 0.0538$	$\beta_1 = 0.0183$	$\beta_2 = 0.0550$
$\gamma_{11} = 0.1399$	$\gamma_{21} = 0.0902$	$\gamma_{11} = 0.1417$	$\gamma_{21} = 0.0868$
$\gamma_{12} = -0.0574$	$\gamma_{22} = -0.0919$	$\gamma_{12} = -0.0582$	$\gamma_{22} = -0.0937$
Group 2 $\mathcal{L} = -206.746$		**Group 2**	
$\alpha_1 = 0.4699$	$\alpha_2 = 0.3998$	$\alpha_1 = 0.5339$	$\alpha_2 = 0.3767$
$\beta_1 = 0.0654$	$\beta_2 = 8.635 \times 10^{-3}$	$\beta_1 = 0.0511$	$\beta_2 = 9.975 \times 10^{-3}$
$\gamma_{11} = 0.0136$	$\gamma_{21} = -0.0429$	$\gamma_{11} = -0.0396$	$\gamma_{21} = -0.0214$
$\gamma_{12} = -0.0878$	$\gamma_{22} = 0.1584$	$\gamma_{12} = -0.0434$	$\gamma_{22} = 0.1307$
Group 3 $\mathcal{L} = -270.006$		**Group 3**	
$\alpha_1 = 0.4434$	$\alpha_2 = 0.3827$	$\alpha_1 = 0.4464$	$\alpha_2 = 0.3968$
$\beta_1 = 0.1026$	$\beta_2 = 0.1595$	$\beta_1 = 0.1041$	$\beta_2 = 0.1477$
$\gamma_{11} = -0.0512$	$\gamma_{21} = -0.0561$	$\gamma_{11} = -0.0496$	$\gamma_{21} = -0.0687$
$\gamma_{12} = 0.0416$	$\gamma_{22} = 0.0265$	$\gamma_{12} = 0.0377$	$\gamma_{22} = 0.0363$

Table 6.18: The MLEs and Bayes estimates for the parametric regression model where the baseline cause-specific hazard is assumed to follow a Power Lindley distribution.

Since the homogeneous model is nested within the regression model, a hypothesis test using a likelihood ratio test statistic is done to determine the significance of the covariates. To do so, the following hypotheses are tested:

$$H_0 : \boldsymbol{\gamma} = 0 \text{ and } H_A : \boldsymbol{\gamma} \neq 0,$$

where the null hypothesis is testing the significance of the reduced model (homogeneous model), and the alternative is testing the significance of the full model (regression model). The test statistic is

$$\Lambda = (-2\mathcal{L}_{\text{Reduced}}) - (-2\mathcal{L}_{\text{Full}}) \sim \chi^2_b,$$

where b is the number of reduced parameters. The values of $\mathcal{L}_{\text{Full}}$ for group 1, 2 and 3 are -201.859, -214.240 and -271.890, respectively. Using these values, we find the test

statistic for each group to be

$$\begin{aligned}
\Lambda_1 &= 403.718 - 393.062 = 10.656 \sim \chi^2_4, \\
\Lambda_2 &= 428.480 - 413.492 = 14.988 \sim \chi^2_4, \\
\Lambda_3 &= 543.780 - 540.002 = 3.786 \sim \chi^2_4.
\end{aligned}$$

Then, for group 1, 2 and 3, we find the p-value to be 0.0307, 0.0047, and 0.4357, respectively. Therefore, we reject the null hypothesis for group 1 and 2, and fail to reject the null hypothesis for group 3 at a significance level of 0.05. These results are consistent with the results found from the Cox model, which found significant covariates for group 1 and 2, and no significant covariates for group 3. To further investigate the significance of the coefficients, a test of hypothesis on the significance of the true coefficients is performed. Consider the following hypotheses:

$$H_0 : \gamma_{jp} = 0 \text{ and } H_A : \gamma_{jp} \neq 0$$

for $j = 1, 2$ and $p = 1, 2$. Using the 95% asymptotic confidence intervals (Table 6.19), we reject the null hypothesis if the interval does not contain the hypothesized value (zero), and conclude the covariate is statistically significant. For group 1, we reject the claim that γ_{11} and γ_{22} are equal to 0. For group 2, we reject the claim that γ_{22} is equal to 0. For group 3, all intervals contain 0 so we fail to reject the claim that the coefficients are equal to 0. Therefore, for group 1, these results suggest that patient's age is significant for the relapse hazard model and donor's age is significant for the remission hazard model at a 5% significance level. For group 2, it suggest that the donor's age is significant for the remission model at a 5% significance level. For group 3, we find no significant covariates at a 5% significance level. The results of this hypothesis test are consistent with the results found from the Cox's cause-specific hazard model.

Par.	**95% CI**	**Par.**	**95% CI**
Group 1			
α_1	[0.4298, 0.4913]	α_2	[0.4509, 0.5178]
β_1	[0.0132, 0.0240]	β_2	[0.0415, 0.0698]
γ_{11}	[0.0268, 0.2530]	γ_{21}	[-0.0292, 0.2098]
γ_{12}	[-0.1493, 0.0345]	γ_{22}	-0.1794, -0.0044]
Group 2			
α_1	[0.4362, 0.5062]	α_2	[0.3784, 0.4225]
β_1	[0.0523, 0.0818]	β_2	[0.0063, 0.0118]
γ_{11}	[-0.1197, 0.1469]	γ_{21}	[-0.1233, 0.0375]
γ_{12}	[-0.2283, 0.0525]	γ_{22}	[0.0426, 0.2742]
Group 3			
α_1	[0.4214, 0.4630]	α_2	[0.3581, 0.4048]
β_1	[0.0857, 0.1161]	β_2	[0.1281, 0.1828]
γ_{11}	[-0.1162, 0.0313]	γ_{21}	[-0.1608, 0.0381]
γ_{12}	[-0.0351, 0.1047]	γ_{22}	[-0.0576, 0.1257]

Table 6.19: 95% asymptotic confidence intervals for the true parameters and coefficients of the regression model, where the risks are assumed to follow the Power Lindley distribution.

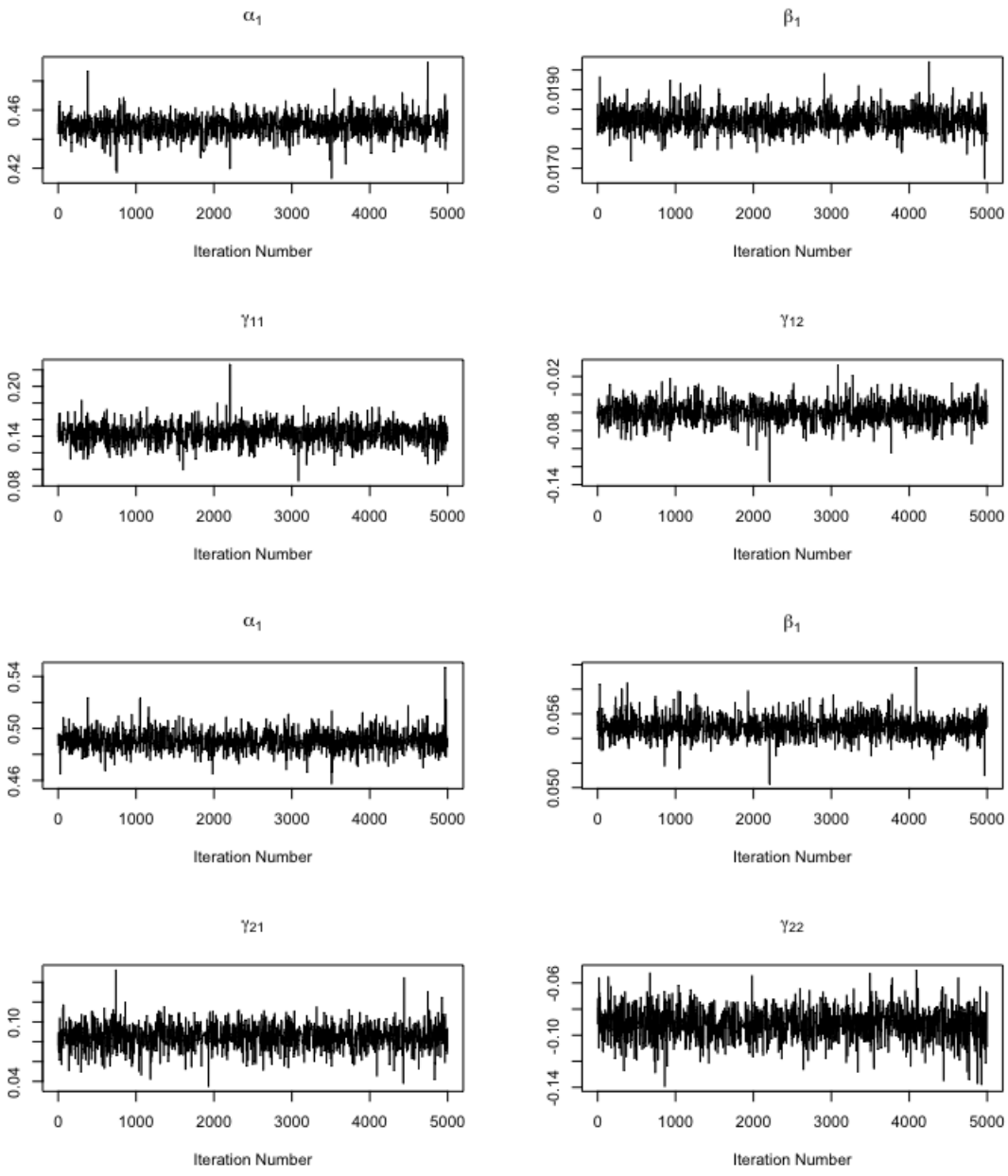


Figure 6.17: The trace plots for the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1).

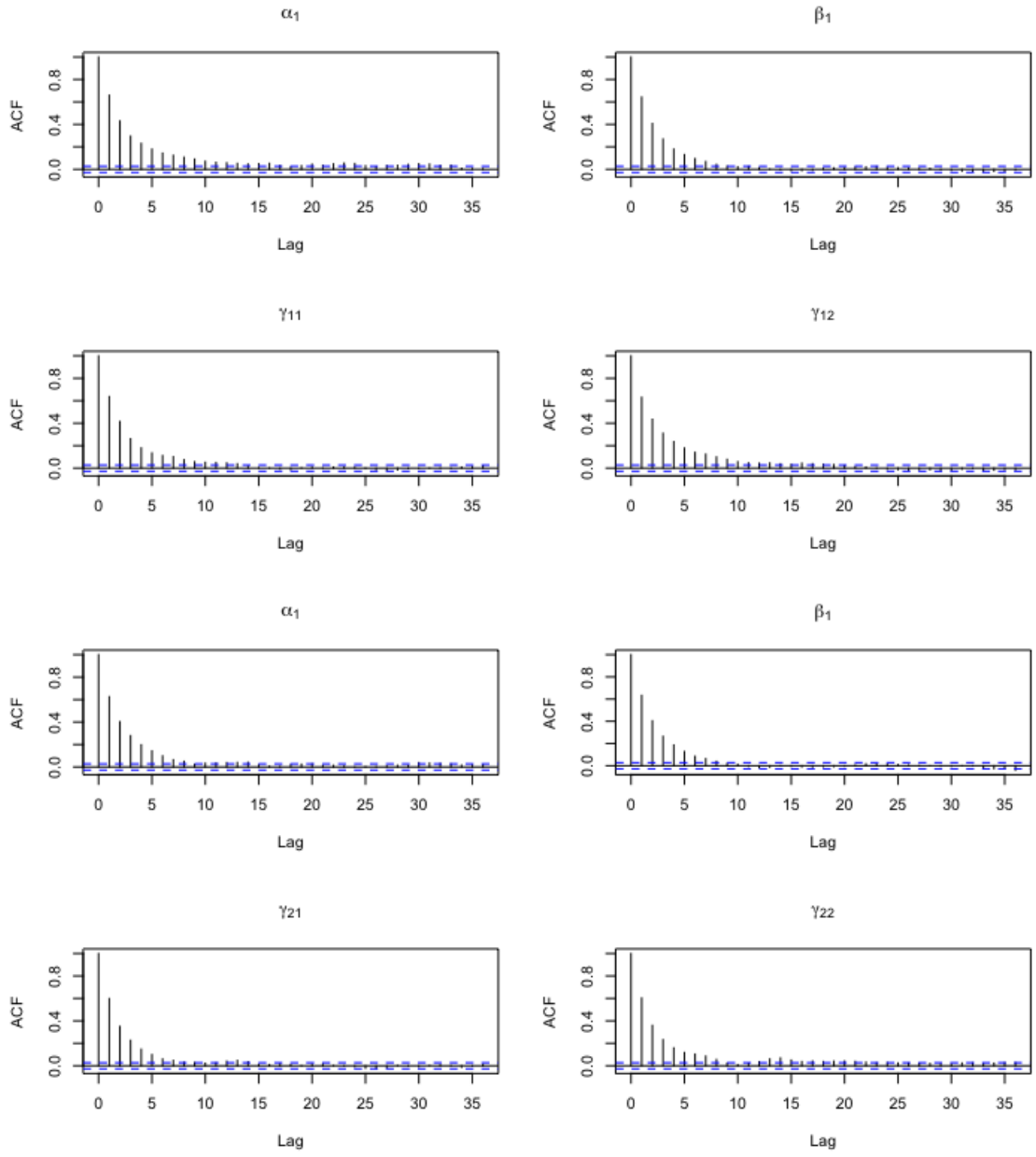


Figure 6.18: The ACF plots for the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1), with regression.

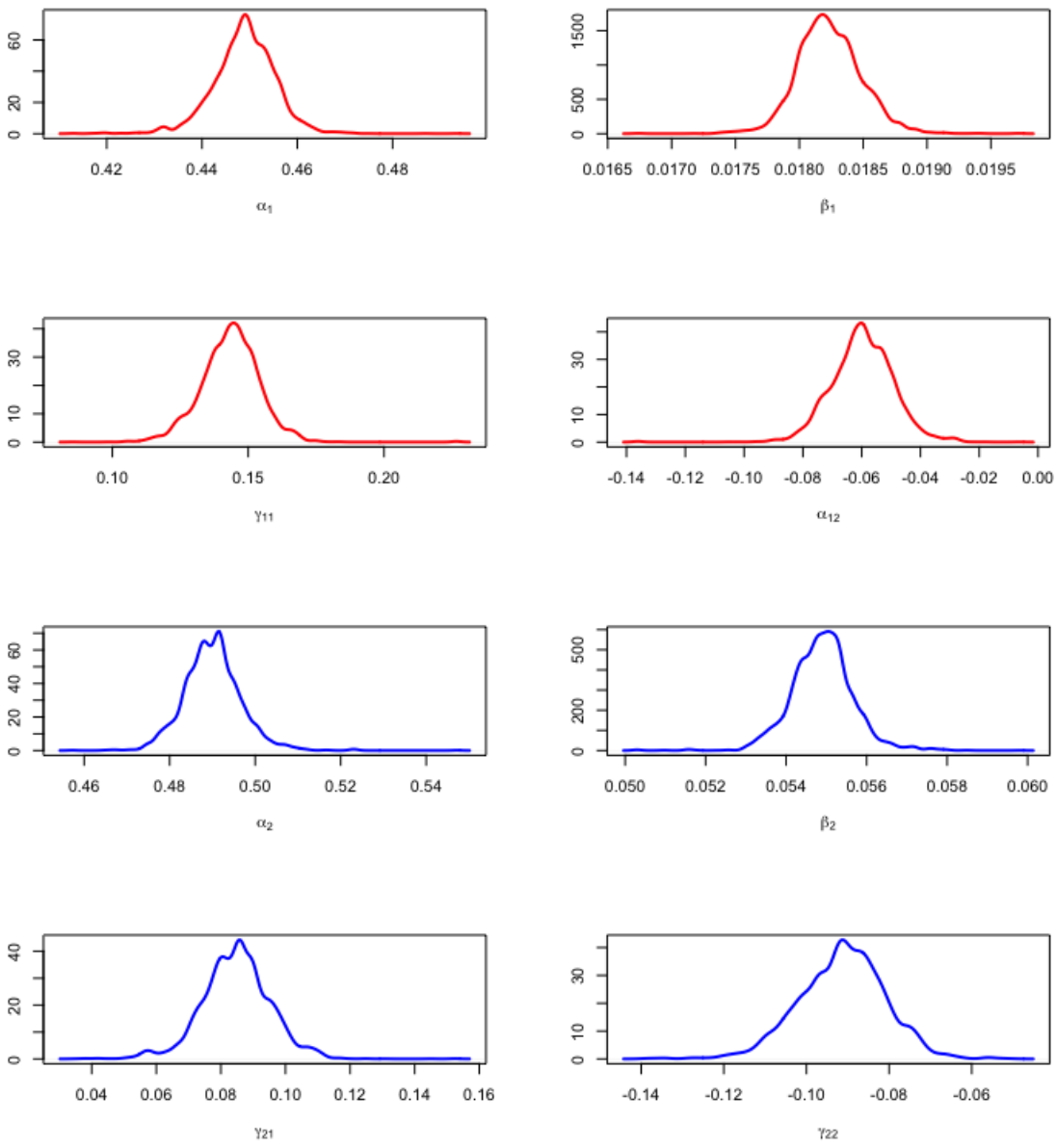


Figure 6.19: The marginal densities for the simulated draws using the Metropolis-Hastings algorithm for Power Lindley (group 1), with regression.

Chapter 7

Conclusion

In the thesis, we presented different techniques in analysing the competing risks models. In this final chapter, we discuss the significant conclusions drawn from the non-parametric and parametric techniques conducted.

Non-Parametric Estimation

In the non-parametric techniques, we used the complement of the Kaplan-Meier (KM) and empirical cumulative incidence function (CIF) to estimate the failure probability of a specific cause. Overall, the complement of the KM estimates the probability of a specific risk in the absence of all other risks. The CIF estimates the probability of a specific risk in the presence of all other risks. When these techniques were applied to a real-life data set that consisted of bone marrow transplant (BMT) patients, we found that the probabilities of relapse leukemia and death in remission were overestimated by the KM when compared to the CIF. A test of hypothesis on the CIF found that the relapse probability were not equal among the three disease groups (ALL, AML-LR, and AML-HR), where the death in remission probability were equal in the three disease groups. The results of the CIF allowed us to estimate the relapse, death in remission, and event-free probabilities at a given time in the range of the observations. The probabilities found from the CIF, as well as the KM, are only valid for the range of the observations. Thus, these non-parametric techniques cannot be used for extrapolation. To overcome this obstacle, we assumed that the risks follow parametric distributions.

Parametric Estimation

The thesis presented seven useful risk distributions which all have unique characteristics from one another. We first considered a homogeneous model in which no covariates are assumed. The methods used to estimate the parameters of these homogeneous models were the maximum likelihood method, and Bayesian method using the Metropolis-Hastings algorithm. The main piece of both estimation techniques required the use of the likelihood function. The likelihood function presented in this thesis accounted for right-censored and failure times due to a known cause of failure, as well as accounting for observations in which the cause of failure is unknown. Parameter estimation was also conducted on a cause-specific hazard regression model, in which the unknown parameters of the baseline cause-specific hazard and the unknown coefficients of the regression needed to be estimated. The parameters of the baseline cause-specific hazard were estimated using maximum likelihood and Bayesian methods, where the regression coefficients were estimated using these methods as well as Cox's partial likelihood function.

In the data analysis, we found the parameter estimates between the maximum likelihood and Bayesian methods to be consistent with one another. According to the AIC and BIC values, the data were best fitted when assuming a Power Lindley distribution for ALL and AML-HR patients and a Weibull distribution for AML-LR patients. The CIF that was the most consistent with the empirical CIF was when assuming a Power Lindley and Weibull distribution. However, the values for the AIC and BIC for the top three models did not differ significantly among each other, and thus assuming a Power Lindley, Weibull, or STH-2 can be deemed appropriate. Unlike the empirical CIF, the parametric CIF can be used to estimate failure probabilities of a specific cause at any given time. When we incorporated the covariates to the cause-specific hazard model, we estimated the coefficients of the regression portion by using a partial likelihood function and left the baseline cause-specific hazard unspecified. When we considered the fully parametric regression model, the parameters and coeffi-

cients were estimated using maximum likelihood and Bayesian methods. The results between all methods were consistent with one another. Testing the significance of the homogeneous model against the regression model found that the regression model was significant for group 1 and group 2, and not significant for group 3. Further testing found that at least one of the covariates, patient's age and donor's age, were significant for patients with ALL and AML-LR, but not for patients with AML-HR.

Limitations

The techniques in the thesis presented some complications. First, the numerical methods used to maximize the likelihood function requires a specification of a starting point. Based on the starting point, the method may fail to return the actual maximum likelihood estimates. This may occur if the starting point is within proximity to the local maxima rather than the global maxima. Secondly, the local maxima may also affect the Metropolis-Hastings algorithm. If the chain enters a local maxima, then the posterior evaluated at this point will have a higher posterior value than the points nearby, and thus the probability that this point gets accepted will be 1. The problem just mentioned can be caused by a bad choice of variance for the proposal distribution. If the variance is very small, then the chain will explore tightly around the initial value which may be near the local maxima. If the variance is very large, then the proposal will suggest poor candidate values leading to a low acceptance rate [18].

Future Work

In this thesis, the results were presented by assuming the random variables were continuous. Therefore, for future work, we will analyse the competing risks model under the assumption of discrete random variables. A competing risks model where each risk is assumed to follow a different lifetime distribution will be presented as well. For future work on the competing risks regression model, the sub-distribution model derived by Fine and Gray will be investigated [2].

Bibliography

[1] Albert, J. (2009). Bayesian Computation with R. Springer Science & Business Media.

[2] Austin, P. C., & Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Statistics in medicine, 36(27), 4391–4400.

[3] Fürstová, J. (2011). Competing Risks of CML-Related Death and Death from Other Causes. ISC Prague. p. 20-27.

[4] Ghitany, M. E., D. K. Al-Mutairi, N. Balakrishnan, L. J. Al-Enezi (2013). Power Lindley distribution and associated inference, Comput. Stat. Data Anal., 64, 20–33.

[5] Harrell, F.E. (2015). Binary logistic regression. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Cham: Springer International Publishing. p. 475.

[6] Klein, J. P. & Moeschberger, M.L. (2003). Survival Analysis: Techniques for Censored and Truncated Data. New York: Springer-Verlag.

[7] Koller MT, Raatz H, Steyerberg EW, Wolbers M. (2012). Competing risks and the clinical community: irrelevance or ignorance? Stat Med. 31:1089–1097.

[8] Lambert, P. C., Dickman, P. W., Nelson, C. P., & Royston, P. (2010) Estimating the crude probability of death due to cancer and other causes using relative survival models. Statistics in Medicine, 29(7-8), 885–895.

[9] Lau, B., Cole, S. R. Gange S. J. (2009). Competing Risk Regression Models for Epidemiologic Data, American Journal of Epidemiology, Volume 170, Issue 2, Pages 244–256,

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

[11] Pintilie, Melania. (2006). Competing Risks: A Practical Perspective. xiii.

[12] Porta, N., Gomez, G. % Luz, M.L. (2008). The Role of Survival Functions in Competing Risks.

[13] Putter, H., Fiocco, M., & Geskus, R.B. (2006). Tutorial in Biostatistics: Competing Risks and Multi-State Events. Statistics in Medicine 26, pp. 2389-2430.

[14] Raftery, A. E. (1995). Bayesian Model Selection in Social Research (With Discussion). Sociological Methodology, 25, 111-195.

[15] 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

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

[17] Sarhan, A. M., Tadj, L. & Hamilton, D. C. (2014) A New Lifetime Distribution and its Power Transformation. Journal of Probability and Statistics.

[18] van Ravenzwaaij, D., Cassey, P. & Brown, S.D. (2018) A simple introduction to Markov Chain Monte–Carlo sampling. Psychon Bull Rev 25, 143–154.