JOINT MODELING OF MULTIVARIATE LONGITUDINAL DATA AND
COMPETING RISKS DATA

by

JEEVANANTHAM RAJESWARAN

Submitted in partial fulfillment of the requirements
For the degree of Doctor of Philosophy

Dissertation Advisor: Dr. Mark D Schluchter

Department of Epidemiology and Biostatistics

CASE WESTERN RESERVE UNIVERSITY

January 2013
CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the dissertation of

Jeevanantham Rajeswaran

candidate for the Doctor of Philosophy degree*

Committee Chair:
Dr. Mark D Schluchter
Professor, Department of Epidemiology and Biostatistics

Committee:
Dr. Jeffrey M Albert
Professor, Department of Epidemiology and Biostatistics

Dr. John Barnard
Adjunct Professor, Department of Epidemiology and Biostatistics &
Staff, Department of Quantitative Health Sciences, Cleveland Clinic.

Dr. Pingfu Fu
Professor, Department of Epidemiology and Biostatistics

January, 2013

*We also certify that written approval has been obtained for any proprietary
material contained therein.
# Table of Contents

List of Tables .......................................................... iv
List of Figures .......................................................... vi
Acknowledgments ........................................................ viii
Abstract .................................................................. xi

1 Motivating Problem and Description of the Data 1
   1.1 Introduction ....................................................... 1
   1.1.1 Heart Failure ................................................. 2
   1.1.2 Heart Transplantation and Mechanical Circulatory Support .... 3
   1.1.3 Renal and Liver Functions ................................. 4
   1.2 Description of Data .............................................. 6
       1.2.1 Longitudinal assessment of Renal and Liver functions while on MCS ........................................................................ 7
       1.2.2 Transplant or Death before transplant ...................... 9
       1.2.3 Renal and Liver functions and Transplant or Death before transplant ........................................................ 10

2 Competing Risks Data ................................................. 13
   2.1 Introduction ........................................................ 13
   2.2 Single Time-to-Event: Concepts and Notations .......... 16
   2.3 Competing Risks: Concepts and Notations ................ 23
       2.3.1 Likelihood estimator ......................................... 30
       2.3.2 Nonparametric estimate of cumulative incidence function .......... 33
       2.3.3 Regression models ........................................... 35
   2.4 Data Analysis ......................................................... 40

3 Multivariate Longitudinal Data .................................... 48
   3.1 Introduction ........................................................ 48
   3.2 Univariate: Notation and Concepts .......................... 51
       3.2.1 Marginal Models ............................................. 51
       3.2.2 Random Effect Models ...................................... 52
   3.3 Multivariate: Notation and Concepts ....................... 56
   3.4 A Non-Linear Mixed Effects Model ............................ 60
       3.4.1 Multi-Phase time function ................................. 62
       3.4.2 Model components .......................................... 65
       3.4.3 Data Analysis ................................................ 67
# List of Tables

1.1  Number of deaths on MCS and Transplants in different time intervals. . . 11

2.1  Parametric estimates of cumulative incidence functions - estimated probability of observing death of MCS ($\hat{F}_{\text{Dead}}(t)$) or Transplant ($\hat{F}_{\text{Texp}}(t)$), or any one of these two ($1 - \hat{S}_{\text{any}}(t)$). ................................. 44

2.2  Cause-specific hazard regression: Regression estimates of covariate effect on the cause-specific hazards. ................................. 45

2.3  Fine and Gray’s model for regression on cumulative incidence function: Regression estimates of covariate effect on the hazards of subdistribution. . . . 47

3.1  Estimates of shaping parameters of the temporal trend of Logarithm of bilirubin. Note that, in the early phase, $m = 0$ is the limiting case of $g(t, \Theta)$ when $\nu < 0$ and $m \to 0^+$ as described in Case 3 in the mathematical formulation subsection 3.4.1, and in the late phase, $m = 0$ means the limiting case of $h(t, \Theta)$ when $\nu > 0$ and $m \to 0^+$ as described in Case 1 in the mathematical formulation subsection 3.4.1. ................................. 68

3.2  Estimates of shaping parameters of the temporal trend of GFR. Note that, in the early phase, $m = 0$ is the limiting case of $g(t, \Theta)$ when $\nu < 0$ and $m \to 0^+$ as described in Case 3 in the mathematical formulation subsection 3.4.1, and in the late phase, $m = 0$ means the limiting case of $h(t, \Theta)$ when $\nu > 0$ and $m \to 0^+$ as described in Case 1 in the mathematical formulation subsection 3.4.1. 72

3.3  Goodness of Fit using Concordance correlation. ................................. 75

3.4  Patient-specific risk factors for higher post-op Bilirubin. ................................. 77

3.5  Patient-specific risk factors for Higher post-op GFR. ................................. 78

3.6  Summary measures of the parameter estimates based on the 2000 simulated data. ................................................................. 81

3.7  Estimates of shaping and scaling parameters of the temporal trends of logarithm of bilirubin and GFR. Note that since $T_2(t, \Theta) = h(t, \Theta)$, $t_{1/2}$ is the scaler (intercept) in this phase. ................................. 93

4.1  Estimates of shape and scale parameters of Weibull baseline cause-specific hazard functions of the competing risks. ................................. 113

4.2  Estimates of scale parameters of Piecewise baseline cause-specific hazard functions of the competing risks. ................................. 114

4.3  Estimates of shape and scale parameters of baseline cause-specific multi-phase hazard functions of the competing risks. ................................. 116
4.4 ML Estimates under completely ignorable missing (Model 1), semi-joint (Model 2) and non-ignorable (Model 3) mechanisms when using Weibull baseline cause-specific hazards. ................................................................. 119
4.5 ML Estimates under completely ignorable missing (Model 1), semi-joint (Model 2) and non-ignorable (Model 3) mechanisms when using Piecewise Exponential baseline cause-specific hazards. ...................................................... 120
4.6 ML Estimates under ignorable missing (Model 1), semi-joint (Model 2) and non-ignorable (Model 3) mechanisms when using baseline cause-specific multi-phase hazards. ................................................................. 133
4.7 ML Estimates of association parameters under 3 different cause-specific multi-phase hazard models. ................................................................. 133
4.8 ML Regression Estimates of baseline covariates and the association parameters in the joint-model (non-ignorable (Model 3) mechanism) when using Weibull baseline cause-specific hazards. ................................................................. 134
4.9 ML Regression Estimates of baseline covariates and the association parameters in the joint-model (non-ignorable (Model 3) mechanism) when using piecewise exponential baseline cause-specific hazards. ................................................................. 135
4.10 ML Regression Estimates of baseline covariates and the association parameters in the joint-model (non-ignorable (Model 3) mechanism) when using multi-phase cause-specific hazards. ................................................................. 136
4.11 True value of association parameter. ................................................................. 136
4.12 Case 1: Summary measures of the parameter estimates based on the 250 Monte Carlo Simulations with two covariates in the joint model. ........ 137
4.13 Case 2: Summary measures of the parameter estimates based on the 250 Monte Carlo Simulations without covariates in the joint model. .... 138
4.14 Case 3: Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on varying sample sizes. ... 139
4.15 Case 4: Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on varying censoring rates. ... 139
4.16 Case 5: Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on varying association between random effects $b^1$ and $b^2$. ................................................................. 140
4.17 Case 6: Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on non-normal frailty. .... 141
4.18 Case 7: Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on non-normal random effect. 142
## List of Figures

1.1 *Profile plot of creatinine values while on MCS. Thick Solid line depicts trend based on smoothing splines.*  

1.2 *Profile plot of GFR while on MCS. Thick Solid line depicts trend based on smoothing splines.*  

1.3 *Profile plot of bilirubin values (logarithm) while on MCS Thick Solid line depicts trend based on smoothing splines.*  

1.4 *Longitudinal profiles of GFR and bilirubin (logarithm) and the competing risks (transplant and death on MCS).*  

1.5 *Longitudinal profiles of GFR and bilirubin (logarithm) for patients who died on MCS, grouped according to time on MCS. Thick solid lines depict the trends based on smoothing splines.*  

1.6 *Longitudinal profiles of GFR and bilirubin (logarithm) for patients who were transplanted while on MCS, grouped according to time on MCS. Thick solid lines depict the trends based on smoothing splines.*  

2.1 *Decomposition of estimated cause-specific hazards. Figure in the left depicts the decomposition of hazard of death on MCS, and the figure on the right is for the intensity of transplant.*  

2.2 *Estimated cause-specific hazards (solid lines) with pointwise 95% confidence intervals (dashed lines).*  

2.3 *Solid lines depict the parametric estimate of cumulative incidence functions and the dashed lines depict the Aalen-Johansen non-parametric estimates.*  

2.4 *Solid lines are parametric estimate of cumulative incidence functions. Dash-dash lines depict 95% confidence intervals.*  

2.5 *Regression model on Cause-specific hazards: Effect of LVAD on cause-specific hazards.*  

2.6 *Translating regression model on Cause-specific hazards to cumulative incidence: Effect of LVAD.*  

2.7 *Regression modeling on the hazard of the subdistribution: Effect of device type LVAD on the cumulative incidence functions of the competing events.*
3.1 Three cases of shapes of \( g(t, \Theta) \): in all cases \( t_{1/2} = 3 \) months. Case I: \( \nu = 0.5 \) and \( m = 0.5 \), we have an early peaking function; Case II: \( \nu = 0.5 \) and \( m = -1 \), the function starts at a finite point and decreases; Case III: \( \nu = -1.5 \) and \( m = 0.5 \), we have a decreasing function starting at infinite.

3.2 Four cases of shapes of \( h(t, \Theta) \): in all cases \( t_{1/2} = 6 \) years. Case I: \( \nu = 1.5 \) and \( m = 1.5 \), we have an early decreasing function starting at infinite; Case II: \( \nu = 0.5 \) and \( m = 0.5 \), we have a late peaking function; Case III: \( \nu = -0.75 \) and \( m = 0 \), we have a late increasing function; and Case IV: \( \nu = -1 \) and \( m = 0 \), the function is a constant phase.

3.3 Decomposition of the temporal trend of Logarithm of bilirubin.

3.4 Temporal trend of Logarithm of Bilirubin. Blue line is for average profile and the red closed circle symbols depict average actual data obtained by binning smoother.

3.5 Temporal trend of Logarithm of Bilirubin. Blue line is for average profile. Thin pink lines are actual profiles of bilirubin (Logarithmic) for 100 randomly selected patients.

3.6 Decomposition of the temporal trend of GFR.

3.7 Temporal trend of GFR. Blue line is for average profile and the red closed circle symbols depict average actual data obtained by binning smoother.

3.8 Temporal trend of GFR. Blue line is for average profile. Thin pink lines are actual profiles of bilirubin (Logarithmic) for 100 randomly selected patients.

3.9 Goodness of Fit: Observed versus Predicted for the temporal trend models (3.11) and (3.12).

3.10 True shapes of the phases for temporal trend in the simulated model.

3.11 The shapes of the early phase \( g(t, \Theta) \) and the late phase \( h(t, \Theta) \) based on the true values (blue line) and the average, 2.5% percentile and 97.5% percentile of the estimated values of the shaping parameters.

4.1 Plot on the left depicts the shapes of the baseline cause-specific Weibull hazards of the competing risks, solid lines in the plot on the right depict parametric estimates cumulative incidence functions of the competing risks, and the symbols are the Aalen-Johansen non-parametric estimates of the cumulative incidence functions.

4.2 Plot on the left depicts the shapes of the baseline cause-specific piecewise hazards of the competing risks, solid lines in the plot on the right depict cumulative incidence functions of the competing risks, and the symbols are the Aalen-Johansen non-parametric estimates of the cumulative incidence functions.

4.3 Plot on the left depicts the shapes of the baseline cause-specific multiphase hazards of the competing risks, solid lines in the plot on the right depict cumulative incidence functions of the competing risks, and the symbols are the Aalen-Johansen non-parametric estimates of the cumulative incidence functions.
4.4 True shapes of the phases for temporal trends of the bivariate longitudinal processes.
Acknowledgement

I would like to express my sincere gratitude to those who helped me complete this dissertation. It would be impossible to list the countless people who have supported throughout my carrier in statistics which started in SriLanka, and continued in Canada, Denmark and finally in USA.

First to my research advisor, Dr. Mark Schluchter, for his guidance in all aspects of this work. I truly appreciate his patience, encouragement, guidance and assistance in helping me to complete this work. I have enjoyed our discussions and comments which immensely helped me to improve my understanding of this research topic.

I would like to thank my dissertation committee members, Dr. Jeffrey Albert, Dr. John Barnard and Dr. Pingfu Fu for their time and helpful comments. Specially, the timely constructive comments during the proposal meeting.

I would like to thank my employer, the Department of Quantitative Health Science and the Cleveland Clinic for the financial support. I would like to also thank my colleagues, “The Cardio Team”, specially, to Ms. Penny Houghtaling and to Ms. Carolyn Apperson-Hansen, for their understanding, constant encouragement, support and accommodating flexible working hours.

I would like to thank Dr. Eugene H. Blackstone, the director of clinical investigations, Heart and Vascular Institute, Cleveland Clinic, for his support and constant encouragement. I, specially thank him for mentoring and teaching me the art and science of data analysis that immensely helped me to advance my carrier. Finally, more importantly, I am very much indebted to Dr. Blackstone for introducing me to the field of multi-phase decomposition modeling which is an integral part of this dissertation.

I am very much appreciate the love and support of my family. To my late parents, Jeevan and Leelawathy, thank you for your love and support. To my sister Mathi, to my late brother Navaneethan, and to my late sister Tharshini, thank you for your love and wonderful memories. Miss you guys. To my aunts and uncles, thank you for the love and support, specially, during tough times. Specially, to my uncle Navaretnarajah and aunt Rasam for the love and support and to my uncle Loganantham who encouraged
me to take a carrier in Statistics.

To my children – Leela and Jaylen, for their sacrifices, patience and unwavering constant love. You are the source of my inspiration.

Of course, nothing would be possible without the love of my life, Melissa. Thank you for being there for me in every step of my carrier, for your patience and to inspire me to go for “one more step”.

x
Abstract

by

JEEVANANTHAM RAJESWARAN

In many clinical studies that involve follow-up, it is common to observe one or more sequences of longitudinal measurements, as well as one or more time to event outcomes. A competing risks situation arises when the probability of occurrence of one event is altered/hindered by another time to event. A classical example is different causes of death. When the missing data mechanism in the longitudinal process is non-ignorable due to informative dropout, one has to jointly model the longitudinal data and the time to event outcomes in order to obtain valid inferences. Recently, there has been much attention paid to the joint analysis of single longitudinal measurements and single time to event data. A natural extension in the joint modeling literature is to the case which involves a joint analysis of multiple failure types and longitudinal process(es). However, to the best of our knowledge only a few such extensions currently exist in the literature. We, in this thesis, propose one such extension where multiple longitudinal responses are jointly modeled with competing-risks time to event data (multiple failure types). Our joint model consists of two sub-models: a system of non-linear mixed effects sub-models for the multiple longitudinal responses; and a system of cause-specific hazards frailty sub-models for competing risks, with associations among multiple longitudinal responses and multiple failure types (competing risks) are modeled using latent parameters. That is, we use the shared parameter modeling approach in this joint model. The joint model is
applied to a data set of patients with end-stage heart failure awaiting heart transplant. Heart failure occurs when heart loses the ability to pump enough blood through the body. Currently, most of the patients with end-stage heart failure who are eligible for heart transplant are inserted with a mechanical circulatory support (MCS) devices (such as left ventricular assists devices) to support the failing heart while waiting for heart transplant. While on the MCS, their liver and renal functions may worsen and these in turn may influence one of the two possible outcomes: i. death before transplant; ii. transplant. These two events can be considered as competing risks. We use longitudinal measurements of bilirubin as a marker for liver function and longitudinal measurements of GFR as a marker for renal function. By using our joint model we assess the effect (association) of liver and renal function on the competing risks.
Chapter 1

Motivating Problem and Description of the Data

In this chapter, we briefly overview the motivating problem and describe the data using non-parametric methods. We start with describing the heart failure (HF) disease and its effect on the U.S population. We then describe heart transplantation (HTx), and the need for monitoring the patients who are listed for HTx. We then describe the effect of renal and liver functions on patients before and after HTx. We briefly describe the effect of renal and liver functions on the decision making process in treating (of HF).

1.1 Introduction

Cardiovascular disease (CVD) is one of the major causes of death globally. According to the World health organization (WHO), more people die from CVD than from any other causes. In 2004, an estimated 17.1 million people died from CVD, representing 29% of all global deaths. The WHO further estimates that around 24 million people will die from CVD by 2030 around the world. CVD is caused by disorders of the heart and blood vessels. The major causes of CVD are tobacco use, physical inactivity, and unhealthy diet. CVDs include coronary heart disease, high blood pressure, peripheral artery disease, congenital heart disease, and heart failure. In addition to behavioral changes, there are several surgical and medical treatment options available to control
CVD. Surgical options include coronary bypass, valve repair or replacement, and heart transplantation.

In 2006, there were around 81.1 million affected by one or more CVDs in the U.S. ([1]). An estimated 74.5 million are affected by high blood pressure. An estimated 5.8 million are affected by heart failure. From 400,000 to 700,000 new cases are diagnosed with heart failure each year, and the frequency is expected to increase so that by 2010 there may be around 7 million affected by heart failure in the U.S. (Kouchoukos et al. [2]).

1.1.1 Heart Failure

Heart Failure (HF) occurs when the heart loses the ability to pump enough blood through the body (Kouchoukos et al. [2]). It should be noted that the heart naturally loses some of its pumping ability as one ages. But for persons with HF, the reduction in the pumping ability is much more severe, and in some cases it is life threatening when medication and lifestyle changes fail to control its symptoms. In heart failure, the heart can not provide sufficient blood supply to meet the demands of the metabolic tissue. It is normally characterized by abnormal function in other organs, such as the lung, and kidneys (Kouchoukos et al. [2]). Treatment for HF includes drug therapy or surgical interventions that are designed to improve patients’ symptoms and survival (Hunt et al. [3]). HF is a progressive disease, hence, there are several stages in HF. An estimated 200,000 patients have American College of Cardiology/American heart association stage D heart failure (Ammar et al. [4]). A stage D heart failure, also called end-stage HF, is defined as heart failure where symptoms do not respond to standard treatments (Ammar et al. [4]). Heart Transplantation is the most effective treatment for selected patients with end-stage heart failure who have exhausted all the other treatment options.
1.1.2 Heart Transplantation and Mechanical Circulatory Support

The first human heart transplant was performed by Professor Christiaan Barnard in 1967 in Cape Town, South Africa. The number of adult heart transplantations performed world-wide and reported to the International Society for Heart and Lung transplantation (ISHLT) transplant registry was 4000 patients per year in mid-1990s and around 3500 per year during the 2000s (Stehlik et al. [5]). There were 2,163 heart transplants performed in the United States in 2008 and 2,210 in 2007. People with long-term heart failure that can not be treated by any other medical or surgical means may be candidates for heart transplants. The median survival for adult and pediatric heart recipients is 10 years, with 1-year survival higher than 80% (Stehlik et al. [5]). However, despite the better prognosis after transplantation, a severe shortage of organs has made this treatment option untenable for many who are eligible. In 2009, approximately 5-10 people were on the waiting list for each of the 2200 heart transplants that were performed in the U.S. (Shreenivas et al. [6]). Therefore, there is a huge discrepancy between the number of patients who have end-stage HF and the number of donors available for heart transplantation. So, there is a long waiting time for patients who were selected for heart transplantation. Also, even if a patient had end-stage HF, some factors, such as older age, and other comorbidity conditions eliminate the patient from being eligible for heart transplant. Hence, the limitations in medical therapy to treat end-stage HF, limitations in organ availability, and the presence of large number of patients who have end-stage HF but do not qualify for heart transplant have led to the use of mechanical circulatory support (MCS) (Shreenivas et al. [6]). MCS are mainly Ventricular Assists Devices (VAD). VADs are designed to provide mechanical support to failing ventricle. Hence, VADs can support the left ventricle (LVAD), right ventricle (RVAD), or both ventricles (BiVAD). A clinical trial that compared LVAD therapy to medical therapy for
end-stage HF (REMATCH trial) showed a superior survival with acceptable safety and quality of life when using LVAD (Rose et al. [7]).

MCS is being used as either a bridge to decision (BTD) or bridge to transplant (BTT) or as a destination therapy (DT) (Shreenivas et al. [6]). MCS is used as BTD, when an end-stage HF patient who does not respond to drug therapy and is in an immediate need for MCS but eligibility criteria for listing to transplant are not met at the time of MCS implant and has to be reevaluated after he/she recovers from the acute phase of HF. That is, under BTD, it is hoped that with adequate support to the failing ventricle, the parameters for listing for transplantation can change and may eventually lead to the patient being eligible for listing for heart transplant. Some patients, because of their age and some other organ dysfunction, would not be eligible for listing for transplant at any point in time. In this scenario, MCSs are implanted as a DT. Currently, most of the MCSs are implanted as BTT, the main reason being to give support while waiting for a suitable donor.

1.1.3 Renal and Liver Functions

Renal function can be evaluated using lab values such as creatinine, blood urea nitrogen (BUN), creatinine clearance, or glomerular filtration rate (GFR). Liver (hepatic) function is evaluated using lab values such as aspartate transaminase, alanine transaminase, albumin, and total bilirubin.

Some of the initial indications for advanced heart failure are impaired renal and liver functions. Renal failure has been identified as a significant predictor for poor outcomes in patients with end-stage HF. Early experience at Cleveland Clinic and Columbia University showed that low urine output and poor liver function in the immediate pre-implant period was a predictor for poor survival after VAD implantation (Oz et al. [8], Reinhartz et al. [9]). It has also been shown that elevated serum creatinine and elevated bilirubin at the time of transplant are risk factors for long term mortality after heart
transplant (Hunt et al. [3]). In end-stage HF patient, liver dysfunction mainly points towards right ventricular failure. Sandner et al. [10] showed that renal function improved after LVAD implantation and is associated with improved survival. McCarthy et al. [11] stated that the degree of liver dysfunction caused by HF before the institution of MCS is an important factor in determining outcomes while on VADs.

The timing of heart transplantation for a patient on MCS is depend on several factors such as availability of donor hearts and device related complications. Gammie et al. [12] stated that transplantation of patient on MCS should occur when the recipient is extubated, ambulating, eating, and outside of the intensive care unit and has normalization of renal and liver function. They further stated that avoidance of transplantation early after VAD insertion yields improved survival after transplantation with optimal survival when transplanted within 2 to 4 weeks of MCS implantation. That is, the general practice is to wait for the patient to recover from acute effects of HF before transplantation. However, a recent finding by Smedira et al. [13] showed that the earlier the transplant, the better the survival on MCS and after transplant.

Therefore, timing of heart transplant is partly influenced by renal and liver functions. If the lab values indicate that there is a severe dysfunction in kidney or liver, the surgeon may wait before deciding on transplantation. On the other hand if the wait is longer, the patient’s condition may worsen by worsening renal, liver functions or other MCS related complications which are risk factors for mortality while on MCS and after transplant. In other words, awaiting full functional recovery on MCS must be balanced with the awareness that the longer duration of support, the potentially worse the pre- and post-transplant survival.

Hence, there is an interplay among renal function and liver function while on MCS and time to heart transplant and time to death before heart transplant. Death before transplant can also be termed as death on MCS.
1.2 Description of Data

The motivating data are from Cleveland Clinic’s experience in using MCS as a bridge to heart transplant. Four hundred and thirty-nine patients who were transplant candidates (listed for transplant) and were bridged with MCS from December 1991 to July 2009 at the Cleveland Clinic are considered in this data analysis. Mean age of the patients at the time of MCS implant was 53 years (with SD=12 years, min=15 years; max=74 years) and 82% were male. Time zero for all the analyses is the time of initiation of MCS. Patients who at the outset were not transplant candidates or were bridged to recovery were excluded.

Primary data were collected from the Electronic Data Interface for Transplant (EDIT) and Cardiovascular Information Registry (CVIR) databases. The databases have been approved for use in research by the institutional review board, with patient consent waived.

**Pre-MCS variables:** Variables such as demography, hemodynamics (for example, pulmonary artery pressure, cardiac index), cardiac related variable, markers of non-cardiac comorbidity (for example, levels of creatinine, total bilirubin, BUN, hematocrit), and index MCS device type were retrieved from the databases. Almost the same set of variables were collected at the time of transplant for the patients who had reached the transplant. In this data analysis, we will focus only on pre-MCS variables.

Among the 439 patients, 342 (78%) had LVAD, 76 (17%) had extracorporeal membrane oxygenation (ECMO), 12 (2.7%) had BIVAD, 3 (0.67%) had RVAD and 13 (2.9%) had Total Artificial Heart (TAH) as MCS. Pre-MCS mean creatinine was 1.7 mg/dL (range: 0.34 to 7.4 mg/dL). Note that the normal range for creatinine is between 0.8 to 1.4 mg/dL. Pre-MCS mean bilirubin was 1.7 mg/dL (range: 0.20 to 10 mg/dL). The normal range for bilirubin is between 0.30 to 1.9 mg/dL.

**During MCS variables:** Patients were closely and very rigorously followed while on
MCS. Variables such as creatinine, bilirubin, BUN systolic and diastolic blood pressure were retrieved from the databases.

1.2.1 Longitudinal assessment of Renal and Liver functions while on MCS

We will use serum creatinine as a marker for renal function and bilirubin as a marker for liver function.

**Creatinine:** A total of 12813 lab records on creatinine are available for 438 patients. Among these, 39 (9%) patients had only one record, 31 (7%) had only two records and the remaining had at least 3 records each. Mean follow-up time was 78 days (median=37 days, \(P_{25}=16\) days, \(P_{75}=88\) days).

![Figure 1.1: Profile plot of creatinine values while on MCS. Thick Solid line depicts trend based on smoothing splines](image)

Note that, these measurements are censored either at the time of transplant, at death on MCS or at the end of the follow-up. Figure 1.1 shows the profile plot of creatinine...
values while on MCS for a randomly selected subset of 100 patients who had at least 3 records.

Since it is a well known fact that Glomerular filtration rate (GFR) is a better marker for kidney dysfunction and has a well behaved distribution, we use GFR in the data analysis instead of creatinine. We estimate GFR based on The Modification of Diet in Renal Disease (MDRD) study formula [14] as follows:

$$GFR(\text{mL/min/1.73m}^2) = 186 \times \text{Creatinine}^{-1.154} \times \text{Age}^{-0.203} \times \text{Female}^{-0.299} \times \text{African American}^{0.192}.$$  

![Figure 1.2: Profile plot of GFR while on MCS. Thick Solid line depicts trend based on smoothing splines](image)

The normal range for GFR varies according to age, sex, and race; in young adults it is approximately 120-130 mL/min/1.73m² and declines with age. A persistent decrease in GFR is an indication for chronic kidney disease (CKD). When the GFR is
below 60 mL/min/1.73m$^2$, the prevalence of complications of CKD increases, as does the risk of cardiovascular disease (CVD). Figure 1.2 shows the longitudinal GFR profile for the patients who have creatinine profile as in Figure 1.1.

**Bilirubin:** A total of 9714 lab records on bilirubin are available for 436 patients. Among these, 44 (10%) patients had only one record, 35 (8%) had only two records and the remaining had at least 3 records each. Mean follow-up time was 76 days (median=34 days, $P_{25}=13$ days, $P_{75}=87$ days) Figure 1.3 shows the profile plot of bilirubin values while on MCS for a randomly selected subset of 100 patients who had at least 3 records. Since there are some extreme values in the bilirubin, it is depicted in logarithmic scale.

![Figure 1.3: Profile plot of bilirubin values (logarithm) while on MCS Thick Solid line depicts trend based on smoothing splines.](image)

**1.2.2 Transplant or Death before transplant**

As of 02/12/2010 (closing date - date of data pull) there were 118 deaths on MCS and 291 had heart transplant. Hence, there were still 30 patients alive, on MCS and waiting
for transplant as of the closing date. In this data analysis, all-cause mortality is defined as death. Hence, after the initiation of MCS, patients were at risk for experiencing either death on MCS or transplant. As we will justify in later sections, we have to treat this type of data as competing risks data.

1.2.3 Renal and Liver functions and Transplant or Death before transplant

We now descriptively investigate the longitudinal profiles and the multiple time to events. Figure 1.4 shows the longitudinal profiles for a randomly selected subset of 75 patients who had at least 2 records.

Figure 1.4: Longitudinal profiles of GFR and bilirubin (logarithm) and the competing risks (transplant and death on MCS).
From Figure 1.4, we see that for this subset of patients, it appears that lower GFR and higher bilirubin followed by death on MCS. Even though there were some early transplants with lower GFR, most of the transplants were preceded with lower bilirubin and higher GFR. We now further descriptively investigate the shape of the profiles and the competing risks. Following Li and Schluchter [15], we first group the patients based on the competing time-to-events in different time intervals as follows in Table 1.1:

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Death on MCS</th>
<th>Transplant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 month or less</td>
<td>94</td>
<td>12</td>
</tr>
<tr>
<td>1 month - ≤ 3 months</td>
<td>33</td>
<td>117</td>
</tr>
<tr>
<td>3 month - ≤ 6 months</td>
<td>14</td>
<td>91</td>
</tr>
<tr>
<td>Beyond 6 months</td>
<td>18</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 1.1: Number of deaths on MCS and Transplants in different time intervals.

Figure 1.5: Longitudinal profiles of GFR and bilirubin (logarithm) for patients who died on MCS, grouped according to time on MCS. Thick solid lines depict the trends based on smoothing splines.

Figures 1.5 and 1.6 shows the profiles of GFR and bilirubin for the patients who experienced death on MCS and transplant under four different time intervals, respectively.

Figure 1.5 clearly shows different profiles for different death groups based on time of death. The same for Figure 1.6 where different profiles observed for different death
Figure 1.6: Longitudinal profiles of GFR and bilirubin (logarithm) for patients who were transplanted while on MCS, grouped according to time on MCS. Thick solid lines depict the trends based on smoothing splines.

groups based on time of transplant. All these point to a non-ignorable mechanism for drop outs or missing data.
Chapter 2

Competing Risk Data

In this chapter, we briefly overview the literature on competing risks. We then describe and illustrate the application of two of the most commonly used approaches, namely, the cause-specific hazard approach and the cumulative incidence approach, using the data from our motivating example.

2.1 Introduction

In many situations, a patient can experience one of several possible events after a procedure or intervention. When the probability of occurrence of one event is altered by the other event(s), we call these events competing events. In other words, a competing risks situation arises when one individual exposed to more than one possible events (risks) and occurrence of one type of event hinders or alters the occurrence of other types of events. A typical example in medical science is different causes of death. Each cause of death can preclude the occurrence of other cause of death. Competing risks theory has a long history which goes back to 18th century. In 1760, Daniel Bernoulli applied Edmund Halley’s life tables method to demonstrate the advantages of smallpox inoculation. He calculated the increase in Halley’s survivor function if smallpox were eliminated as a cause of death. A complete description of his work can be found in the Appendix of
David and Moeschberger [16]. Note that in time to event analysis, the process is called “risk” before the occurrence and “event” afterwards. So, in the competing risks setup, the risks compete to be events. For example, after a cardiac surgery, a patient is exposed to two possible risks: i. reoperation; ii. death before reoperation. Hence, “risk” of reoperation and risk of ”death before reoperation” are competing until we observe either the event ”reoperation” or the event “death before reoperation”. Similarly, in our bridge-to-transplant data, a patient on mechanical circulatory support (MCS) is exposed to two possible risks: i. death on MCS; ii. transplant. After the occurrence, we can potentially observe one of these two events.

Competing risks data have been widely encountered in medical science. For example, Pollock-BarZiv et al. [17] investigated competing outcomes in infants waiting for heart transplant. They have investigated the following five possible competing outcomes: i. heart transplantation; ii. death on the transplant wait-list; iii. delisting: too sick; iv. delisting: clinically improved; and v. delisting for surgical palliation.

Note that our bridge-to-transplant data can be viewed as a more general multi-state
model as follows: i. death on MCS; ii. transplant; iii. death after transplant. Note that, here we have a two-transition state: initiation of MCS to transplant; and then from transplant to death. In this setup transplant is a transient state and death is an absorbing state. However, in competing risks setup, one only interested in the first transition states. Hence, the competing risks model can be viewed as a special case of multi-state models. In our motivating problem, our interest is only on the first transition states right after initiation of MCS (see diagram below). Hence, we are not interested in death after transplant. A brief description and an application of multi-state models can be found in Andersen and Keiding [18].

In general, competing risks data analysis focuses on answering one of the two questions: 1. Describing the time courses of different risks without focusing on any one specific risk; 2. Focusing only on one risk while treating the remaining as nuisance. Hence, in this situation, even though they are not of interest, one can not ignore the remaining risks.

The Kaplan-Meier ($K M$) method [19] is the most widely used technique to estimate the freedom from a certain outcome of interest. Hence, $1 - KM$ estimate can be interpreted as the probability of the outcome. However, in the presence of competing risks, using the usual Kaplan-Meier method to estimate the probability of occurrence of an outcome is often criticized. In fact, as we will show in later sections, it will over estimate the probability of an outcome in the presence of other competing outcomes. This problem has been pointed out in medical literature by many (for example, Arriagada et
al [20], McGiffin et al. [21], Satagopan et al. [22], Grunkemeier et al. [23]).

An earlier text book exclusively devoted to the subject of competing risks is by David and Moeschberger [16]. The more recent text books that are exclusively devoted to this subject are by Crowder [24] and Pintilie[25]. In addition to this, there are several classical books on survival analysis contain chapters on this subject, for example, Kalbfleisch and Prentice [26], and Anderson et al. [27].

2.2 Single Time-to-Event: Concepts and Notations

We now briefly discuss the basic concepts and notations in the presence of only a single outcome. Comprehensive details of the concepts, notations, and derivations can be found, for example, in Kalbfleisch and Prentice [26].

Time to event data are said to be incomplete when censoring occurs. Censoring occurs when the exact time to event is not observed for some individuals. An observation is said to be right censored if it is known that the time to event is greater than or equal to some cut-off point (normally the end of follow-up). That is, when the event of interest does not occur during their observation period, these individuals are said to be right censored. When time to event is known to be less than or equal to some cut-off point but the actual time of the event is otherwise unknown, it is said to be left censored. Note that, we may also get incomplete data when some subjects’ times to event are less than some time point (normally, start of the follow-up) and we may not observe these patients at all. This is called truncation. If we observe these subjects, these subjects are subject to delayed entry or left truncation. It should be noted that truncation is different from left censoring. In left censored data, we know the subjects, but for truncated data, we may not know the subject. Sometimes, we only know that an event occurred within a certain time interval, in this situation, it is called interval censored. Censoring mechanism can be of Type I or Type II. In the Type I mechanism,
all the individuals are followed for a fixed period of time or to a fixed calendar date of time. Type II censoring occurs when it is decided to observe the exact times of the first \( r \) events. Sometimes, censoring can happen due to some non-ignorable causes, for example, failing health or better health. This is called informative censoring when dropout is non-ignorable (as defined by Rubin [28]) and the censoring mechanism can not be assumed to be independent of the event of interest. On the other hand, non-informative censoring happens when time to censoring and time to event are independent. That is, a censoring mechanism is said to be independent, if the event (failure) rate applied to individuals at each time are the same as those that would have applied had there been no censoring.

Let \( T \) denote a non-negative random variable for the time to event, considered as continuous. The survivor function is defined as the probability that \( T \) exceeds \( t \) in its range:

\[
S(t) = P(T > t), \quad 0 < t < \infty.
\]

If \( F(t) \) is the cumulative distribution function (cdf), then \( S(t) = 1 - F(t) \). Hence, the survivor function gives the probabilities of right tail and the \( cdf \) gives the probabilities of left tail. The probability density function (pdf) is given by

\[
f(t) = \frac{-dS(t)}{dt}.
\]

Alternatively, if the pdf is known then the survivor function and the \( cdf \) can be found by integrating the \( pdf \). The hazard function is the instantaneous event rate for an individual who surviving to time \( t \) and is defined as:

\[
(2.1) \quad \lambda(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t} \right\}.
\]
It can be shown that,

\[ \lambda(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t)}{P(T \geq t) \delta t} \right\} = \frac{1}{P(T \geq t)} \lim_{\delta t \to 0} \left\{ \frac{F(t + \delta t) - F(t)}{\delta t} \right\} = \frac{f(t)}{S(t)}. \]

Hence, the survivor function can be expressed in terms of the hazard function as \( \lambda(t) = -d \log S(t)/dt \). The cumulative hazard function (or integrated hazard function) is defined as \( \Lambda(t) = \int_0^t \lambda(t)dt \). Hence, survivor function can be expressed as:

\[ S(t) = \exp [-\Lambda(t)]. \]

**Nonparametric estimates:** The survivor function can be nonparametrically estimated by the product-limit method proposed by Kaplan-Meier [19] as follows. Suppose that \( t_1, t_2, \ldots, t_k \) are observed event times such that \( t_1 < t_2 < \ldots < t_k \). Let \( d_j \) be the number of events occurs at time \( t_j \) and \( n_j \) be the number of individuals at risk just prior to time \( t_j \) then the product-limit estimator of the survivor function is given as:

\[ \hat{S}(t) = \prod_{t_j \leq t} \left( \frac{n_j - d_j}{n_j} \right). \]

It can be shown that \( \hat{S}(t) \) is the nonparametric maximum likelihood estimate of survivor function (see, for example, Kalbfleisch and Prentice [26]). It can be further shown that the asymptotic variance estimate is given by:

\[ \text{Var}(\hat{S}(t)) = \left[ \hat{S}(t) \right]^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}. \]
This formula is known as Greenwood formula (who first derived the asymptotic variance). For a large sample size, using a normal approximation, a confidence interval for \(S(t)\) can be estimated using \(\hat{\text{Var}}(\hat{S}(t))\). However, Kalbfleisch and Prentice [26] noted that the confidence interval may include values that are outside of \([0, 1]\). Hence, they applied the asymptotic normal distribution to a transformation of \(S(t)\), \(\hat{\eta}(t) = \log[-\log \hat{S}(t)]\). This yielded the following confidence interval estimate for \(S(t)\):

\[
\left[\hat{S}(t)\right]^{\exp[\pm Z_{1-\alpha/2} A]}, \text{ where } A = \frac{\left\{\hat{\text{Var}}(\hat{S}(t))\right\}}{\hat{S}(t) \log(\hat{S}(t))}
\]

An estimate of the cumulative hazard function can be obtained using the product-limit, \(\hat{S}(t)\) as \(\hat{\Lambda}(t) = -\log \left[\hat{S}(t)\right]\). Alternatively, \(\Lambda(t)\) can be estimated using the Nelson-Aalen estimator (Nelson [29], Aalen [30]), given by

\[
\hat{\Lambda}(t) = \sum_{t_j \leq t} \frac{d_j}{n_j},
\]

where \(\hat{\Lambda}(t)\) is 0 prior to the time of the first uncensored event. For large samples, the Nelson-Aalen estimator and the product-limit based estimates are similar. A variance estimator for the Nelson-Aalen estimator is given by \(\sum_{t_j \leq t} \frac{d_j(n_j-d_j)}{n_j^2}\). By assuming asymptotic normality, one can calculate a confidence interval for \(\Lambda(t)\).

**Regression models:** The effect of covariates on an event can be assessed using regression models. By assuming a certain distribution for the time \(T\), one can generate a complete parametric model. For example, under the assumption of exponential time distribution with parameter \(\lambda\), hazard at time \(t\) for an individual with covariate \(Z\) can be written as

\[
\lambda(t; Z) = \zeta(Z).
\]

Here the failure rate depends on the covariate \(Z\). Suppose the effect of \(Z\) is only through
a linear function $Z^\top \beta$, one can write

\begin{equation}
\lambda(t; Z) = \lambda \omega(Z^\top \beta),
\end{equation}

where $\lambda$ is a constant. Choice of $\omega$ depends on the data. A popular choice is $\omega(\cdot) = \exp(\cdot)$. Hence (2.2) can be written as

\begin{equation}
\lambda(t; Z) = \lambda \exp(Z^\top \beta).
\end{equation}

The model (2.3) states that the log of the failure rate is a linear function of the covariates. Similarly, assumption of Weibull time distribution with parameters $\gamma$ and $\lambda$ yields

\begin{equation}
\lambda(t; Z) = \gamma \lambda \lambda(t)^{\gamma-1} \exp(Z^\top \beta).
\end{equation}

See, for example, Kalbfleisch and Prentice [26] for details. However, one can model survival time directly using log-linear models. For example, $Y = \log T$, where $T$ is from an exponential distribution, the model (2.3) can be written as

\[ Y = \alpha - Z^\top \beta + W, \]

where $\alpha = -\log \lambda$ and error $W$ has an extreme value distribution. Hence, based on the error distribution of the log-linear model, one can deduce the distribution of the survival time. For example, normal, logistic and Gumbel error distributions lead to lognormal, logistic, and Weibull distributions for survival time, respectively. In general, a regression model on the hazard function can be written as a non-proportional hazard model as follows:
\[ \lambda(t; Z(t)) = \lambda_0(t) \exp(Z^\top(t)\beta(t)), \]

where \( \lambda_0(\cdot) \) is a baseline hazard function for continuous \( T \). This can be fully parametric, if we assume a distribution for the survival time. For example, assumption of exponential distribution yields a regression model similar to (2.3). \( Z(t) \) is a vector of time-varying and baseline covariates and \( \beta(t) \) is a vector of time-varying coefficients. The regression effect is different at different points of time. When \( \beta(t) = \beta \), the model is said to be a proportional hazard model.

The Cox proportional hazard model (Cox [31]) is the most often used model in time to event analyses. This is a semi-parametric model. Cox showed that a hazard of a subject with a given set of covariates can be decomposed into two parts: i. a part with time but not covariates; ii. a part with covariates but not time. Let \( \lambda(t; Z) \) denote the hazard function at time \( t \) for an individual with baseline covariates \( Z \). Then the Cox proportional hazards model is specified by:

\[ \lambda(t; Z) = \lambda_0(t) \exp(Z^\top \beta), \]

where \( \lambda_0(\cdot) \) is an arbitrary unspecified baseline hazard function for continuous \( T \). Cox proposed a partial likelihood approach to estimate the regression coefficients. This can be done without knowing the form of the baseline hazard, \( \lambda_0(t) \).

Blackstone et al. [32] proposed a full parametric nonproportional hazard model which is based on decomposing the cumulative hazard (hazard) into multiple overlapping phases of risk which are additive. Each phase has its own set of covariates. In hazards domain, their model is given as:

\[(2.5) \quad \lambda(t, Z) = \sum_{j=1}^{k} \mu_j(Z_j, \beta_j)(t, \Theta_j). \]
where $Z_j$ is a vector of covariates, $\beta_j$ is a vector of regression coefficients, and $\mu_j(Z_j, \beta_j)$ is a model for covariates for phase $j$. They mainly used log-linear models, $\mu_j(Z_j, \beta_j) = \exp(Z_j^T\beta_j)$. $g_j(t, \Theta_j)$ is a non-linear function of time where $\Theta_j$ is a vector of shaping parameters for the time function for phase $j$. They defined 3 phases ($k = 3$), namely, $g_1$ as an early phase, $g_2$ as a constant phase $g_3$ as a late phase and used different forms of time function for $g_1$ and $g_3$, and $g_2$ is a special case of $g_3$. Details of the time functions and estimations are given in Blackstone el al. [32] and we will discuss this time function in detail in the next chapter where we extend this multi-phase approach to model longitudinal outcomes. It can be clearly seen that the effect of $Z_j$ changes with time function $g_j(t, \Theta_j)$. Further, $Z_j$ is not influenced by the time phase $g_k(t, \Theta_k)$ for $k \neq j$. We also can have different risk factors for different phases. Hence this is a nonpropotional hazards model. Note that, when $k=1$, this model reduces to a proportional hazards model.

**Frailty models:** Frailty in a survival model is a convenient way of introducing random effects, association and unobserved heterogeneity into the models for survival data. In the case of omitted or unobserved covariates, the frailty approach is a statistical modeling concept that can account for omitted or unobserved covariates. There are two broad class of frailty models:

1. Univariate survival time models;

2. Multivariate survival endpoints models (eg. competing events, repeated events on the same individual, occurrence of a disease in relatives).

A natural way to model dependence of clustered event times is by introducing a cluster-specific random effect (frailty). This random effect explains the dependence in the sense that given the frailty, the events are independent. That is, the life times are conditionally independent. For example, using a proportional hazard model, we can define a frailty model as follows:
\[
\lambda_{ij}(t; Z) = \lambda_0(t) \exp(Z^\top \beta + w_i),
\]

where \(\lambda_0(\cdot)\) is an arbitrary unspecified baseline hazard function for continuous \(T\). \(\lambda_{ij}(t; Z)\) is the conditional hazard function for the \(j^{th}\) subject from the \(i^{th}\) cluster and \(w_i\) is the random effect of the \(i^{th}\) cluster. The above model can be written as:

\[
\lambda_{ij}(t; Z) = \lambda_0(t) u_i \exp(Z^\top \beta),
\]

where \(u_i = \exp(w_i)\) is called the frailty for the \(i^{th}\) cluster. Based on whether \(u_i > 1\) (or \(u_i < 1\)), one can interpret that the individuals in group \(i\) are frail (or strong) or higher risk (or lower risk). Note that the above model is also called a shared frailty model because subjects in the same cluster share the same frailty term (Clayton [34]). One of the main features in the frailty model is the choice of distributions for the random effect, the frailty. The most often applied distributions are the Gamma, positive stable distribution, inverse Gaussian, and the log-normal distribution. See, for example, Duchateau and Janssen [35] for further details on frailty models.

An interesting and important extension of univariate frailty is in the area of joint modeling, where longitudinal covariate information and an event time are modeled jointly. As we will discuss in a later section, the hazard process and the longitudinal process of time-varying covariates are jointly modeled using a bivariate latent process.

### 2.3 Competing Risks: Concepts and Notations

Let us suppose that there are \(D\) mutually exclusive competing events, denoted as \(\{1, 2, \ldots, D\}\). There are two main approaches in presenting a competing risks data:

1. as a bivariate random variable; 2. using a latent failure time approach.
Bivariate random variable: A usual survival data problem is normally presented as a bivariate random variable as \((T, C)\) problem, where \(T\) is the survival time, and \(C\) is a censoring variable with \(C = 1\) if the event is observed and \(0\) otherwise. We are interested in the joint distribution of \((T, C)\). This definition can be extended to the competing risks data where there is more than one event of interest. In this scenario, the data can be presented as \((T, C)\) with \(C = 0\) if the censoring occurred and \(C = i\) if event \(i \in \{1, 2, \ldots, D\}\) is observed. When \(C = i\), \(T\) is the time at which event \(i\) is observed.

Latent failure time: In the latent failure time approach, we assume an existence of \(D\) potential failure times, \(T_1, T_2, \ldots, T_D\). In the uncensored case where we observe an event, we only observe \(T = \min\{T_1, T_2, \ldots, T_D\}\) and \(C = \arg\min\limits_{i \in \{1, 2, \ldots, D\}}\{T_1, T_2, \ldots, T_D\}\), is an index variable specifies which event occurred. If an individual is censored, it is only known that all \(T_1, T_2, \ldots, T_D\) are larger than the observation and we do not know any information about the type of event. In this scenario, \(C = 0\) and a censoring distribution is introduced which is assumed to be independent of the other events. In the latent failure time approach, the data are assumed to follow a multivariate failure time model and the focus is on joint distribution of the times to \(D\) competing events. The joint survivor function is given by:

\[
Q(t_1, t_2, \ldots, t_D) = P(T_1 > t_1, T_2 > t_2, \ldots, T_D > t_D).
\]

The marginal (or overall) survivor function of the \(T = \min\{T_1, T_2, \ldots, T_D\}\) is

\[
S(t) = P(T > t) = Q(t, t, \ldots, t),
\]

because \(\min\{T_1, T_2, \ldots, T_D\} > t \Rightarrow T_j > t \ \forall j\). The marginal survivor function of event
type $j$ is given by,

$$S^j(t) = P(T_j > t) = Q(t_1 = 0, t_2 = 0, \ldots, t_j = t, \ldots, t_D = 0).$$

The fundamental problem in the joint distribution approach is that, without additional assumptions, the joint survival function is not identifiable from the observed data (single time-to-event). That is, for the same two marginal distributions, one can find more than one joint distribution. See, for example, Pintilie [p.50,25] for a counter example. That is, marginal distributions do not define the joint distribution (Tsiatis [33]). However, under the assumption of independence of $T_1, T_2, \ldots, T_D$, one can uniquely define the joint distribution. Even in this scenario, because we are observing only the first event, it is impossible to test the assumption of independence. The latent failure time approach with the assumption of independent times is vastly applied in the field of reliability where different types of component failure can be assumed to be independent. However, it is hard to justify the assumption of independence in medical science, where the event types are biologically or in some other way connected. In fact, the earlier textbook on the competing risks by David and Moeschberger [16] devotes most of it chapters to the case of independent risks. Crowder [24] approached this problem by specifying a multivariate distribution for $\{T_1, T_2, \ldots, T_D\}$. However, he notes that even for this case, since not all $T_i$ s are observed, it is hard to verify the appropriateness of the selected multivariate distribution.

The fundamental concept of the analytical approach in the competing risks data analysis is the cause-specific (or risk-specific) hazard function. This, as defined by Prentice et al. [36], is given as
\[
(2.6) \quad \lambda_j(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t, C = j | T \geq t)}{\delta t} \right\},
\]

for \( j = 1, \ldots, D \). The function \( \lambda_j(t) \) gives the instantaneous event rate for event \( j \) at time \( t \) in the presence of the other competing events. That is, \( \lambda_j(t) \) is the failure rate from a specific cause conditional on surviving up to time \( t \).

In the presence of covariate vector \( \mathbf{Z}(t) \) the above cause-specific hazard function can be written as:

\[
\lambda_j(t; \mathbf{Z}(t)) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t, C = j | T \geq t; \mathbf{Z}(t))}{\delta t} \right\}.
\]

The overall hazard function can be expressed in terms of cause-specific hazard functions as follows:

\[
\lambda(t) = \sum_{j=1}^{D} \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t, C = j | T \geq t)}{\delta t} \right\} = \sum_{j=1}^{D} \lambda_j(t).
\]

Note that different authors have used different terminology, for example, Crowder [24] called \( \lambda_j(t) \) as the sub-hazard function. Crowder [24] further defined the so-called sub-distribution function as \( F_j(t) = P(T \leq t, C = j) \), sub-survivor function as \( S_j(t) = P(T > t, C = j) \) and sub-density function as \( f_j(t) = \frac{\partial F_j(t)}{\partial t} \). The sub-survivor function can be interpreted as the probability that an event \( j \) does not occur by time \( t \). It is worth noting that \( S_j(t) \) is not equal to the probability that \( T > t \) for the event type \( j \). That probability is a conditional one and is written as \( P(T > t | C = j) = \frac{S_j(t)}{P(C = j)} \).

Note that the overall distribution function \( F(t) \) reaches 1 as \( t \to \infty \). However, in the
presence of competing risks, $F_j(t) \to p_j$ as $t \to \infty$, where $p_j = P(C = j)$ with $p_j > 0$ and $\sum_1^D p_j = 1$. It is further noted, $F_j(t) + S_j(t) = P(C = j)$ as $t \to \infty$. Hence, $F_j(t)$ is not a proper distribution function. Thus the terms sub-distribution, sub-hazard and sub-survivor. Note that the most familiar term in the medical literature for the sub-distribution function is cumulative incidence function. Some called this function as the crude cumulative incidence function or crude transition probability (see for example, Crowder [24], Putter et al. [37]).

The overall survivor function is given by $S(t) = P(T > t) = \sum_1^D P(T > t, C = j) = \sum_1^D S_j(t)$. Likewise, the marginal density is given by $f(t) = \sum_1^D f_j(t)$. Here the marginal survivor function $S(t)$ is interpreted as the probability of not having experienced any event at time $t$. We now write the cause-specific hazard function in terms of marginal survivor and sub-density functions,

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

$$= \frac{1}{P(T \geq t)} \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t, C = j)}{\delta t} \right\}$$

$$= \frac{f_j(t)}{S(t)} \quad (2.7)$$

We now express the cumulative incidence function for the event $j$ in terms of marginal survivor and cause-specific hazard functions as follows:

$$F_j(t) = P(T \leq t, C = j)$$

$$= \int_0^t f_j(u) du$$

$$= \int_0^t \lambda_j(u) S(u) du, \quad (2.8)$$
where (2.8) was obtained by substituting (2.7). $F_j(t)$ is the probability of experiencing event type $j$ before time $t$. In the multi-state transition probability terminology, $F_j(t)$ can also be viewed as the transition probability of an individual in state 0 (alive) at time 0 will in state $j$ at a time $t$. It is noted that, since $S(t) = \exp\{-\sum_{j=1}^{D} \int_0^t \lambda_j(u)du\}$, the cumulative incidence function for the event type $j$ is dependent on the cause-specific hazards for all the event types. Hence, the derivation of the cumulative incidence function for event type $j$ depends (or involves) on other event types.

A conditional probability of interest that incorporates both the event of interest and the competing risks is the probability of observing an event of interest by time $t$ given a subject did not experience a competing event. Pepe and Mori [38] called this the conditional incidence function, defined as:

$$
CP_j(t) = P(T \leq t, C = j|\text{no other type of events by time } t).
$$

$$
= P(T \leq t, C = j|T \leq t, C \neq j)^C)
$$

$$
= \frac{P(T \leq t, C = j)}{1 - P(T \leq t, C \neq j)}
$$

$$
= \frac{F_j(t)}{1 - \sum_{k \neq j} F_k(t)}.
$$

The conditional incidence function essentially tries to answer the question of the probability of an event of interest, when one removes the other causes. Another conditional probability of interest is the probability of experiencing event of interest $j$ given one has
experienced an event before time $t$. This can be written as

$$CP^F_j(t) = P(C = j | T \leq t).$$

$$= \frac{P(T \leq t, C = j)}{P(T \leq t)}$$

$$= \frac{F_j(t)}{\sum_k F_k(t)}. $$

It is noted that both conditional probabilities are functions of the cumulative incidence functions.

A naive estimator for the cumulative incidence function can be obtained by using the Kaplan-Meier estimate or some parametric survival function $\hat{S}^*_j(t)$, where the subjects who have experienced events other than event type $j$ are censored. We then can use the $1 - \hat{S}^*_j(t)$ as the probability of experiencing an event type $j$ before or at time $t$. This can be expressed as,

$$1 - S^*_j(t) = \int_0^t f_j(u)du$$

$$= \int_0^t \lambda_j(u)S_j(u)du. $$

(2.9)

The difference between equation (2.8) and (2.9) is that $S(t)$ is replaced by $S_j(t)$. Since $S(t)$ is the probability of not experiencing any events and $S_j(t)$ is the probability of not experiencing event type $j$, we have $S(t) \leq S_j(t)$. Thus $F_j(t) \leq 1 - S^*_j(t)$, with equality at $t$ if there is no competing events. Hence, there is a bias in the naive Kaplan-Meier estimator when using it as a cumulative incidence function.
2.3.1 Likelihood estimator

We now investigate the likelihood function and its relationship to the cause-specific hazard function as described by Prentice et al. [36]. Suppose there are \( n \) subjects with data \((t_i, j_i, \delta_i, z_i^*)\), \( i = 1, \ldots, n \), where \( t_i \) is the event time or censored time, \( j_i \in \{1, \ldots, D\} \) is the event type, \( \delta_i \) is the censoring indicator with 1 if event observed and 0 otherwise, and \( z_i^* = z_i^*(t) = \{z(u); u \leq t\} \) is a vector of covariate information for the \( i^{th} \) subject.

For notational convenience, we use \( z_i \), where \( z_i = z_i^*(t) \). Note that when \( \delta_i = 0 \), \( j_i \) can be arbitrarily specified. Under non-informative censoring the likelihood function can be written as the product of sub-densities and sub-survivor functions. A censored subject contributes the probability of being not experiencing any events at time \( t_i \) and a subject with an observed event (type \( j_i \)) at time \( t_i \) contributes the sub density of event type \( j_i \).

Suppose there are \( o \) cases for which events have been observed and \( c \) cases were censored. That is, \( n = o + c \). Then the likelihood can be written as,

\[
\mathcal{L} = \prod_{i=1}^{o} f_{j_i}(t_i; z_i) \Pi_{c} S(t_i; z_i).
\]

(2.10)

\[
= \prod_{i=1}^{o} \{\lambda_{j_i}(t_i; z_i)S(t_i; z_i)\} \Pi_{c} S(t_i; z_i).
\]

(2.11)

\[
= \prod_{i=1}^{n} [\lambda_{j_i}(t_i; z_i)]^{\delta_i} S(t_i; z_i),
\]

where we have used equation (2.7) to obtain equation (2.10). Now using the fact that, \( S(t) = \exp[-\Lambda(t)] = \exp[- \int_0^t \lambda(u)du] = \exp[- \int_0^t \sum_{j=1}^D \lambda_j(u)]du \), equation (2.11)
can be written as

\[ L = \prod_{i=1}^{n} \left[ \{ \lambda_{ji}(t_i; z_i) \}_i \delta_i \exp\left\{ - \int_{0}^{t_i} \sum_{j=1}^{D} \lambda_j(u; z_i)du \right\} \right] \]

\[ = \prod_{i=1}^{n} \left[ \{ \lambda_{ji}(t_i; z_i) \}_i \delta_i \exp\left\{ - \int_{0}^{t_i} \lambda_j(u; z_i)du \right\} \right] \]

\[ = \prod_{j=1}^{D} \left[ \Pi_{i: j_i = j} \lambda_j(t_i; z_i) \left\{ \Pi_{i=1}^{n} \exp\left\{ - \int_{0}^{t_i} \lambda_j(u; z_i)du \right\} \right\} \right] \]

(2.12)

Now let,

\[ L_j = \prod_{i: j_i = j} \lambda_j(t_i; z_i) \Pi_{i=1}^{n} \exp\left\{ - \int_{0}^{t_i} \lambda_j(u; z_i)du \right\}, \]

(2.13)

where the first product in equation (2.13) taken over the set of subjects who experienced event type \( j \). Then the whole likelihood is written as,

(2.14) \[ L = L_1 L_2 \ldots L_D. \]

Equations (2.12), (2.13), and (2.14) give two important results:

1. The overall likelihood is a product of \( D \) likelihoods, one for each event type. Further the overall likelihood can be completely specified by the cause-specific hazard functions \( \lambda_j(t; z) \);

2. If there are non-overlapping parameter sets among different cause-specific hazard models, we can maximize each likelihood separately. Hence, numerically, it may be easier to maximize each likelihood separately, as there are fewer parameters. So, if one is interested in, say, event type \( j \), the parameters associated with \( \lambda_j(t; z) \) can
be estimated and examined without reference to other competing events. In fact, as Prentice et al. [36] describe, the likelihood factor for $\lambda_j(t; z)$ is exactly same as would be obtained by regarding all event types other than $j$ as censored at their time of experiencing the other events. They further note that, this provides the formal justification for the most used approach in the estimation of $\lambda_j(t; z)$, where events from other event types are regarded as censored when studying factors that affect a certain event of interest. See, for example, Crowder [24] and Prentice et al. [36] for further discussion on this topic.

Note that, the cause-specific hazard function is defined in terms of the latent failure time approach as follows; since, under the condition that overall survival time $T \geq t$, $t \leq T_j < t + \delta t$ is equivalent to $t \leq T < t + \delta t$ and $C = j$

$\lambda_j(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T_j < t + \delta t | T \geq t)}{\delta t} \right\}$

$= \frac{1}{P(T \geq t)} \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T_j < t + \delta t)}{\delta t} \right\}$

$= \frac{1}{Q(t, \ldots, t)} \frac{\partial F_j(t)}{\partial t}$

$= \frac{1}{Q(t, \ldots, t)} \frac{-\partial Q(t_1, t_2, \ldots, t_D)}{\partial t_j}$

$= -\frac{\partial \log Q(t_1, t_2, \ldots, t_D)}{\partial t_j} \bigg|_{t_1 = t_2 = \ldots = t_D = t}.$

That is, the cause-specific hazard is equal to the negative of partial derivative of the log of joint survivor function w.r.t $t_j$ evaluated at $t_1 = t_2 = \ldots = t_D = t$.

Because the likelihood entirely depends on the cause-specific hazards, only these hazards or functions of these can be estimated, other quantities are unidentifiable. Therefore, based on (2.15), only the diagonal derivatives of $\log Q$ will be estimable, not the
off diagonals, $\frac{\partial^2 \log Q}{\partial t_j \partial t_k}$. That is, when we observe only the minimum of the latent times, one can not distinguish between independent competing risks. Thus, infinitely many dependent risks produce the same cause-specific hazard.

### 2.3.2 Nonparametric estimate of cumulative incidence function

Let $t_1 < t_2 < \ldots < t_r$ be the unique ordered event times. Define $d_{ij}$ be the number of events of type $j$ that occurred at time $t_i$ and $n_i$ be the number at risk just prior to time $t_i$, $i = 1, \ldots, r$. Further let, $\hat{S}(t)$ be the Kaplan-Meier estimator of the probability of being free of all the events by time $t$. Then it can be shown that the nonparametric maximum likelihood estimate of the cumulative incidence function for the event $j$ is given by

$$
\hat{F}_j(t) = \sum_{\forall i, t_i \leq t} \frac{d_{ij}}{n_i} \hat{S}(t_{i-1})
$$

(2.16)

Note that the equation (2.16) is known as the Aalen-Johansen estimator (Aalen and Johansen [39]). Hence, the Aalen-Johansen estimator is obtained by substituting the Kaplan-Meier estimate for the survivorship function and the Nelson-Aalen estimator for the cause-specific hazard function in (2.8). As was discussed in the introduction, if one is tempted to use the naive estimator $1 - KM_j(t)$ for the event type $j$ as the cumulative incidence function of event type $j$ at time $t_k$, we have
\[
KM_j(t_k) = \prod_{i=1}^{k} \frac{n_i - d_{ij}}{n_i}
= \frac{n_k - d_{kj}}{n_k} KM_j(t_{k-1})
= KM_j(t_{k-1}) - \frac{d_{kj}}{n_k} KM_j(t_{k-1}).
(2.17)
\]

Now subtracting both sides of (2.17) from 1, we have

\[
1 - KM_j(t_k) = 1 - KM_j(t_{k-1}) + \frac{d_{kj}}{n_k} KM_j(t_{k-1}).
(2.18)
\]

Now using the recursive equation (2.18), we obtain

\[
1 - KM_j(t_k) = \sum_{i=1}^{k} \frac{d_{ij}}{n_i} KM_j(t_{i-1}).
(2.19)
\]

The difference between equations (2.16) and (2.19) is \(\hat{S}(t_{i-1})\) which is the probability of not experiencing any events and \(KM_j(t_{i-1})\) is the probability of not experiencing event type \(j\). Hence clearly, \(KM_j(t_{i-1}) \geq \hat{S}(t_{i-1})\). Thus \(1 - KM_j(t) \geq F_j(t)\) with equality at \(t\) if there is no competing events (that is, if there is no competition). That is, using \(1 - KM\) as the estimator of the cumulative incidence function overestimates the cumulative probability of event type \(j\). The reason behind this bias is that the Kaplan-Meier estimate for an event of interest is estimated under the assumption that the censoring distribution and the event distribution are independent. That is, the hazard
of an event of interest is the same for a subject who has not experienced the event and still under follow-up and a subject who had experienced the event. However, in the presence of competing events, the Kaplan-Meier estimate for an event of interest is estimated by treating the subjects who experienced other events as the same as censored subjects. Clearly, because of the competition, the hazard of an event of interest is different for a subject who experienced other events and a subject who experienced the event of interest.

2.3.3 Regression models

The most widely used regression modeling approaches in the analyses of competing risks data in medical research are

1. regression on cause-specific hazard functions

2. regression on cumulative incidence functions

The choice between two approaches should be determined by the research questions. In some situations, it is wise to investigate both functions. Simply speaking, if we are interested in the number of subjects who experience a certain event and its determinants, we should use the cumulative incidence function of that event. On the other hand, if we are interested in the risk of experiencing a certain event and its determinants, we should use the cause specific hazard function. In some instances, it is better to perform both approaches, so that it is easy to check whether the inference conceived by the cumulative incidence function also holds for the cause specific hazard functions. In terms of sample size, while cumulative incidence functions can be drawn based on relatively few events, cause specific hazard functions need larger number of events to identify time trends.

**Regression on cause-specific hazard functions**: When investigating the effects of covariates on a rate of an event of interest, one can use the classical survival regression methods such as log-rank or Cox-type models. These methods ignore the other events
(competing events) and test the “pure” effect. That is, cause-specific hazards models an event of interest in the absence of other competing events. Hence, in this approach, we test rate of an event of interest in the virtual situation when other competing events did not exist. Thus the cause-specific hazard approach is beneficial, for example, when the goal is to test whether a specific factor has a biological relevance. In this regression setup, each time a subject experiences the event of interest \( j \), the covariate values of this subject are compared with the covariates of all other subjects still event-free and in follow-up. Subjects experiencing events other than \( j \) are censored at the time of their respective events. We can model the cause-specific hazard for event \( j \) for a subject with covariates \( Z \) as follows:

\[
\lambda_j(t; Z(t)) = \lambda_{j,0}(t) \exp(Z^\top(t)\beta_j),
\]

where \( \lambda_{j,0}(t) \) is the baseline cause-specific hazard for event type \( j \), and the vector \( \beta_j \) is the vector of covariate effects on event type \( j \). The analysis can be performed using standard software. A proportional risk model can be obtained when the following model is used,

\[
\lambda_j(t; Z(t)) = \lambda_0(t) \exp(\zeta_j + Z^\top(t)\beta_j).
\]

In (2.21), the baseline hazard functions are restricted to be proportional to each other competing events with proportionality factor \( e^{\zeta_j} \), with, for uniqueness, \( \zeta_1 = 0 \). That is, the cause-specific hazards are assumed to have the same shape. The assumption of proportionality among the competing risks can be tested using (2.20).
Suppose that, $Z$ is a time independent covariate vector then this model has an interesting property that

$$(2.22) \quad P(D = j | Z) = \frac{\exp\{\zeta_j + Z^\top \beta_j\}}{\sum_{l=1}^D \exp\{\zeta_l + Z^\top \beta_l\}}, \quad j = 1, \ldots, D$$

regardless of $\lambda_0(\cdot)$. That is, the probability that a subject with fixed covariates $Z$ has the event type $j$ is $P(D = j | Z)$. The equation (2.22) can be obtained using (2.8) and $P(C = j) = \lim_{t \to \infty} F_j(t)$.

Hence, using cause-specific hazard regression approach is straightforward and somewhat simple. However, transforming the cause-specific regression functions, say, by using (2.8), to obtain cumulative incidence functions always yields a fairly complicated non-linear function of regression coefficients and the effect of covariates on the cumulative incidence functions are not described by simple parameters. For example, when we translate cause-specific Cox proportional hazard models into cumulative incidence functions, the proportionality is lost and the covariate effects on the cumulative incidence functions are no longer expressed in simple formats. With this in mind a number of models have been proposed to model the cumulative incidence function directly.

**Regression on cumulative incidence functions:** Fine and Gray [40] proposed a model to allow regression directly on the cumulative incidence functions. In that they proposed a regression model on the hazard of subdistribution. In analogy to $\lambda(t) = -d \log S(t)/dt$ in a single time-to-event setup, the hazard of subdistribution based on the cumulative incidence (subdistribution) function is defined by Gray [41] as follows:
Gray then used this function and introduced a k-sample test to compare the weighted averages of the hazard of the subdistribution for the event of interest. The hazard of subdistribution is defined as the probability of observing an event of interest in \([t, t + \delta t)\) given that either the event of interest did not happen before \(t\) or a competing event was observed.

Fine and Gray [40] defined a “Cox-type” regression model on the hazard of subdistribution as follows:

\[
\gamma_j(t) = \lim_{\delta t \to 0} \frac{P(t \leq T < t + \delta t, C = j|T \geq t \text{ or } (t \leq T\&C \neq j))}{\delta t} = \frac{f_j(t)}{1 - F_j(t)} = - \frac{d \log(1 - F_j(t))}{dt}.
\]

In this modeling approach, there are two important points of interest that are different from the usual Cox-type regression approach: (i). The risk set is defined differently; (ii). inclusion of weights based on the censoring distribution is used in the “Cox-type partial” likelihood for regression coefficient estimation.

The risk set is defined as

\[
R_j(i) = \{i : T \geq t \text{ or } (t \leq T\&C \neq j)\}.
\]

The risk set is formed of those who did not experience the event type \(j\) and those who experienced a competing event by time \(t\). That is, in addition to subjects who
did not experience the event, subjects who experienced another competing event also remain in the risk set. That is, if there is no censoring, once these subjects are given a censoring time that is larger than all event times of event type \( j \), they remain in the risk set forever. However, for the subjects who experienced another competing event before time \( t \), an inverse probability weighting of the censored observations is used in the estimation procedure. Once we have defined the modified dataset, one can simply perform a standard Cox-type regression.

Fine [42] extended the Fine and Gray model (2.23) to more general transformations as follows:

\[
g(F_j(t)) = h_{jo}(t) - Z^\top \beta_j.
\]

Note that (2.23) is obtained by using \( g(x) = \log(-\log(1-x)) \).

Another interesting modeling approach that modeled cumulative incidence function (probability) of an event of interest directly was proposed by Andersen et al [43]. They used pseudo-values from jacknife statistics constructed from cumulative incidence functions. Suppose \( \hat{F}_j(t) \) is the Aalen-Johansen estimator based on the entire sample (of size \( n \)) and \( \hat{F}_j^{-i}(t) \) is based on the sample with subject \( i \) removed. They defined the pseudo-observation for subject \( i \) at time \( t \) as

\[
\hat{F}_j^i(t) = n\hat{F}_j(t) - (n-1)\hat{F}_j^{-i}(t).
\]

They then used these observations and proposed the following generalized linear model

\[
g\{E(\hat{F}_j^i(t))\} = \alpha_j(t) + Z_i^\top \beta_j,
\]

where \( g \) is a link function. For example, \( g(x) = c \log \log x = \log(-\log(1-x)) \) or \( g(x) = \logit(x) \). Regression coefficients were estimated using GEE.
Larson and Dinse [44] proposed a mixture modeling approach for regression analysis of competing risks data. The joint model of event time and event type \((T, C)\) is modeled as the mixture of the probability of experiencing event type \(j\), \(P_j(Z)\), and the conditional survivor function \(Q_j(t|Z)\) where

\[
P_j(Z) = P(D = j|Z) = \frac{\exp(Z^\top \eta_j)}{\sum_{l=1}^D \exp(Z^\top \eta_l)},
\]

is a logistic regression and

\[
Q_j(t|Z) = P(T > t|D = j, Z) = \exp[- \int_0^t h_{jo}(x) \exp(Z^\top \beta_j)],
\]

where \(h_{jo}(x)\) is the null \(Z = 0\) hazard function for event type \(j\).

**Remark:** Parameter estimation for the above mentioned models was carried out using readily available estimation approaches such as, Newton-Raphson, or GEE methods. Ng and McLachlan [45] proposed an EM algorithm approach to estimate the parameters. The complete data for the EM algorithm approach is obtained by introducing \(z_{ij}\) for the censored subjects where \(z_{ij} = 1\) or \(0\) according as the \(i^{th}\) subject would have experienced event type \(j\) or not.

## 2.4 Data Analysis

We now perform a competing risk analysis using the bridge-to-transplant data described in section 1. We have 439 patients with 118 observed deaths on MCS and 291 had transplants. For simplicity, we focus on the following four baseline covariates: Age at the time of initiation of MCS; type of MCS (LVAD vs. other); baseline GFR; baseline bilirubin. We first estimate cause-specific hazards and cumulative incidence functions without covariates, for the competing events using parametric and non-parametric methods described above. We use the parametric model (2.5) proposed by Blackstone et al. [32] to
estimate the *cause-specific hazards*. We then use these estimates in (2.8) to estimate the *cumulative incidence functions*. We further use the *Aalen-Johansen* estimator (2.16) to estimate the non-parametric *cumulative incidence functions*. The multi-phase hazard modeling of Blackstone et al. [32] is implemented using a SAS macro interfaced with Fortran programming. See http://www.clevelandclinic.org/heartcenter/hazard for details of the programming and downloads. We have used the *cuminc* function in R-package *cmprsk* [46] to estimate the *Aalen-Johansen* estimator for the *cumulative incidence functions*.

![Figure 2.1: Decomposition of estimated cause-specific hazards. Figure in the left depicts the decomposition of hazard of death on MCS, and the figure on the right is for the intensity of transplant.](image)

**Cause-Specific Hazards:** We have used parametric modeling (2.5) to obtain cause-specific hazards for transplant and death on MCS. Figure 2.1 shows the decomposition of cause-specific hazards. Briefly, parametric modeling involves maximum likelihood estimation of the scaling, $\mu(\cdot)$ and shaping parameters of the non-linear function of time $g(\cdot)$ in (2.5). Starting values of these parameters are obtained from the respective Kaplan-Meier estimates. See http://www.clevelandclinic.org/heartcenter/hazard for details of the modeling approach. Decomposition of the hazard for death on MCS ($\lambda_{\text{Dead}}$) yielded an early peaking phase and a constant phase, the hazards (or the intensity) of
transplant ($\lambda_{T\text{xp}}$) yielded an intermediate peaking phase and a constant phase. The estimated cumulative hazard models ($\Lambda(t)$) are:

\[
\hat{\Lambda}_{\text{Dead}}(t) = 0.275 \left[ \exp\left\{-\left(\frac{t}{0.087}\right)^{-1/1.72}\right\} \right] + 0.367t,
\]

and

\[
\hat{\Lambda}_{T\text{xp}}(t) = 2.25 \left[ 1 + \left(\frac{t}{0.484}\right)^{-1/0.56} \right]^{-1} + 0.131t.
\]

Note that, the estimated cause-specific hazards ($\hat{\lambda}_{\text{Dead}}$, and $\hat{\lambda}_{T\text{xp}}$) are obtained by differentiating (2.24) and (2.25) w.r.t $t$.

Estimated cause-specific hazards are given in Figure 2.2. While the risk for death on MCS ($\hat{\lambda}_{\text{Dead}}$) peaked at around 2 weeks after the initiation of MCS, the intensity of transplant ($\hat{\lambda}_{T\text{xp}}$) peaked at around 3 months after the initiation of MCS.

![Figure 2.2: Estimated cause-specific hazards (solid lines) with pointwise 95% confidence intervals (dashed lines).](image)

**Cumulative Incidence Functions**: We use equation (2.8) and the estimates (2.24) and (2.25) to estimate the cumulative incidence functions, $F_{\text{Dead}}(t)$ and $F_{T\text{xp}}(t)$. Kaplan-
Meier non-parametric estimates of freedom from experiencing any one of the events (that is, neither death on MCS nor transplant) at 6-months and at 1 year are 20% and 8%, respectively. Hence, the probability \( (1 - \hat{K}M_{\text{any}}(t)) \) of experiencing one of these two events are 80% and 92% at the respective times. However, the naive estimate of the probability of death on MCS, \( (1 - \hat{K}M_{\text{Dead}}(t)) \) and the naive estimate of probability of transplant, \( (1 - \hat{K}M_{\text{Txpl}}(t)) \) at 6-months and 1-year are 31% and 44%; and 71% and 80%, respectively. This clearly shows that \( 1 - \hat{K}M_{\text{any}}(t) < \left[ 1 - \hat{K}M_{\text{Dead}}(t) \right] + \left[ 1 - \hat{K}M_{\text{Txpl}}(t) \right] \). This solidifies our assertion that occurrence of one event can affect or hinder the occurrence of the other event and hence, these two events are needed to be considered as competing events. Figure 2.3 shows parametric estimates of cumulative incidence functions superimposed on the Aalen-Johansen non-parametric estimates. The fact that the estimated parametric curves and the estimated non-parametric curves are nearly on top of each other implies that the goodness-of-fit of the parametric multi-phase hazard modeling is very good.

![Figure 2.3: Solid lines depict the parametric estimate of cumulative incidence functions and the dashed lines depict the Aalen-Johansen non-parametric estimates.](image)

Table 2.1 shows the estimates of cumulative incidence function at 3, 6-months and 1 year. It is clearly shown that the probability observing any type of event is partitioned
into two probabilities. That is, \( 1 - \hat{S}_{\text{any}}(t) = \hat{F}_{\text{Dead}}(t) + \hat{F}_{\text{Txpl}}(t) \).

<table>
<thead>
<tr>
<th>Time (t)</th>
<th>( \hat{F}_{\text{Dead}}(t) )</th>
<th>( \hat{F}_{\text{Txpl}}(t) )</th>
<th>( 1 - \hat{S}_{\text{any}}(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>19</td>
<td>26</td>
<td>85</td>
</tr>
<tr>
<td>6 month</td>
<td>23</td>
<td>57</td>
<td>80</td>
</tr>
<tr>
<td>1 year</td>
<td>26</td>
<td>66</td>
<td>92</td>
</tr>
</tbody>
</table>

Table 2.1: *Parametric estimates of cumulative incidence functions - estimated probability of observing death of MCS (\( \hat{F}_{\text{Dead}}(t) \)) or Transplant (\( \hat{F}_{\text{Txpl}}(t) \)), or any one of these two (\( 1 - \hat{S}_{\text{any}}(t) \)).*

Figure 2.4 shows parametric estimate of cumulative incidence functions. From Figure 2.4, it can be observed that as the probability of death on MCS and probability of transplant increases, the probability of being free of any event decreases. This shows how the transition occurred from event free state to experiencing one of the events. It is also noted that, during the first month, the probability of death is higher than the probability of transplant.

*Regression modeling:* First, we use the parametric model (2.5) to model the cause-specific hazards (Blackstone et al [32]) to study the effect of the selected covariates on the competing events. We then use Fine and Gray’s model [40] (2.23) to study the effect of covariates directly on the hazard of the subdistribution (*cumulative incidence*...
Fine and Gray’s regression model [40] is implemented using the `crr` function in the R-package `cmprsk` [46].

Table 2.2 shows the cause-specific hazard regression estimates for death of MCS and Transplant.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Covariate</th>
<th>Death on MCS</th>
<th>Transplant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estimate ± SE</td>
<td>P-value</td>
</tr>
<tr>
<td>Early</td>
<td>Age</td>
<td>1.1 ± 1.1</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>LVAD</td>
<td>-1.9 ± 0.504</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>GFR</td>
<td>-0.58 ± 0.45</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Bilirubin</td>
<td>-0.015 ± 0.16</td>
<td>&gt;0.9</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>Age</td>
<td>0.52 ± 0.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LVAD</td>
<td>0.86 ± 0.703</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GFR</td>
<td>-0.38 ± 0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bilirubin</td>
<td>0.13 ± 0.099</td>
</tr>
</tbody>
</table>

Table 2.2: Cause-specific hazard regression: Regression estimates of covariate effect on the cause-specific hazards.

LVAD appears to have a protective effect on the rate of early deaths on MCS (Figure 2.5). It also appears at least marginally, that higher the bilirubin, higher the intensity of early transplants and lower the late transplants. The higher the bilirubin, the higher the risks of death on MCS and lower the intensity of transplant. In the cause-specific hazard approach, the whole analyses of a certain event are performed as if the other types of events did not exit. Therefore, there may be some internal relationship between LVAD.
and risk of death. Higher bilirubin means sicker patients. Therefore, higher intensity of early transplants means sicker patients are getting transplant sooner at the early stages after the initiation of MCS. However, in the later stages (say after 6 months), the intensity is lower for the sicker patients. That is, higher bilirubin forces the transplant to be postponed until the patient gets stabilized.

Figure 2.6: Translating regression model on Cause-specific hazards to cumulative incidence: Effect of LVAD.

Figure 2.6 is obtained by translating the cause-specific hazard regression estimates to cumulative incidence function by using equation (2.8).

We now perform regression modeling directly on the hazard of the subdistribution using Fine and Gray’s model [40] (2.23). Table 2.3 shows the estimates of the regression coefficients. While lower GFR and having non-LVAD device are appeared to have an adverse effect on the death on MCS, the effects are in the opposite direction for the intensity of transplant. This basically states that these two competing events are negatively correlated and can not be regarded as independent events.

If we had just focused regression modeling on the hazard of the subdistribution, we might not know which one of the following is actually the true inference for the effect of LVAD (Figure 2.7): LVAD has a protective effect on the death on MCS or it increases the intensity of the transplant or both. But by combining the inferences from both modeling
Table 2.3: *Fine and Gray’s model for regression on cumulative incidence function: Regression estimates of covariate effect on the hazards of subdistribution.*

 approaches, it can be safely said that LVAD has a protective effect on the death on MCS and thus patients on LVAD survive to reach the transplant. On the other hand, it may not be clear, if the patients with higher GFR (normal cases) are getting the transplant and thus the remaining cases (lower GFR) are at more risk for death on MCS. Note that the effect of LVAD using direct modeling of the hazard of the subdistribution on the competing events (Figure 2.7) is almost similar to the effect obtained by translating the effect on the cause-specific hazard modeling (Figure 2.6) into the cumulative incidence function.

![Figure 2.7](image-url)

*Figure 2.7: Regression modeling on the hazard of the subdistribution: Effect of device type LVAD on the cumulative incidence functions of the competing events.*
Chapter 3

Multivariate Longitudinal Data

This chapter focuses on modeling of longitudinal data. We first briefly provide an overview of the current literature on longitudinal modeling using linear and non-linear modeling. We then propose a multiphase non-linear mixed effects model for univariate longitudinal outcome and extend this approach to model a multivariate longitudinal data.

3.1 Introduction

In medical science, we often encounter data collected repeatedly over time. For example, after aortic valve repair, investigators are often interested in determining the long-term durability of the repaired valve. Durability of the aortic valve, for example, can be determined by investigating hemodynamics of the valve, such as aortic peak gradient, mean gradient, and aortic valve regurgitation (Banbury et al [47]). These investigations often require collection of such data repeatedly over time for each subject and analyzing them to determine if there is a temporal change in the response of interest. Data collected repeatedly over time for each subject in a study are called longitudinal data. In contrast, in cross-sectional data, the response of interest is measured only at a single time point for each subject. Since longitudinal data are collected repeatedly over time, they can
be thought of as *time series* data. However, in principle, in a typical time series dataset we have large number of observations collected in a small number of subjects (or units), often only on one subject. An example would be monthly temperature for the last 20 years in a single city. On the other hand, in a longitudinal data set up, we collect a relatively small number of observations on a large number subjects. Sociologists and economists often refer to longitudinal data as *panel data*. Replacing the time dimension in longitudinal data with *spatial* dimension leads to *spatial data*.

The important feature of longitudinal data is that while, as with cross-sectional data, the observations between subjects are assumed to be uncorrelated, the observations from a single subject tend to be correlated. Hence any statistical analyses or model that involves longitudinal data should take such correlation into consideration in order to draw valid inferences.

In general, the objectives of a longitudinal data analysis can be either population based inferences (marginal), or subject-specific inferences or transitional. Diggle et al. [48] suggested there may be three sources of variability in longitudinal data: i. inter-individual variability, which can be accounted for by using *random effects*; ii. serial correlation, which presents when residuals closer to each other in time are more similar than residuals far apart in time, and can be accommodated, for example, using time-series type correlation structures; and iii. measurement error, this occurs when one uses some kind of sampling to obtain measurements within a subject.

Longitudinal observations can be continuous, ordinal, binary or count data. For example, serial measurements of aortic valve gradient [47] after aortic valve repair is a continuous longitudinal data. Serial observation of aortic valve regurgitation after aortic valve replacement [49] where the regurgitation, which is recorded on an ordinal scale as none (0), mild (1+), moderate (2+), moderately severe (3+), severe (4+) and this scale reflects the severity of leakage in the valve [49], can be regarded as ordinal longitudinal data. Serial observations of presence (1) or absence (0) of atrial fibrillation based of
12-lead EKG after surgical ablation for atrial fibrillation [50] are longitudinal binary data. Serial observation of the number of units of red blood cell transfusion after lung transplant can be considered as longitudinal count data.

In many medical studies, we observe more than one longitudinal outcome. That is, we encounter multivariate longitudinal data. In a multivariate longitudinal data set up, we observe more than one outcome for the same subject repeatedly over time. For example, we observe serial measurements of postoperative normalized forced expiratory volume in 1 second (FEV1) and forced vital capacity (FVC) after lung transplant [51]. However, we often tend to analyze these outcome separately using readily available univariate longitudinal methods. When analyzing each outcome separately, one ignores the association structure between different outcomes, and this may induce bias in the inferences, if we are interested in the effect of some factor on all the longitudinal responses. Further, we may also be interested in associations among the longitudinal responses, particularly, evolution of the association among the responses [53]. In these situations, one has to analyze these multivariate longitudinal outcomes jointly. By using a multivariate model, one could increase the efficiency and power [52] by borrowing strength across the multiple longitudinal response variables.

Note that these different longitudinal outcomes may be of different types. For example, after mitral valve repair, we serially observe left ventricular ejection fraction which is continuous and graded severity of mitral regurgitation which is ordinal. Hence, in this example, we have mixed (continuous and ordinal) multivariate longitudinal data. In our motivating problem, bridge-to-transplant data, we have serially observed glomerular filtration rate (GFR) a marker for renal function and bilirubin a marker for liver function. Both are continuous type longitudinal data and thus, we limit our discussion in the following sections to multivariate continuous longitudinal data.
3.2 Univariate: Notation and Concepts

Let $Y_{ij}$ represent a response variable and $X_{ij}$ a vector of length $p$ ($p$-vector) of explanatory variables observed at time $t_{ij}$, for observation $j = 1, \ldots, n_i$ on subject $i = 1, \ldots, m$. So that, $Y_i = (Y_{i1}, \ldots, Y_{in_i})$ is $n_i$-vector of longitudinal (repeated) outcomes and $X_i = (X_{i1}, \ldots, X_{in_i})^\top$ is a $n_i \times p$ matrix of explanatory variables collected for subject $i$. Therefore, the responses of all subjects are denoted as $Y = (Y_1, \ldots, Y_m)$ which is a $N$-vector, where $N = \sum_{i=1}^{m} n_i$. Baseline explanatory variables will not change over time for a subject. Note that the time to the measurement is external to the process and can be treated as one of the explanatory variables. We now describe the two important modeling approaches that are frequently used to analyze longitudinal data.

3.2.1 Marginal Models

In marginal regression models, the regression of the response on explanatory variables is modeled separately from within-subject correlation. In a marginal model, the marginal expectation of the response is $E(Y) = \mu$ and it depends on the explanatory variables through a function $g$ as $g(\mu) = f(\beta; X)$, where $\beta$ is a $p$-vector of regression coefficients. The function $g$ is called the link function. When $Y$ is a normally distributed response, $g$ can be an identity function; for a binary response, $g$ can be a logistic function. For a linear model $f(\beta; X) = X\beta$. In general, $f(\beta; X)$ can be nonlinear. In marginal models, it is further assumed that the marginal variance is a function of the marginal mean, $\mu$ and the correlation between responses for a subject is also a function of the marginal mean. When the response is normally distributed and the model is linear, the model is called a general linear model (GLM). If the response is non-normal and the model is linear, it is called a generalized linear model (GLIM).

The regression coefficients, $\beta$ in a GLM can be estimated using weighted least squares estimates [48, p.59]. Since the usual maximum likelihood estimation produces biased
estimates of the variance, Patterson and Thompson [54] introduced restricted maximum likelihood estimation (REML). See, for example, Verbeke and Molenberghs [55, p.44] for further details on REML. Since the maximum likelihood approach to estimate regression coefficients, \( \beta \) in a GLIM setup (non-normal response) requires assumptions about higher order moments, Liang and Zeger [56] proposed generalized estimation equations (GEE) where they have extended the quasi-score function [57] of a univariate exponential family to a multivariate setup. GEE requires only the first two moments without explicitly defining the joint distribution of the longitudinal responses. GEE is, hence, a semi-parametric approach.

Therefore, in summary, in marginal models, marginal distributions are used for the longitudinal response and the correlation among the responses within a subject can be accounted for by a fully parametric approach or by a semi-parametric approach such as, for example, GEE.

### 3.2.2 Random Effect Models

When the subjects are samples from a large population, one can model the natural heterogeneity due to unmeasured factors by introducing random effects in the model. In other words, the natural heterogeneity across subjects is represented by a probability distribution. In modeling of longitudinal response data, the correlation among the observations within a subject is accounted for by using the unobservable random effects. Hence, this type of model is also called a latent variable model. Laird and Ware [58] introduced the random effect models, more appropriately called mixed effect models, in a longitudinal setup.

A simple linear mixed effect model is given as follows:

\[
E(Y_i|b_i) = X_i\beta + Z_i b_i, 
\]
where $\mathbf{\beta}$ is a vector of regression coefficients of the fixed effects associated with fixed effect design matrix $\mathbf{X}_i$ and $\mathbf{b}_i$ is a vector of random effects associated with the random effect design matrix $\mathbf{Z}_i$ and assumed to follow a probability distribution. In general, the random effects models can be defined as follows:

\begin{equation}
\begin{aligned}
g(\mu_i(b_i)) &= f(\mathbf{\beta}, \mathbf{b}_i; \mathbf{X}_i, \mathbf{Z}_i),
\end{aligned}
\end{equation}

where $\mu_i(b_i) = E(Y_i|b_i)$ and $g$ is a link function. The random effects $b_i$ ($i = 1, \ldots, m$) are mutually independent with a common multivariate distribution $F_b(G)$, often a multivariate normal. Conditional on the random effects $b_i$, the longitudinal responses are assumed to be independently, identically distributed, often as a distribution from the exponential family. Hence, these models often called conditional independence models.

In general, the interpretation of the fixed effects $\mathbf{\beta}$ is conditional on $b_i$. Hence, in a random effects model, estimates of $\mathbf{\beta}$ are usually called subject-specific estimates. However, when $g$ is the identity function $f(\mathbf{\beta}, \mathbf{b}_i; \mathbf{X}_i, \mathbf{Z}_i) = \mathbf{X}_i \mathbf{\beta} + \mathbf{Z}_i \mathbf{b}_i$ and $\mathbf{b}_i$ is a multivariate normal with zero mean vector, it can be shown that $E(Y_i|b_i) = \mathbf{X}_i \mathbf{\beta} + \mathbf{Z}_i \mathbf{b}_i$ and $E(Y_i) = \mathbf{X}_i \mathbf{\beta}$. Hence, under this scenario the fixed effect regression coefficients $\mathbf{\beta}$ have the same interpretation for both conditional models and marginal models. Note that, since $E_b\{E(Y_i|b_i)\} = E(Y_i)$, one can obtain estimates of the marginal expectation of $\mathbf{Y}$ from conditional models [59]. In fact, Lee and Nelder [60] argue that the conditional model is the basic model and any conditional model leads to a specific marginal model.

Further, in practice, longitudinal data are often highly unbalanced in the sense that, each subject has different number of longitudinal responses observed at non-fixed time points. In this scenario, a mixed effect modeling approach is preferable to a marginal modeling approach.

In principle, estimation of fixed effects parameters, $\mathbf{\beta}$ and parameters of the distribution of the random effects, $\mathbf{G}$ can be obtained by the method of maximum likelihood.
estimation. Precisely, using marginal maximum likelihood estimation involves finding estimates of $\beta$ and $G$ that maximize:

$$L(\beta, \psi, G; y) = \prod_{i=1}^{m} \int \prod_{j=1}^{n_i} f_{y|b}(y_{ij}|b_i; \beta, \psi) f_b(b_i; G) db_i,$$

where $f_{y|b}$ is the conditional density of the longitudinal response, $\psi = var(y_i|b_i)$, and $f_b$ is the density of the random effects. Equation (3.2) is simply the marginal distribution of $Y$ obtained by integrating the joint distribution of $Y$ and $b$ with respect to $b$. Except for some special cases, the integral in (3.2) does not have a closed form. Hence, first some numerical methods, such as, for example, numerical integration, may have to be implemented to evaluate the integral, before maximizing the marginal likelihood, again using some numerical methods such as, for example, Newton-Raphson method. For generalized linear mixed models, McCulloch [61] provides some details on the usage of an extension of the EM algorithm for parameter estimation. Estimation of random effects can be done by using empirical Bayes inference where we use the posterior density function of $b_i$ given $y_i$, $f_{b|y}$ which is obtained using Bayes theorem. See, for example, Verbeke and Molenberghs [55, p.78] for further details.

Laird and Ware [58] used the EM algorithm for estimation in linear mixed effect models. For linear mixed effect models for unbalanced data, Jennrich and Schluchter [62] discussed the implementation of Newton-Raphson and Fisher’s scoring algorithms for computing maximum likelihood estimates and a generalized EM algorithm to estimate restricted and unrestricted maximum likelihood estimates. In general, analytical methods for analyzing longitudinal data using linear mixed effect models and general linear models are well developed in current literature. See for example, Diggle et al. [48], Crowder and Hand [63].

Because of the need to have a flexible modeling approach to model unbalanced (or balanced) longitudinal data, particularly, in temporal relationships in biological, and
pharmacokinetics data, the usage of non-linear mixed effect models have been increasing recently. For example, Wu and Ding [64] proposed a non-linear temporal trend of viral load after antiretroviral therapy. Because of computational difficulties in modeling categorical data, most of the literature focuses on the case of continuous longitudinal data. Sheiner and Beal [65] proposed a non-linear one compartment model to describe the plasma concentration over time. A thorough review of non-linear models and estimation methods can be found, for example, in Davidian and Giltinan [66].

Estimation in nonlinear random effects models can be carried out by maximum likelihood estimation which can be implemented either using exact methods or by using approximation methods. In the exact approach, one can use numerical integration as in equation (3.2) or use Markov Chain Monte Carlo (MCMC) methods such as the Gibbs sampler or Metropolis-Hastings algorithm. Thus, in general, “exact” methods involve approximating the integral in the marginal likelihood in equation (3.2) using methods such as Gaussian quadrature numerical integration method, or Laplacian approximation, or using some MCMC method and then optimizing the approximation. See, for example, Pinheiro and Bates [67] and Vonesh [68] for further details.

In the approximation methods, one can use first order or second order Taylor series expansion of the nonlinear function \( f(\beta, b_i; X_i, Z_i) \) [65]. The Taylor series expansion of \( f(\beta, b_i; X_i, Z_i) \) can be taken around the random effect \( b_i = 0 \) or around \( b_i = \hat{b}_i \). Vonesh and Carter [69] implemented a four-step approach that is based on first-order Taylor series expansion to estimate parameters of a nonlinear mixed effects model. See, for example Davidian and Giltinan [16, ch 6.] for the description of the approximations.

Another widely used estimation approach is the so called “two-stage” estimation procedure. In general, the approach is as follows. In the first stage \( y_i = f(x_i, B_i) + \epsilon_i \), using \( i^{th} \) subject’s data and for example, nonlinear ordinary least squares estimator, subject-specific regression estimates \( \hat{B}_i \) are obtained. Hence, in the first stage, we model the \textit{intra-subject variation}. In the second stage \( B_i = \beta + b_i \), using \( \hat{B}_i \) as “observed” data,
\(\beta\) and variance parameters are estimated. Hence, in this stage, we account for *inter-subject variation*. It is noted, however, that we need sufficient number of data points for each subject to estimate the subject-specific estimates. See, for example Davidian and Giltinan \[16, \text{ch } 5.\] for a detailed description of the approach.

### 3.3 Multivariate: Notation and Concepts

In our motivating example, one of the main points of interest is how the evolution of renal function is related to the evolution of liver function (“association of evolutions”). We further want to investigate how the association between the renal function and liver function evolves over time (“evolution of the association”). To answer these questions, one needs to model these correlated longitudinal responses jointly. An important application of mixed effect modeling is in the area of joint modeling. When we have more than one response of interest, for example, more than one longitudinal responses, or one or more longitudinal responses and time related event(s), by introducing a common distribution for the random effects these longitudinal responses can be modeled jointly. One of the advantages of this approach is that not only the same type of responses, as in our motivating example where we have two continuous longitudinal responses, but also different types of longitudinal responses, for example, a continuous and an ordinal, can be modeled jointly. See, for example, Fieuws and Verbeke \[53\] for further details. Shah et al. \[70\] extended linear mixed models to allow for multiple outcomes in longitudinal responses. Note that by modeling the multivariate longitudinal responses jointly, and hence, utilizing the underlying covariance structure of the multivariate responses, we may obtain efficient estimation of the parameters.

We now extend the notation from the previous section to have multivariate longitudinal responses. We suppose, for subject \(i\), that there are \(K\) correlated longitudinal continuous responses \(Y_{1i}, Y_{2i}, \ldots, Y_{Ki}\) with design matrices for fixed and ran-
dom effects denoted by $X_1^i, X_2^i, \ldots, X^K_i$ and $Z_1^i, Z_2^i, \ldots, Z^K_i$ respectively. Note these covariate matrices need not necessarily be mutually exclusive. Note that, for subject $i$, the longitudinal response $k$ is observed at $n^k_i$ distinct time points. That is, $Y^k_i = (Y^k_{i1}, Y^k_{i2}, \ldots, Y^k_{in^k_i})$, $(k = 1, \ldots, K)$. It should be noted that in observational studies, in general, not all $n^k_i$ ($k = 1, \ldots, K$), are necessarily equal. That is, for a given subject, only for a subset of time points all $K$ responses are observed. Therefore, in multivariate longitudinal data, in terms of number of data points, we may have the following two features:

1. As in the univariate case, the number of time points and time intervals at which the observations are collected may differ from subject to subject;

2. For a given subject, the number of time points and time intervals at which observations are collected may differ from response to response.

An extension of the univariate mixed model approach to multivariate longitudinal modeling is to define different mixed effect models for each longitudinal response and then model their association by introducing a common distribution for the random effects.

In general, we can write the joint model as follows:

\[
\begin{align*}
    Y^1_i &= f^1(\beta^1, b^1_i; X^1_i, Z^1_i) + \epsilon^1_i \\
    Y^2_i &= f^2(\beta^2, b^2_i; X^2_i, Z^2_i) + \epsilon^2_i \\
    \vdots &= \vdots \\
    Y^K_i &= f^K(\beta^K, b^K_i; X^K_i, Z^K_i) + \epsilon^K_i,
\end{align*}
\]
where \( f^k \) is a linear or non-linear function of covariates and regression parameters related to response \( Y^k \) \((k = 1, \ldots, K)\). Note that, \( f^k \) can be of different functions of covariates and regression parameters for different responses. \( X^k_i \) is an \( n^k_i \times p^k \) covariate matrix for the response \( k \), and \( \beta^k \) is a \( p^k \)-vector of fixed effects, \( Z^k_i \) is an \( n^k_i \times q^k \) matrix, usually a subset of \( X^k_i \), of covariates related to \( q^k \)-vector of subject-specific random effects \( b^k_i \) with the random effects with \( q^k \leq p^k \). Now let,

\[
(3.4) \quad \mathbf{b}_i = \begin{bmatrix} b^1_i \\ b^2_i \\ \vdots \\ b^K_i \end{bmatrix}, \quad \mathbf{e}_i = \begin{bmatrix} e^1_i \\ e^2_i \\ \vdots \\ e^K_i \end{bmatrix},
\]

then, in general, one can specify some multivariate distributional assumptions for random effects \( \mathbf{b}_i \) and measurement errors \( \mathbf{e}_i \) as follows: \( \mathbf{b}_i \sim F_{\mathbf{b}}(\mathbf{G}) \) and \( \mathbf{e}_i | \mathbf{b}_i \sim F_{\mathbf{e}}(\Sigma_i) \), similar to univariate non-linear mixed effect models and distributions of \( \mathbf{b}_i \) and \( \mathbf{e}_i \) are independent. In practice, most assume, \( \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}) \), a multivariate normal. For continuous responses, as in our motivating problem, in practice, most assume \( \mathbf{e}_i | \mathbf{b}_i \sim N(\mathbf{0}, \Sigma_i) \).

In practice, most ([71], [72], [73], [74]) of the multivariate longitudinal modeling approaches assume a linear model as \( \mathbf{Y}_i = \mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i \), where \( \mathbf{b}_i \) and \( \mathbf{e}_i \) are as in (3.4) and

\[
(3.5) \quad \beta = \begin{bmatrix} \beta^1 \\ \beta^2 \\ \vdots \\ \beta^K \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{X}^1_i \\ 0 & \mathbf{X}^2_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{X}^K_i \end{bmatrix}, \quad \mathbf{Z}_i = \begin{bmatrix} \mathbf{Z}^1_i \\ 0 & \mathbf{Z}^2_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}^K_i \end{bmatrix}.
\]
One important modeling assumption that most adopt concerns the variance covariance matrix, $\Sigma_i$, of the measurement errors, where it is assumed that $\Sigma_i$ is a diagonal matrix. That is, it is assumed that conditional on the random effects, the multivariate longitudinal responses are independent. For example, Chakraborty et al. [75] estimated correlation between two longitudinal responses under the assumption of independent measurement error. However, care should be taken when one makes assumption about the covariance structure of the residuals. Fieuws and Verbeke [53] relaxed this conditional independence assumption on modeling of a bivariate longitudinal hearing data and showed a better fit for the evolution of marginal correlation. Marshall et al. [76] proposed a non-linear random effect model under the assumption of correlated measurement errors.

Parameters of the model can be estimated using likelihood based inference using readily available statistical software, for example, PROC MIXED or PROC NLMIXED in SAS. See for example, Thiebaut, et al. [77] for an application using PROC MIXED.

As the number of longitudinal responses increases, the dimension of the covariance matrices $G$ and $\Sigma$ increases and this in turn may introduce computational problems. To reduce this computational problem, Fieuws and Verbeke [74] proposed a pairwise fitting approach where they fitted $K(K - 1)/2$ number of possible pairs of longitudinal responses and then combined the estimates to obtain overall estimates for the $K$ longitudinal responses. Thus instead of maximizing the log likelihood of the joint mixed model of dimension $K$, they maximized the likelihood of bivariate mixed models and then averaging the all pair-specific maximum likelihood estimates. The variances and covariances of these estimates are obtained by maximizing a pseudo-likelihood.
3.4 A Non-Linear Mixed Effects Model

We first define a flexible non-linear mixed effect model for a single longitudinal response and apply to our motivating example. We then extend the model to multivariate longitudinal responses.

Let \( y_{ij} \) be the \( j \)th (\( j = 1, \ldots, n_i \)) continuous response for subject \( i \) (\( i = 1, \ldots, m \)).

We now consider the following non-linear mixed effect model,

\[
y_i = X_{io} \beta_o + \log \left( \sum_{l=1}^{L} \mu_l(X_{il}; \beta_l)T_l(t_i, \Theta_l) \right) + b_i + \epsilon_i,
\]

where, \( \log(\cdot) \) is a \( n_i \times 1 \) vector with \( \log \left( \sum_{l=1}^{L} \mu_l(X_{il}; \beta_l)T_l(t_{ij}, \Theta_l) \right) \) as its \( j \)th element; \( t_i \) is a \( n_i \times 1 \) vector of time points at which longitudinal response \( y_i \) are observed or measured for subject \( i \); \( X_{il} \)s are matrices of design matrix of fixed effects that are not necessarily equivalent, and \( \beta_l \)s are the corresponding fixed-effect regression parameters.

The subject-specific random effect \( b_i \) is assumed to follow a normal distribution, \( N(0, \sigma_b^2) \) and \( \epsilon_i \) is a \( n_i \times 1 \) vector of measurement errors. It is further assumed that \( \epsilon_i | b_i \sim N(0, \sigma^2 \epsilon I) \). That is, we assume conditional independence.

\( T_l(t, \Theta_l) \) is a function of time that depends only on time \( t \) and a parameter vector \( \Theta_l \), and this function can be any of the forms or transformations of (3.8) given below.

The model (3.6) has two main fixed components:

1. An overall model (namely, \( X_{io} \beta_o \)) that does not depend on time;

2. A series of log-linear models \( \mu_l(X_{il}; \beta_l) = \exp \{ X_{il} \beta_l \} \) (\( l = 1, \ldots, L \)) that are modulated by the time functions \( T_l(t, \Theta_l) \). Hence, the effects of covariates \( X_{il} \)s are time varying. Note that, as we explain later in the discussion of time phases, the scalers \( \mu_l(X_{il}; \beta_l) \) are related to the corresponding time function \( T_l(t, \Theta_l) \) through some property of \( T_l(t, \Theta_l) \).
Note that, by taking exponentiation of the conditional expectation of the model (3.6), it can be written as

\[
(3.7) \quad \exp\{E(y_i|b_i)\} = \exp\{X_{io}\beta_o + b_i\} \times \sum_{l=1}^{L} \mu_l(X_{il}; \beta_l)T_l(t_i; \Theta_l),
\]

where \(\exp\{\cdot\}\) is a vector with \(n_i\) elements. Hence, multiple overlapping time phases of outcome are additive in the conditional expectation domain, with each phase individually shaped by a function of time \(T_l(t, \Theta_l)\) and scaled by a function of concomitant information \(\mu_l(X_{il}; \beta_l)\). Note that we can use as many phases as the data warrants, but in our data analysis experience, at most 2 phases – early, and late – are usually adequate and very rarely at most 3 phases – early, constant and late.

Following the spirit of Blackstone and colleagues’ [32] multiphase parametric hazard modeling strategy to model time-to-event data, the motivation for the nonlinear mixed effect model (3.6) is two fold:

1. In medical sciences, the temporal trend of a longitudinal biological response of interest plays an important role. It essentially describes how a patient or a group of patients behave after an intervention and most of the post-intervention patient management protocols are based on these temporal trends. Hence, it is important to model the temporal trend of a longitudinal response accurately. It is also known that most of the biological responses (such as, pharmacokinetics, lab results from blood samples, biomarkers [78], [79], [80]) are nonlinear in trend. The additive flexible linear and non-linear components in (3.6) can handle any non-linear and linear trends.

2. It is known that the magnitude of the effect of a risk factor on a longitudinal response can change with time. Especially, after a medical intervention (for e.g., surgery), different risk factors can influence the longitudinal response at differ-
ent time segments. For example, after a mitral valve repair, the type of mitral valve ring and experience of the surgeon may influence the early measurements of ejection fraction, say, measured within one month after surgery. However, patient comorbidities, such as diabetes and history of smoking may be two of the risk factors that influence late ejection fraction, say, one year after the repair. Older age, on the other hand, may be a predictor for ejection fraction regardless of the time. Hence, while the effect of some factors on a longitudinal response may diminish and become negligible or started increasing and become noticeable after some time, some other factors’ effect may stay the same regardless of the time. In other words, there is an interaction effect between some factors and time on the longitudinal response. The model (3.6) can accommodate this scenario.

It is noted that our model is similar to a model proposed by Wu [81] where Wu [81] proposed a similar but a focused biphasic model using exponential decay equations to fit a specific shape of temporal trend of virus load in an AIDS study, our model is more flexible, in the sense that they can handle more than 2 time phases.

We now briefly describe the time function and the 3 commonly occurring phases in the following sub sections.

3.4.1 Multi-Phase time function

We now describe the time function $T(t, \Theta)$ and some commonly used special cases of the function.

Mathematical Formulation

The generic equation of the time for the multiphase model was originally described as a model of cumulative mortality by Hazelrig and colleagues [82] and then used in a multiphase hazard model to fit time-to event data by Blackstone and colleagues [32].
We transform these equations into the conditional expectation domain. The generic function \( G(t, \Theta) \) or any of its transformation can be used as \( T(t, \Theta) \). The family of equations is given as:

\[
G(t, \Theta) = \frac{|\nu| - \nu}{2|\nu|} + \frac{\nu}{|\nu|} \left[ 1 + \phi(m) \left( \frac{|m| - m}{2|m|} + \frac{|\nu|t}{\rho(t_{1/2})} \right)^{-1/\nu} \right]^{-1/m} - 1/\nu, \]

where \( m > 0 \) and/or \( \nu > 0 \), \( \phi(m) = m \) if \( m > 0 \), and \( \phi(m) = -1 \) if \( m \leq 0 \). \( \Theta \equiv (m, \nu, t_{1/2}) \) is the shaping parameter vector, \( \rho(t_{1/2}) \) is a function of \( t_{1/2} \), \( m \), and \( \nu \). We define the parameter \( t_{1/2} \) as the time point \( t \) such that \( G(t_{1/2}) = 1/2 \). Based on this definition one can deduce \( \rho(t_{1/2}) \) in terms of \( m \), \( \nu \), \( t_{1/2} \). In survival terminology, Blackstone and colleagues [32] called \( t_{1/2} \) the half life time. That is, \( t_{1/2} \) is the time point at which cumulative mortality is 1/2. Natural constraints on \( G \) are that \( G(0, \Theta) = 0 \) and \( G(t, \Theta) \rightarrow 1 \) as \( t \rightarrow \infty \).

The first derivative of \( G \) with respect to \( t \) is

\[
g(t, \Theta) = \left[ 1 + \phi(m) \left( \frac{|m| - m}{2|m|} + \frac{|\nu|t}{\rho(t_{1/2})} \right)^{-1/\nu} \right]^{-1/m - 1/\nu} \left( \frac{|m| - m}{2|m|} + \frac{|\nu|t}{\rho(t_{1/2})} \right)^{-1/\nu} \frac{\phi(m)}{\rho(t_{1/2})m},
\]

which is non-negative. Thus, \( G \) is non-decreasing.

Note that, when \( m < 0 \) and \( \nu < 0 \), \( G(0, \Theta) \neq 0 \), and this violates the constraints. Thus, this generic formulation (3.8) does not exist for \( m < 0 \) and \( \nu < 0 \). Hence, the generic formulation (3.8) simplifies into three cases, depending on the signs of \( m \) and \( \nu \):

**Case 1:** \( m > 0 \) and \( \nu > 0 \):

\[
G(t, \Theta) = \left[ 1 + m \left( \frac{\nu t}{\rho(t_{1/2})} \right)^{-1/\nu} \right]^{-1/m} - 1/\nu,
\]

\[
\rho(t_{1/2}) = \nu t_{1/2} \left( \frac{2m - 1}{m} \right)^{\nu}.
\]
and the limiting case as $m \to 0^+$ is

$$G(t, \Theta) = \exp \left[ - \left( \frac{v t}{\rho(t_{1/2})} \right)^{-1/v} \right],$$

$$\rho(t_{1/2}) = vt_{1/2} \log^\nu(2).$$

**Case 2:** $m < 0$ and $\nu > 0$:

$$G(t, \Theta) = \left[ 1 - \left( 1 + \frac{vt}{\rho(t_{1/2})} \right)^{-1/m} \right]^{-1/m},$$

$$\rho(t_{1/2}) = vt_{1/2} \left( (1 - 2^m)^{-\nu} - 1 \right)^{-1}.$$

and the limiting case as $\nu \to 0^+$ is

$$G(t, \Theta) = 1 - \exp \left( - \frac{t}{\rho(t_{1/2})} \right) \right]^{-1/m},$$

$$\rho(t_{1/2}) = -t_{1/2} \log^{-1}(1 - 2^m).$$

**Case 3:** $m > 0$ and $\nu < 0$:

$$G(t, \Theta) = 1 - \left[ 1 + m \left( \frac{-vt}{\rho(t_{1/2})} \right)^{-1/m} \right]^{-1/m},$$

$$\rho(t_{1/2}) = -v t_{1/2} \left( \frac{2^m - 1}{m} \right)^\nu.$$

and the limiting case as $m \to 0^+$ is

$$G(t, \Theta) = 1 - \exp \left[ - \left( - \frac{vt}{\rho(t_{1/2})} \right)^{-1/v} \right],$$

$$\rho(t_{1/2}) = -vt_{1/2} \log^\nu(2).$$
3.4.2 Model components

We now briefly describe the 3 commonly occurring time phases in the following subsections.

Overall model

\[ X_o\beta_o \] is a model that identifies the risk factors that are related to the subject-specific mean response in an overall fashion and do not involve time \( t \).

Early phase

In the early time phase, \( g(t, \Theta) \), the first derivative of \( G(t, \Theta) \) is the most commonly used function. That is, we can use \( g(t, \Theta) \) as \( T(t, \Theta) \). Under this scenario, the scaling parametric function \( \mu_l(X_l, \beta_l) \) is related to the area beneath function \( g(t, \Theta)(= T(t, \Theta)) \).

Various shapes of \( g \) for different values of \( m \) and \( \nu \) are given in Figure 3.1. In Figure 3.1, while we vary \( m \) and \( \nu \), we keep the \( t_{1/2} \), the time at which the area under curve is 1/2, fixed at 3 months.

![Figure 3.1: Three cases of shapes of \( g(t, \Theta) \): in all cases \( t_{1/2} = 3 \) months. Case I: \( \nu = 0.5 \) and \( m = 0.5 \), we have an early peaking function; Case II: \( \nu = 0.5 \) and \( m = -1 \), the function starts at a finite point and decreases; Case III: \( \nu = -1.5 \) and \( m = 0.5 \), we have a decreasing function starting at infinite.](image-url)
Note that, for the early peaking function, by changing \( t_{1/2} \), one can change the peaking point.

**Constant phase**

This phase is time independent. Hence, it will only have concomitant information, \( \mu_l(X_t, \beta_t, b_u) \). Note that, when \( \nu = -1 \) and \( m = 0 \), \( h(t, \Theta) \) described below changes to a constant phase (Figure 3.2).

**Late phase**

The most commonly used function for the late phase is

\[
T(t, \Theta) = h(t, \Theta) = \frac{g(t, \Theta)}{1 - G(t, \Theta)}
\]

Hazelrig and colleagues [82] used this transformation as a hazard function and their motivation for this transformation is that, suppose \( G(t, \Theta) \) is a CDF and \( g(t, \Theta) \) the corresponding pdf then \( h(t, \Theta) \) is a hazard function. It is noted that the function \( h(t, \Theta) \) is of the form \( \frac{1}{t_{1/2}} f\left(\frac{t}{t_{1/2}}\right) \). That is, \( \frac{1}{t_{1/2}} \) is a scaling parameter. Hence, wherever one uses \( h(t, \Theta) \) or any of its transformation as \( T(t, \Theta) \), for identifiability, we take the intercept in \( \mu_l(X_t, \beta_t) \) as 0. That is, \( \frac{1}{t_{1/2}} \) acts as the intercept. Therefore, under this scenario, only \( m \) and \( \nu \) are the shaping parameters. Further, the covariate information connected to the time function \( T(t, \Theta) \) through the scaling parameter \( \frac{1}{t_{1/2}} \). Note that, this is a preferred approach in survival modeling (for example, Odell et al. [83]). Note that, Wu [81] also linked the covariate information to exponential decay through the parameters of the decay equation.

Four different shapes of \( h(t, \Theta) \) for different values of \( m \) and \( \nu \) are given in Figure 3.2. Note that, \( h(t, \Theta) \) becomes a constant phase when \( m = 0 \) and \( \nu = -1 \).
Figure 3.2: *Four cases of shapes of $h(t, \Theta)$: in all cases $t_{1/2} = 6$ years.* Case I: $\nu = 1.5$ and $m = 1.5$, we have an early decreasing function starting at infinite; Case II: $\nu = 0.5$ and $m = 0.5$, we have a late peaking function; Case III: $\nu = -0.75$ and $m = 0$, we have a late increasing function; and Case IV: $\nu = -1$ and $m = 0$, the function is a constant phase.

3.4.3 Data Analysis

In this section, we analyze the motivating data using the multiphase nonlinear mixed effect model.

**Temporal trend**

Following the motivation of this type of modeling strategy, we first focus on the temporal trend of the longitudinal outcome and then focus on the baseline covariates that influence the trend. Thus, we first consider the model (3.6) without covariates and determine the number of phases $L$ and estimate the shaping parameter vector $\Theta$ for each phase. With only time, the model (3.6) can be written as,

$$
\begin{align*}
\mathbf{y}_i &= \log \left( \sum_{l=1}^{L} \mu_l(\beta_0|T_i(t_i, \Theta_l) + b_i + \epsilon_i \\
(3.10) &= \log \left( \sum_{l=1}^{L} \exp\{\beta_0|T_i(t_i, \Theta_l) \right) + b_i + \epsilon_i,
\end{align*}
$$
where $\beta_{0l}$ is a phase-specific intercept (fixed effect) and $b_i$ is a patient-specific random intercept. We use Laplacian approximation to approximate the integral in (3.2) and then use quasi-Newton optimization method to estimate the maximum likelihood estimate of the shaping parameter vectors $\Theta_l$. See, for example, Pinheiro and Bates [67] and Vonesh [68] for detail on the Laplacian approximation. This process can be carried out by using SAS procedure, PROC NLMIXED. The implementation of this SAS procedure is based on Pinheiro and Bates [67].

**Temporal trend $\text{Log}(\text{Bilirubin})$:** The data suggested a bi-phase model for the temporal trend of log(Bilirubin). We describe our approaches to identifying the number of phases and estimation of the parameters in more detail at the end of this section. Estimates of the shaping parameters are given in the Table 3.1. Note that, since we have used $T(t, \Theta) = h(t, \Theta)$ as the late phase time function, the $t_{1/2}$ acts as an intercept (scaling, not shaping) parameter in the late phase.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ± SE</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early decreasing phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}_{01}$</td>
<td>-1.4±0.063</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\hat{\nu}$</td>
<td>-0.99±0.017</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{t}_{1/2}$</td>
<td>0.069</td>
<td>-</td>
</tr>
<tr>
<td><strong>Late increasing phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\nu}$</td>
<td>1.11±0.0365</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{t}_{1/2}$</td>
<td>1.41</td>
<td>-</td>
</tr>
<tr>
<td>$\text{Var}(\epsilon) = \hat{\sigma}^2_\epsilon$</td>
<td>0.31±0.0046</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\text{Var}(b) = \hat{\sigma}^2_b$</td>
<td>0.75±0.056</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table 3.1: Estimates of shaping parameters of the temporal trend of Logarithm of bilirubin. Note that, in the early phase, $m = 0$ is the limiting case of $g(t, \Theta)$ when $\nu < 0$ and $m \to 0^+$ as described in Case 3 in the mathematical formulation subsection 3.4.1, and in the late phase, $m = 0$ means the limiting case of $h(t, \Theta)$ when $\nu > 0$ and $m \to 0^+$ as described in Case 1 in the mathematical formulation subsection 3.4.1.

Note that, we have used $h(t, \Theta)$ as time function $T(t, \Theta)$ for the late phase of the temporal trend of bilirubin (Table 3.1). Hence, the parameter $\hat{t}_{1/2}$ acts as the intercept for this phase. Based on the estimates in Table 3.1, the estimated multiphase temporal
trend equation can simplified as follows:

$$E(\log(bilirubin)|b_i) = \log\left[\frac{2.4}{\exp\{10t\}} + \frac{(1.06t)^{-1.9}}{1.05[\exp\{-(1.06t)^{-0.9}\} - 1]}\right] + b_i,$$

where $b_i$, the patient-specific random effects are assumed to be a normal variate.

The bi-phase model yielded an early decreasing phase and a late increasing phase which plateaus after 6 months (Figure 3.3). For these data, the early phase is based on equation $g(t, \Theta)$ and the late phase is based on equation $h(t, \Theta)$. That is we used $T_1(t, \Theta) =$
The equations are based on limiting of \( m \to 0^+ \) as in the Case 1 and Case 3 in the mathematical formulation subsection 3.4.1, respectively.

The temporal trend of the logarithm of bilirubin is given in Figures 3.4 and 3.5. The mean response (the average profile) is obtained by setting \( b_i = 0 \). Note that since the random effect enters the model linearly, the mean response \( E(Y) = E(E(Y|b)) \) estimated in closed form. The average data are obtained by using a bin smoother as follows: Based on the time of the measurement, we first partitioned the response values into a number of disjoint groups and taking the mean of the response of interest, in this case, mean of the logarithm of bilirubin. Note that, when binning, the possibility of a given bin having multiple observations from the same patient was not taken into the consideration. Hence, average actual data are used here as a crude verification of the model fit. The average patient profile appears to more closely follow the average actual data.

**Remark:** Estimation of the temporal trend is a data driven procedure. That is, based on the binned smoothers, we first get an overview of the shape of the non-linear (if any) temporal trend of the logarithmic transformation of bilirubin. We then, for these data, started with 3 phases an early \( (g(t, \Theta)) \), a constant and a late \( (h(t, \Theta)) \).

1. We started with \( m = 1 \) and \( \nu = 1 \) for the early and late phase, with a small
\( t_{1/2} = 0.1 \) for the early and a large \( t_{1/2} = 20 \) for the late phase as initial value, and estimated the intercept of the constant phase. A non-significant very small value of the intercept was observed for the constant phase, for this data. Based on this observation, we have removed the constant phase from the model.

2. Now, keeping \( m = 1 \) and \( \nu = 1 \) for the late phase, we tried 3 possible combinations of starting values for \( m \) and \( \nu \) ((1,1), (-1,1), (-1,1)) for the early phase and observed the convergence and likelihood estimates under these 3 scenarios. Based on the convergence and likelihood values (a larger one), it is noted \( m = 0 \) (that is, \( m \) with very small value closer to 0) and \( \nu < 0 \) was a best fit for the early phase. Note that, PROC NLMIXED provide both information criteria (AIC, BIC) and the actual likelihood. Since, we have not changed the number of parameters, we can either use information criteria or the likelihood for our model selection.

3. Now, keeping \( m = 0 \), \( \nu = -1 \) and \( t_{1/2} = 0.1 \). for the early phase, we tried 3 possible combination of starting values, as described above, for the late phase. Based on the convergence and likelihood values (a larger one), it was noted that \( m = 0 \) and \( \nu > 0 \) was a best fit for the late phase.

4. We now using \( m = 0 \), \( \nu = -1 \) and \( t_{1/2} = 0.1 \) for early phase and \( m = 0 \), \( \nu = 1 \) and \( t_{1/2} = 20 \) for late phase, we have obtained the final estimates given in Table 3.1.

**Temporal trend GFR:** Temporal trend analysis again suggested a bi-phase model. The estimates of the shaping and scaling parameters of each phase are given in Table 3.2. Based on the estimates in Table 3.2, the estimated multiphase temporal trend equation for GFR can simplified as follows:

\[
E(GFR^2|b_i) = \log \left[ \frac{1.99(2.55t)^{0.28}}{\exp(2.55t)^{1.28}} + \frac{1.05(0.72t)^{-2.46}}{\exp((0.72t)^{-1.46}) - 1} \right], \tag{3.12}
\]
### Table 3.2: Estimates of shaping parameters of the temporal trend of GFR.

Note that, in the early phase, \( m = 0 \) is the limiting case of \( g(t, \Theta) \) when \( \nu < 0 \) and \( m \to 0^+ \) as described in Case 3 in the mathematical formulation subsection 3.4.1, and in the late phase, \( m = 0 \) means the limiting case of \( h(t, \Theta) \) when \( \nu > 0 \) and \( m \to 0^+ \) as described in Case 1 in the mathematical formulation subsection 3.4.1.

\[
\text{GFR}_i^Z = \frac{\text{GFR}_i - 80}{40}, \quad \text{and } b_i, \text{ the patient-specific random effects for the phases, are assumed to be a normal variate. Note that, to make computations easier and faster (integration and optimization of the non-linear model), we have used the standardized value of GFR. The multiphase model yielded an early peaking and a late increasing phase (Figure 3.6). The early phase is from equation } g(t, \Theta) \text{ with limiting case 3 and the late phase is from equation } h(t, \Theta) \text{ with limiting case 1.}
\]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ± SE</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early decreasing phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}_{01} )</td>
<td>-0.49±0.046</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>( \hat{\nu} )</td>
<td>-0.78±0.0053</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>( m )</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \hat{t}_{1/2} )</td>
<td>0.29</td>
<td>-</td>
</tr>
<tr>
<td><strong>Late increasing phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\nu} )</td>
<td>0.68±0.023</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>( m )</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \hat{t}_{1/2} )</td>
<td>1.79</td>
<td>-</td>
</tr>
<tr>
<td>( \text{Var}(\epsilon) = \sigma^2_\epsilon )</td>
<td>0.37±0.0047</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>( \text{Var}(b) = \sigma^2_b )</td>
<td>0.46±0.034</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Figure 3.6: Decomposition of the temporal trend of GFR.
The temporal trend of the GFR data is shown in Figures 3.7 and 3.8. The average patient profile, average profile and average actual data were obtained as described above.

Figure 3.7: Temporal trend of GFR. Blue line is for average profile and the red closed circle symbols depict average actual data obtained by binning smoother.

Figure 3.8: Temporal trend of GFR. Blue line is for average profile. Thin pink lines are actual profiles of bilirubin (Logarithmic) for 100 randomly selected patients.

The average profile appears to follow the bin average actual data points closely.

**Goodness of Fit**

In general, information criteria used as a tool for comparing different models. That is, given more than one possible model, using information criteria, one can choose a
model. However, they do not test or provide some insight into how well a given model fits the data. In this section, we briefly describe and apply a goodness-of-fit measure to assess our temporal fits. Vonesh et al. [84] proposed a measure of concordance between predicted and observed response as a measure of goodness-of-fit for the response function. This measure is similar to the $R^2$ measure used in linear regression. Their guidelines for a good measure of goodness-of-fit are that it should: have an intuitively good interpretation, independent of unit of measurement, have well-defined points for perfect fit or complete lack of fit, have positive and negative values weighted equally, be applicable to any models regardless of underlying distributional properties, and be comparable across different models fit to the same data.

Lin [85] proposed a concordance correlation between two bivariate measurements $Y_1$ and $Y_2$ with means $\mu_1$, $\mu_2$, variances $\sigma^2_1$, $\sigma^2_2$ and correlation $\rho$, as follows:

\[
\rho_c = 1 - \frac{E(Y_1 - Y_2)^2}{\sigma^2_1 + \sigma^2_2 + (\mu_1 - \mu_2)^2}
\]

(3.13)

\[
= \frac{2\rho \sigma_1 \sigma_2}{\sigma^2_1 + \sigma^2_2 + (\mu_1 - \mu_2)^2}.
\]

Each pair $Y_1$ and $Y_2$, in the population are in perfect agreement (concordance) when $E(Y_1 - Y_2)^2 = 0$, and perfect disagreement occurs when $Y_1 = -Y_2$ and $\mu_1 - \mu_2 = 0$. For $n$ independent pairs of samples, the point estimate of $\rho_c$ is given by,

\[
\rho_c = \frac{2S_{12}}{S^2_1 + S^2_2 + (\bar{Y}_1 - \bar{Y}_2)^2},
\]

where $\bar{Y}_j$ and $S^2_j$ ($j = 1, 2$) are sample means and variances, and $S_{12} = \frac{1}{n} \sum_{i=1}^{n} (Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)$. Vonesh et al. [84] modified equation (3.13) to compare the observed and predicted values for a given model, under the setting of $n$ subjects each with $n_i$ number
of repeated measurements, as follows;

\[
(3.14) \quad r_c = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^T (y_i - \bar{y}) + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^T (\hat{y}_i - \bar{y}) + N(\bar{y} - \hat{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i) (y_i - \bar{y}_i)}
\]

where \(\bar{y}\) is the grand mean of the observed \(y_{ij}\), \(\hat{y}\) is the grand mean of the predicted \(\hat{y}_{ij}\), \(1_i\) is the \(n_i \times 1\) unit vector of 1’s and \(N = \sum n_i\) is the total number of observations. Note that, predicted value, \(\hat{y}_{ij}\) is obtained using parameter estimates and empirical bayes estimates of the random effects. Vonesh [84] pointed out the following advantages in using \(r_c\) as the concordance correlation coefficient between observed and predicted values: it directly measures the level of agreement between observed and predicted; unlike \(R^2\), we do not need a null model for reference, line of identity serves as the point of reference; \(r_c\) ranges from -1 to 1. That is, \(-1 \leq r_c \leq 1\), with perfect fit at 1 and lack of fit \(\leq 0\).

We now, estimate the concordance correlation (3.14) for the two temporal models.

<table>
<thead>
<tr>
<th>Model</th>
<th>(r_c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Trend of Log(Bilirubin) (Model 3.11)</td>
<td>0.87</td>
</tr>
<tr>
<td>Temporal Trend of GFR (Model 3.12)</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table 3.3: Goodness of Fit using Concordance correlation.

It can be seen from Table 3.3, that the concordance correlation for both temporal models are high and closer to 1 (perfect fit). Observed versus predicted plots of each data point are given in Figure 3.9. The fit of individual data points is appears to be very close in the temporal trend of Log(Bilirubin). In the temporal trend of GFR, however, there appears to be a lack of fit for very high values of GFR. It should be noted that, unlike the concordance correlation, the observed versus predicted plot, which plots the individual data points, does not indicate how well each patient profile is fitted.
Figure 3.9: Goodness of Fit: Observed versus Predicted for the temporal trend models (3.11) and (3.12).

Covariate effect on the temporal change

For simplicity, we have considered the following 5 variables in the multivariate analyses: Age, Gender, Device (LVAD vs other), baseline bilirubin, and baseline GFR. We have performed multivariate analyses to determine which variables influence the early phase, which influence the late phase and which influence both phases at an equal magnitude (at least approximately), so that these variables can be considered as overall common covariates. Since there is no built-in variable selection procedure in PROC NLMIXED, we have performed the following ad-hoc strategy in the data analysis to identify the phase-specific and overall covariates: we first forced in a variable in each phase to see if it was significant in that phase. If that variable was significant in a phase, we keep that variable in that phase, on the other hand, if that variable is significant in both phases and the parameter estimates were at least approximately equal, the variable was kept in the overall phase. Note that we have performed this ad-hoc variable selection by keeping shaping parameters fixed. However, once we have identified the phase-specific and overall covariates, in the final model we re-estimate all the covariate regression and shaping parameters.

Baseline covariate influence on the temporal trend of Log(Bilirubin)
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ± SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (^a)</td>
<td>0.86±0.14</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Baseline GFR (^b)</td>
<td>0.045±0.082</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>Early peaking phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender: Male</td>
<td>0.25±0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>MCS device: LVAD</td>
<td>-0.35±0.092</td>
<td>0.0002</td>
</tr>
<tr>
<td>Baseline bilirubin</td>
<td>0.37±0.031</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Shaping parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu)</td>
<td>-0.94±0.013</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>(t_{1/2})</td>
<td>0.0496</td>
<td>-</td>
</tr>
<tr>
<td><strong>Late increasing phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender: Male</td>
<td>0.22±0.10</td>
<td>0.03</td>
</tr>
<tr>
<td>MCS device: LVAD</td>
<td>-0.12±0.097</td>
<td>0.2</td>
</tr>
<tr>
<td>Baseline bilirubin</td>
<td>0.11±0.035</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Shaping parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu)</td>
<td>1.96±0.12</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table 3.4: Patient-specific risk factors for higher post-op Bilirubin.

\(^a\)Age/50  
\(^b\)GFR/60

Phase-specific and overall effects of covariates are given in Table 3.4. It appears older age associated with higher post-op bilirubin. Baseline GFR values does have any effect on the post-op bilirubin. Gender appears to have an effect on the both early and late post-op bilirubin values, with slightly lower effect on the late post-op than the early post-op bilirubin values. Mechanical device LVAD (compared to other devices) appeared to have a significant protective effect on the early post-op bilirubin, and the magnitude of the effect appears to be higher on the late post-op bilirubin values. Baseline bilirubin values appear to have an impact on the both early and post-op values of the bilirubin, with effect appears to be larger on the early post-op values. Further, note that there is only a slight change in the shaping parameters between the temporal trend model (model with only time, Table 3.1) and complete multivariate model (model with time and covariates, Table 3.4). That is, the shape of the underlying temporal trend has not changed when the covariate added to the model. Note that, since we have used \(T(t, \Theta) = h(t, \Theta)\) as the late phase time function, the \(t_{1/2}\) acts as an intercept (scaling,
not shaping) parameter in the late phase.

**Baseline covariate influence on the temporal trend of GFR**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ± SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age $^a$</td>
<td>-0.66±0.12</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gender: Male</td>
<td>0.012±0.076</td>
<td>0.8</td>
</tr>
<tr>
<td>Baseline Bilirubin</td>
<td>-0.0014±0.022</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>Early peaking phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCS device: LVAD</td>
<td>0.22±0.070</td>
<td>0.001</td>
</tr>
<tr>
<td>Baseline GFR $^b$</td>
<td>0.70±0.068</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Shaping parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>-0.77±0.0055</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>0.26</td>
<td>-</td>
</tr>
<tr>
<td><strong>Late peaking phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCS device: LVAD</td>
<td>0.65±0.11</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Baseline GFR$^b$</td>
<td>0.26±0.083</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Shaping parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.74±0.064</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table 3.5: *Patient-specific risk factors for Higher post-op GFR.*

$^a_{\text{Age/50}}$

$^b_{\text{GFR/60}}$

Phase-specific and overall effects of covariates are given in Table 3.5. Note that the direction of model coefficients are towards the higher GFR. However, since lower GFR is a marker for poor renal function, we will interpret the covariates in the direction of lower GFR. Older age appeared to be associated with lower GFR. Gender and baseline bilirubin does not have any effect on the post-op GFR. Mechanical device LVAD (compared to other devices) appeared to have a protective effect on both the early and late post-op GFR, and is associated with higher post-op GFR values and the effect appeared to be larger for the late post-op GFR. Lower baseline GFR was associated with lower post-op GFR, and the magnitude of the effect appears to be higher on the early post-op GFR values. Further, again note that there is only a slight change in the shaping parameters between the temporal trend model (model with only time, Table 3.2) and complete multivariate model (model with time and covariates, Table 3.5). That is, the shape of the underlying temporal trend has not changed when covariates added to the model. Again
here note that, the $t_{1/2}$ acts as an intercept (scaling, not shaping) parameter in the late phase.

### 3.4.4 A Simulation Study

We now carry-out a simulation study to assess performance of our multi-phase non-linear mixed effects model using the estimates of shaping parameters and of regression coefficients of the overall and phase-specific covariates. Note that, it should be noted that in this simulation study, we are not focusing on model building which is out of the scope of this analysis, rather, given a model (with given phases), we would like to assess how well the parameters are estimated.

**Simulation model**

We generate a continuous longitudinal response for 500 subjects at the following 13 time points over a 5-year time period: 1-day, 2-day, 3-day, 1-week, 2-week, 1-month, 3-month, 6-month and 5 time points at year 1 to year 5. We generated 2000 monte-carlo samples.

**Phases of the temporal trend**

![Figure 3.10: True shapes of the phases for temporal trend in the simulated model.](image)

Following the shapes of the temporal trend of the GFR, for the simulated temporal trend, we assume a two-phase model with $T_1(t, \Theta) = g(t, \Theta)$, with $\Theta = (m = 0.5, \nu = 0.5, t_{1/2} = 0.25)$, as the early phase and $T_2(t, \Theta) = h(t, \Theta)$, with $\Theta = (m = 0, \nu = 0.5)$. 

...
−0.75, \( t_{1/2} = 6 \), as the late phase. Note that, since \( T_2(t, \Theta) = h(t, \Theta) \), in this phase \( t_{1/2} \) acts as the intercept (a scaling parameter). Equation \( g(t, \Theta) \) with the selected values for the shaping parameters shows an early peaking function and \( h(t, \Theta) \) with the selected values gives a late increasing function (Figure 3.10).

For simplicity, we generated 4 binary covariates as follows: \( x_1 \sim \text{Binary}(0.6) \), \( x_2 \sim \text{Binary}(0.5) \), \( x_3 \sim \text{Binary}(0.3) \), and \( x_4 \sim \text{Binary}(0.4) \); \( x_4 \) is an overall covariate that positively influences the response \( y \) regardless of the time; while \( x_1 \) influences early values of \( y \) positively, it is negatively associated with late values of \( y \); \( x_2 \) is positively associated with only early values of \( y \); and \( x_3 \) is positively associated with only late values of \( y \). The random effect distributed as \( b_i \sim \mathcal{N}(0, 0.75^2) \) and the measurement error as \( \epsilon_i \sim \mathcal{N}(0, 0.5^2) \).

We then simulate the following 2-phase on-linear mixed effects model with 4 covariates as follows:

\[
y_i = 0.75x_{4i} + \log \{ \exp(-1 + x_{1i} - 0.5x_{2i})g(t, \Theta) + \exp(0.5x_{1i} + 0.5x_{3i})h(t, \Theta) \} \]

\[
(3.15) \quad b_i + \epsilon_i, \quad i = 1, \ldots, 500.
\]

**Simulation Results**

We assess the performance of the multi-phase non-linear mixed effects model based on simulated data using the following summary measures: suppose there are \( B \) simulated datasets and \( \beta \) is the true value and \( \hat{\beta}_i \) is the estimate from the \( i^{th} \) simulated dataset; \( \text{Bias} = \beta - \bar{\hat{\beta}} \), where \( \bar{\hat{\beta}} = \frac{\sum_{i=1}^{B} \hat{\beta}_i}{B} \); Average within standard error, \( \text{AvgSE} = \frac{\sum_{i=1}^{B} SE(\hat{\beta}_i)}{B} \); Empirical standard error, \( \text{EmpSE} = \sqrt{[1/(B-1)] \sum_{i=1}^{B} (\hat{\beta}_i - \bar{\hat{\beta}})^2} \); Coverage probability.

The summary measures of the shaping and regression coefficients of the covariates based on the 2000 simulated data are given in Table 3.6.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias</th>
<th>AvgSE</th>
<th>EmpSE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_4$</td>
<td>-0.0004</td>
<td>0.06949</td>
<td>0.07085</td>
<td>0.951</td>
</tr>
<tr>
<td>Early peaking phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_1$</td>
<td>-0.0012</td>
<td>0.07327</td>
<td>0.07236</td>
<td>0.948</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.0000</td>
<td>0.03188</td>
<td>0.03095</td>
<td>0.954</td>
</tr>
<tr>
<td>Shaping parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.0000</td>
<td>0.00882</td>
<td>0.00879</td>
<td>0.950</td>
</tr>
<tr>
<td>$m$</td>
<td>-0.0003</td>
<td>0.01444</td>
<td>0.01436</td>
<td>0.956</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>0.0000</td>
<td>0.00253</td>
<td>0.00252</td>
<td>0.950</td>
</tr>
<tr>
<td>Late increasing phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_1$</td>
<td>-0.0001</td>
<td>0.07050</td>
<td>0.06972</td>
<td>0.952</td>
</tr>
<tr>
<td>$x_3$</td>
<td>-0.0010</td>
<td>0.03424</td>
<td>0.03473</td>
<td>0.949</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>-0.0178</td>
<td>0.28188</td>
<td>0.28504</td>
<td>0.956</td>
</tr>
<tr>
<td>Shaping parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.0000</td>
<td>0.00162</td>
<td>0.00158</td>
<td>0.954</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_b$</td>
<td>0.0039</td>
<td>0.03655</td>
<td>0.03709</td>
<td>0.944</td>
</tr>
<tr>
<td>$\sigma^2_\epsilon$</td>
<td>0.0003</td>
<td>0.00456</td>
<td>0.00451</td>
<td>0.954</td>
</tr>
</tbody>
</table>

Table 3.6: Summary measures of the parameter estimates based on the 2000 simulated data.

All the estimated parameters are closer to the true values of the parameters and in some cases the average of the estimates are almost equal to the true values. This suggests a good performance of our non-linear mixed effect model in estimating the parameters.

We now use the same simulated data sets, but fit models without the covariates $(x_1, x_2, x_3, x_4)$ in the model. The main objective of this exercise is just to illustrate that even though the data were simulated using the multi-phase model with 4 covariates (3.15), model fitted without covariates in the model still keep the same underlying shape. That is, a temporal trend model has the same shape as the shape of the model with the covariates in the model. The Figure 3.11 compares the true shape with the shapes obtained under the following 3 summary measures of the estimates from the simulated data; i). average of the estimates; ii). 2.5% percentile of the estimates; iii). 97.5% percentile of the estimates. It is clearly shows the model without covariates still kepps the same underlying shapes of the phases.
3.5 A Multivariate Non-Linear Mixed Effect Model

We now extend the univariate non-linear multiphase model in equation (3.6) to a multivariate set up where we have $K$ correlated longitudinal responses as follows:

$$
Y_1^i = X_{i0}^1 \beta_o^1 + \log \left( \sum_{l=1}^{L_1} \mu_l(X_{il}^1; \beta_l^1) T_l(t_i, \Theta_l^1) \right) + b_i^1 + \epsilon_i^1
$$

$$
Y_2^i = X_{i0}^2 \beta_o^2 + \log \left( \sum_{l=1}^{L_2} \mu_l(X_{il}^2; \beta_l^2) T_l(t_i, \Theta_l^2) \right) + b_i^2 + \epsilon_i^2
$$

$$
: = :$

$$
Y^K_i = X_{i0}^K \beta_o^K + \log \left( \sum_{l=1}^{L_K} \mu_l(X_{il}^K; \beta_l^K) T_l(t_i, \Theta_l^K) \right) + b_i^K + \epsilon_i^K,
$$

where $\log(\cdot)$ is a vector as the same dimension as $Y^K_i$ ($k = 1, \ldots, K$). The main features of this multivariate extension are:

1. In general, the longitudinal responses can be of different measurement scale. For example, $Y_1^i$ can be a longitudinal continuous response and $Y_2^i$ can be longitudinal binary response. However, for this paper, we consider the model with all the $K$
longitudinal responses are continuous.

2. As described in section 3, for each patient \( i, i = 1, \ldots, n \); \( n_i^j \), the number of repeated measurements and \( T_i^j \), the set of time points at which longitudinal response \( j \) is observed, may be different from \( n_i^k \), the number of repeated measurements and \( T_i^k \), the set of time points at which longitudinal response \( k \) is observed, where \( j \neq k \in \{1, \ldots, K\} \). However, as in our motivating example and in most of the observational studies, there may be some overlapping time points. That is, both responses \( j \) and \( k \) may have been observed simultaneously at some time points. In other words, \( T_i^j \cap T_i^k \neq \emptyset \);

3. Each response can have a different number of phases. That is, in general, we can have \( L_k \neq L_{k'} (k \neq k' \in \{1, \ldots, K\}) \) for responses \( k \) and \( k' \);

4. Each response can have different or overlapping covariate sets;

5. There are two general approaches for linking the multivariate responses together:

(a) By the joint distribution of the random effects. Let \( b_i = [b_i^1, \ldots, b_i^K] \) then \( b_i \sim N(0, G) \), where \( G \) is the \( K \times K \) variance-covariance matrix of the random effects. Off-diagonal elements of \( G \) provide the measure of association among the \( K \) longitudinal responses;

(b) By the joint distribution of the measurement errors among the \( K \) responses. In addition to linking the \( K \) responses through the variance-covariance matrix of the random effects, the \( K \) responses can be further linked by the measurement errors \( e_i = [e_i^1, \ldots, e_i^K] \). We assume the measurement errors from different subjects are not correlated. We further assume random effects \( b_i \) and measurement errors \( e_i \) from the same subject \( i \) are not correlated. That is, \( e_i \perp b_i, \forall i \in \{1, \ldots, n\} \). In our case of continuous \( K \)-multivariate responses, we assume \( e_i \sim N(0, \Sigma \otimes I_{n_i}) \), where \( \Sigma \) is the \( K \times K \) covariance
matrix of measurement errors for the $K$ responses measured at that same time point a given subject $i$ and $n_i$ is the number of time points at which all $K$ responses are observed. The Kronecker product $\Sigma \otimes I_{n_i}$ implies while the measurement errors between the subjects are uncorrelated, measurement errors for different responses from the same subject are correlated.

When the variance-covariance matrix of measurement errors is a diagonal matrix,

$$
\Sigma = \begin{pmatrix}
\sigma_{11} & 0 \\
0 & \sigma_{22} \\
& & \ddots \\
0 & & & \sigma_{KK}
\end{pmatrix},
$$

it is assumed that we have a conditional independence (or pure random effect) model. That is, we have $\epsilon_i^j \perp \epsilon_i^k$, $\forall j \neq k \in \{1, \ldots, K\}$. Because of the ease in estimation and application, especially using readily available software, many have proposed different joint models under the assumption of conditional independence (for example [71],[77],[86]). Further, under this assumption, for each subject, the time points at which the $K$ responses were collected, can be completely different for different responses. That is, we can have $\mathcal{T}_i^j \cap \mathcal{T}_i^k = \emptyset$, $\forall j \neq k$. In addition, under the conditional independence assumption, since we do not link the error terms from different responses, conditional independence models have straightforward interpretation when using responses of different nature (example, continuous and binary).

When all the $K$ responses are observed at the same time points for subject $i$ ($i = 1, \ldots, n$), that is, when the sets $\mathcal{T}_i^1 \equiv \mathcal{T}_i^2 \equiv \ldots \equiv \mathcal{T}_i^K$ (thus, $n_i^1 = n_i^2 = \ldots = n_i^K = n_i$), Faes et al [87] proposed a joint model that fits longitudinal outcomes of different nature (example, continuous and binary) under the assumption of condi-
tional independence. Marshall et al. [76] fitted a non-linear random effect model for multivariate responses with missing data under the assumption of correlated measurement errors. However, if $T_i^j \cap T_i^k = \emptyset$, $\forall j \neq k$, then the off-diagonal terms of $\Sigma$ are not identifiable and hence, can not be estimated.

### 3.5.1 Estimation

As described above, assume that we have $K$ multivariate longitudinal responses $Y_1^i, \ldots, Y_K^i$, with each observed at $n_i$ number of time points for subject $i \in (1, \ldots, n)$. Then the marginal likelihood is given by,

\[
L(\beta, G, \Sigma; y^1, \ldots, y^K) = \prod_{i=1}^{n} \int_{\mathbb{R}^K} \prod_{j=1}^{n_i} f_{y^1_j, \ldots, y^K_j | b}(y^1_j, \ldots, y^K_j | b_i; \beta, \Sigma) f_b(b_i; G) db_i,
\]

where $f_{y^1_j, \ldots, y^K_j | b}(\cdot | \cdot)$ is the joint distribution of $Y^1_j, \ldots, Y^K_j$ conditional on the random effects $b$. In this study, we assume that the responses are continuous and the conditional joint distribution is a multivariate normal. $f_b(\cdot)$ is the joint distribution of the random effects. Since $b_i$ is a $K$ vector, the integration in (3.17) involves with a $K$-dimensional integration.

We use Gaussian quadrature technique to approximate the marginal likelihood by numerical integration. This technique essentially approximate the marginal density given in (3.17) by a weighted average of the integrant evaluated at $Q$ pre-determined abscissas (quadrature points) $z_q$ over the random effects $b_i$ [67]. The standard Gauss-Hermite weights $w_q$ and abscissas $z_q$ ($q = 1, \ldots, Q$) can be obtained from tables from the *Handbook of Mathematical Functions* [88]. That is, the integral in (3.17) can be approximated as,
\[ \mathcal{L}(\beta, G, \Sigma; y^1, \ldots, y^K) \approx \prod_{i=1}^{n} \sum_{r_1}^{Q} \cdots \sum_{r_K}^{Q} \left[ \prod_{j=1}^{n_i} f_{y^1_1, \ldots, y^K_1 | b} (y^1_{ij}, \ldots, y^K_{ij} | a_{r_1, \ldots, r_K; \beta, \Sigma}) \right] \]

where, if Gaussian quadrature is adaptive \( a_{j_1, \ldots, j_K} \) is a function of empirical bayes estimate of random effects, abscissas and hessian matrix. If Gaussian quadrature is non-adaptive \( a_{j_1, \ldots, j_K} \) is a function of abscissas and hessian matrix. See, for example, Pinheiro and Bates \[67\] for details.

Note that, since our responses are assumed to be normally distributed, one can use a first-order approximation \[89\] \[90\] to the conditional distribution \( f_{y^1, \ldots, y^K | b} (\cdot | \cdot) \) in the integral in (3.17). The approximation is essentially a Taylor series expansion of conditional mean response with respect to random effects around zero, which is the expected value of the random effects. When the conditional distribution is a multivariate normal and the random effects are from a multivariate normal distribution, using the first-order approximation, the integral (3.17) can be approximated in a closed form. We then can optimize the approximation using the Quasi-Newton method. This approximation can be implemented using PROC NLMIXED. Note that, while the Gaussian quadrature technique usually called the exact method approach likelihood, the latter is called the approximate approach.

**Remark:** In the case where the conditional distribution is a bivariate normal, one can rewrite the conditional distribution as a the product of two univariate normals. That is, \( f_{y^1, y^2 | b} (\cdot | \cdot) = f_{y^1 | b} (\cdot | \cdot) f_{y^2 | y^1, b} (\cdot | \cdot) \). Now these univariate representations can be easily implemented using readily available software (PROC NLMIXED). That is, suppose
\[
\begin{pmatrix}
Y^1 \\
Y^2
\end{pmatrix} \sim N \left( \begin{pmatrix} M_1(b) \\ M_2(b) \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right),
\]

where \( M_k(b) = \log \left( \sum_{t_1=1}^{L_k} \mu_{t_1}(X_{t_1}; \beta^k)T(t, \Theta^k_t) \right) + b^k, \ k = 1, 2. \) Then
\[
Y^1 \sim N (M_1(b), \sigma_{11}) \quad \text{and} \quad Y^2|Y^1 \sim N (M_2(b) + \sigma_{12}(Y^1 - M_1(b))/\sigma_{11}, \sigma_{22} - \sigma_{12}^2/\sigma_{11}).
\]

Note that, under the assumption of conditional independence, likelihood further simplifies, since \( f_{Y^1,Y^2|b}(\cdot|\cdot) = f_{Y^1|b}(\cdot|\cdot)f_{Y^2|b}(\cdot|\cdot). \) That is, under the assumption of conditional independence, we have \( \sigma_{12} = 0. \)

In a related note, Fieuws and Verbeke [74] proposed a pair-wise model fitting approach when there are high-dimensional responses. Hence, in principal, for estimation purpose, one can reduce any multivariate longitudinal problem into a series of bivariate problems and each problem is solved individually and can be combined to obtain the overall estimation using a "Pseudo likelihood" approach. See for example, Fieuws et al.[74] and Faes et al. [87] for applications of high-dimensional longitudinal responses.

Since in the data analysis we perform model selection using information criteria, below we briefly describe the information criteria.

**Information Criteria**

Before we use an information criteria for model selection, we first briefly describe the information criterion we plan to use and its origins. In terms of statistical principles, model selection involves a trade-off between bias and variance. While inference based on too few variables (parameters) can be biased, a model with too many variables (parameters) can have higher variance. Hence attaining a correct model using some information criteria is an important part of model building. Information criteria were originally founded in information theory and are based on the Kullback-Leibler definition [91] of information lost as follows: Suppose \( f \) is the true model and \( g(x|\beta) \) is the proposed model
then the *Kullback-Leibler* information lost is defined as:

\[ I(f, g(.|\beta)) = \int_\Omega f(x) \log \left( \frac{f(x)}{g(x|\beta)} \right) dx. \]

It can be noted, when the proposed model \( g(x|\beta) \) is closer to the true model \( f(x) \), \( I(f, g(.|\beta)) \) gets smaller. See, for example, Burnham and Anderson [92], for a detailed description of \( I(f, g(.|\beta)) \). Akaike [93], adopting the *Kullback-Leibler* definition of information lost and using likelihood theory showed that, for large samples, the minimization of the expected information lost is equal to the maximization of \( E(\int \log g(x|\beta)f(x)dx) \) and is approximately equals to

\[ \log \left( \mathcal{L}(\hat{\beta}) \right) - K, \]

where \( \mathcal{L} \) is the likelihood function, \( \hat{\beta} \) is the maximum likelihood estimate of \( \beta \), and \( K \) is the number of estimated parameters. The Akaike’s information criteria is given by

\[ AIC = -2 \log \left( \mathcal{L}(\hat{\beta}|y) \right) + 2K. \]

The smaller the \( AIC \), the better the model. It can also be viewed as the extension of likelihood with a penalty for each added parameter in the model. When \( K \) is large relative to the sample size \( n \) or when we have a smaller sample size \( n \), a corrected version of \( AIC \) is given as,

\[ AICc = AIC + \frac{2K(K + 1)}{n - K - 1} \]
\[ = -2 \log \left( \mathcal{L}(\hat{\beta}|y) \right) + \frac{2nK}{n - K - 1}. \]

When \( n \) is large it is shown that \( AIC \approx AICc \). Hence, it is suggested to use \( AICc \) in any situation. As \( AIC \) values are influenced by the sample size, \( AIC \) or \( AICc \) values
are usually rescaled as follows before comparing a set of models: Suppose there are $R$ models with $AIC_r$ for $r = 1, \ldots, R$ with $\min\{AIC_r\} = AIC_{\min}$. Then $AIC$ or $AIC_c$ are compared using the following relative differences:

$$\Delta_i = AIC_i - AIC_{\min}.$$ 

Some simple rules of thumb often applied to $\Delta_i$ are; models having $\Delta_i \leq 2$ have strong evidence of being best models; models having $4 \leq \Delta_i \leq 7$ have considerably less support; models having $\Delta_i \geq 10$ have essentially no support of being best models.

Schwartz \[94\] introduced the Bayesian Information Criteria (BIC or SBIC) as:

$$BIC = -2 \log(\mathcal{L}(\hat{\beta}|y)) + 2K \log(n).$$

It is noted that the penalty term in $BIC$ is larger than the penalty term in $AIC$. Based on simulation studies, Burnham and Anderson \[92\] suggested, in general $AIC$ performs better than $BIC$. The main advantages of information criteria such as $AIC$ is that it is valid for both nested and nonnested models, can compare models with different error distributions. But it can not be used for different data sets.

3.5.2 Data Analysis

Four hundred and thirty-four patients (5 patients were excluded from the original cohort of 439, because of no post-MCS GFR and bilirubin is available for these 5 patients) who were transplant candidates (listed for transplant) and were bridged with MCS from December 1991 to July 2009 at the Cleveland Clinic are considered in this data analysis. A total of 9620 lab records on post-MCS bilirubin and GFR are available for 434 patients. Both bilirubin and GFR are available in all 9620 records. Among these, 43 (10%) patients had only one record, 38 (9%) had only two records and the remaining had at least 3
The mean follow-up time was 77 days (median=34 days, P_{25}=14 days, P_{75}=87 days).

The main objectives of this data analysis are to jointly model the temporal evolutions of renal function (GFR) and liver function (bilirubin) while on MCS, and to evaluate different types of modeling assumptions. We use the logarithm of bilirubin ($Y^1$) and GFR ($Y^2$) as the longitudinal responses and assume a bivariate normal distribution. We define the joint model as follows:

\[
\begin{align*}
Y^1_i &= X^1_{i0}\beta^1_o + \log \left( \sum_{l=1}^{L_1} \mu_l(X^1_{i1}; \beta^1_l)T_l(t_i, \Theta^1_l) \right) + b^1_i + \epsilon^1_i \\
Y^2_i &= X^2_{i0}\beta^2_o + \log \left( \sum_{l=1}^{L_2} \mu_l(X^2_{i1}; \beta^2_l)T_l(t_i, \Theta^2_l) \right) + b^2_i + \epsilon^2_i.
\end{align*}
\]

We first performed temporal trend modeling for each response separately to identify the phases and baseline variable associated with each responses. We then used the resulting models in the joint modeling. The data suggested an early decreasing and a late increasing phase for the temporal trend of the logarithm of bilirubin and an early peaking and late increasing phase for the temporal trend of GFR, we have used these phases in the joint modeling (see Section 3.4.2 for details). That is, we have $b_i = [b^1_i, b^2_i]^\top$ with $b_i \sim N(0, G)$, where $G$ is the $2 \times 2$ variance-covariance matrix of the random effects.

We further assume $\epsilon_i = [\epsilon^1_i, \epsilon^2_i]^\top \sim N(0, \Sigma \otimes I_{n_i})$, where $\Sigma$ is a $2 \times 2$ matrix specifies the associations between the measurement errors between the 2 responses for a given subject $i$ and $n_i$ is the number of time points at which both responses are observed. Based on different formations of $G$ and $\Sigma$, one can obtain different models that have different correlation structures as follows:

1. **Independence model**: Under this scenario, $G$ is a block diagonal and $\Sigma$ is a
diagonal. That is, this is equivalent to fitting separately;

2. **Joint model: Conditional Independence model**: Under this scenario, $G$ is an unstructured $2 \times 2$ matrix and $\Sigma$ is a diagonal; That is, the two responses are linked using a joint distribution of random effects;

3. **Joint model: Full**: Under this scenario, $G$ is an unstructured $2 \times 2$ matrix and $\Sigma$ is a $2 \times 2$ unstructured matrix; In this case, the two responses are joined using joint distributions of random effects and measurement errors.

The estimates of the shaping parameters for these 3 models are given in Table 3.7. The estimates of the shaping and the scaling parameters are almost identical for all three models. Comparison of AICC information criteria shows the **Joint full model** appeared to be a better model than the other two. However, while the difference in the AICC between **Joint conditional independence model** and **Joint full model** is only 79, the difference between **Joint conditional independence** and **Independent** models is much larger at 284. Further, the correlation of the measurement errors between the two responses appeared to be very small (-0.11) and relaxing of the conditional independence assumption did not appear to change the variance estimates in $\Sigma$. However, there appears a slight change in the variance-covariance estimate of the random effects. Under the conditional independence assumption, the correlation between the random effects is -0.55. That is, there is a moderate negative association between bilirubin and GFR. That is, the higher the bilirubin (severe the liver dysfunction), the lower the GFR (severe the renal dysfunction). Under the **Joint full model** this correlation between the random effects slightly reduces to -0.49.

**Remark**: In addition to the above three types of modeling approaches, we have considered a fourth type where the two responses are linked only through the measurement errors. That is, $\Sigma$ is an unstructured $2 \times 2$ matrix and $G$ is a diagonal matrix. The
AICC information criterion for this model was 38353. Note that this is only better than the *Independence model (separate models)*.
Table 3.7: Estimates of shaping and scaling parameters of the temporal trends of logarithm of bilirubin and GFR.

Note that since $T_2(t, \Theta) = \theta(t, \Theta)$, $t_1/2$ is the scaler (intercept) in this phase.
Chapter 4

Joint Modeling of Multivariate Longitudinal & Competing Risks Data

In this chapter, we briefly review the current literature on joint modeling of time to event(s) and longitudinal outcome(s). We then propose an approach for joint modeling of multivariate longitudinal data and competing risks data using a multivariate non-linear mixed effects model and cause-specific hazards frailty models.

4.1 Introduction

In many medical studies that involve longitudinal follow-up, it is common to observe one or more sequences of longitudinal measurements, as well as one or more time to event outcomes. For example, in a study of durability of mitral valve after the valve repair, it is common to focus on the longitudinal graded observation of mitral valve regurgitation (severity of valve leakage). We continue to collect this longitudinal observation until the end of administrative follow-up, until the patient undergoes reoperation on the valve,
or until the patient dies. A classical example is longitudinal CD4 counts and time to progression to AIDS or death (for example, Pawitan and Self [95]). In our motivating example, longitudinal measurements of renal and liver function were collected until the patient died or until he/she received a heart transplant. Since these patients are listed for heart transplant they are systematically followed until they reach one of the two competing risks. When one encounters these two types of data, the usual scientific objectives can be one or more of the following:

i. To assess the effect of longitudinal measurement(s) on the time-to-event(s) of interest. Here, the main focus is on the time-to-event(s), and one would like to make inference on or prediction of time-to-event(s) based on the longitudinal measurement(s) observed thus far. When this prediction is of interest, one basically treats the longitudinal outcome as a marker for the time-to-event outcome;

ii. To assess the pattern (or temporal trend) of the longitudinal measurement(s) after adjusting for the fact that longitudinal measurements were censored due to the time-to-event(s). Hence, here the main focus is on longitudinal measurements, treating the time-to-event(s) as a nuisance variable;

iii. To assess the relationship between the longitudinal outcome(s) and time-to-event(s) simultaneously.

When the objective is (i), the simplest and most straightforward approach is to treat the longitudinal process as a time-varying covariate in the time-to-event(s) analysis. Following the notation of Kalbfleisch and Prentice (Ch.6, [26]), suppose $X_i(t) = \{x_i(u) : 0 \leq u < t\}$ is the longitudinal covariate history for subject $i$ at time $t$. The longitudinal process can be incorporated into the hazard function of the event at time $t$, for example, as follows, $\lambda(t) = \lambda_0(t) \exp{\gamma X_i(t) + \eta^\top Z_i}$, where $Z_i$ baseline covariates for subject $i$ and $\lambda_0(t)$ is the baseline hazard. The parameter $\gamma$ signifies the effect of
the longitudinal data on the time to event. However, in practice, there is variability in the longitudinal measurements due to measurement error and biological variation. For example, FEV1 is a marker for pulmonary function and there may be variation in this marker over time due to minor fluctuations in health status and to a number of sources of laboratory measurement error including variation among technicians, changes in equipment, brands of laboratory kits and repetition error. Therefore, if one does not account for the variability due to measurement error, the regression coefficient $\gamma$ in the hazard model is biased. Prentice [96] argued that when this is the case, the estimated regression coefficient $\gamma$ will be biased towards the null. To address this problem, several authors have proposed different two-step approaches. See, for example, Raboud et al. [97], Tsiatis et al. [98] and Bycott and Taylor [99], for details of different two-step approaches. Simply speaking, the two-step approach is as follows; suppose $X_i(t)$ is the observed longitudinal process and $X^*_i(t)$ is the corresponding true process. In the first step, we estimate the true process using some parametric or non-parametric longitudinal regression method as $X_i(t) = X^*_i(t) + \epsilon_i(t)$, where $\epsilon_i(t)$ the measurement error. In the second step, we incorporate the estimated true process as a time-varying covariate in the time-to-event hazard model as $\lambda(t) = \lambda_0(t) \exp\{\gamma \hat{X}^*_i(t) + \eta^\top Z_i\}$. However, even though the two-step approach reduces the majority of the bias compared to naive direct approach, it still has some drawbacks. Wulfson and Tsiatis [100] argued that the two-stage model does not use any time-to-event information in modeling the longitudinal process, hence, it may not be efficient, and there will still some residual bias in the regression coefficient $\gamma$ if there is a violation in the normality assumption of the random effect; Hence, one needs a joint modeling approach where joint maximization of the likelihood from both the longitudinal process and the time-to-event(s) data is carried out in order to make more efficient use of all the data simultaneously.

Under the scientific objective (ii), our focus is in making inference on the longitudinal process. However, the presence of time-to-event(s) may influence the longitudinal
process. Note that, in this case, time-to-dropout (or withdraw from a study) can also be considered as a time-to-event. For example, in an observational study of durability of mitral valve repair, for the first 6 months after the repair, because regular visits are scheduled ejection fraction is measured (based on echocardiogram) for (almost) all the patients. However, only the patients who may feel sick or deemed to be sicker by their physician may have their ejection fraction measured after six months. Simple longitudinal methods without appropriate adjustment may reveal that there is a sharp decrease in the ejection fraction after six months in this cohort. However, this is due to the fact that normal patients drop out of the study after six months and may not reflect the true ejection fraction in the whole cohort. On the other hand, in an HIV study, subjects who experience the event AIDS may drop out of the study; hence, only healthier patients stay in the study and thus, we observe an increase in the CD4 count. Again, using simple longitudinal techniques without appropriate adjustment can lead to biased estimation of some average quantity of interest. In our motivating example, sicker patients with very high bilirubin values may die early, hence we observe bilirubin values of the healthier patients after some time. The opposite is true for the transplanted patients. Hence, one needs to take into the account of drop-out process (for example, death, transplant, healthy or sicker) when estimating some quantity of the longitudinal process. This can be done through some conditional (on the drop out) or joint modeling approach. Once a drop-out occurs, the longitudinal process is deemed to be missing.

Rubin [28] categorized the missing value process into 3 categories as follows: A dropout process is missing completely at random (MCAR), if the dropout process is independent of observed and unobserved data; missing at random (MAR), if the dropout process is dependent on the observed but not on the unobserved values; non missing at random (NMAR), if otherwise. The last case is also called informative or nonignorable dropout. When the drop out process is MAR, one can use likelihood based methods to obtain valid inference (see for example, Little and Rubin [101], Little [102] for details).
However, when the drop-out process is nonignorable, the drop-out mechanism needs to be modeled along with the longitudinal process.

Suppose $Y$ is the longitudinal process and $T$ is the drop-out (or time-to-event) mechanism, Rubin [28] introduced the following two types of factorizations of the joint distribution of $(Y, T)$:

1. **Selection models**: $f(Y, T) = f(T|Y)f(Y)$, Little [102] argued that this approach models the hypothetical complete data and then appends a model for missing data process conditional on the hypothetical complete data. Wu and Carroll [103] used probit regression (conditional on the longitudinal process regression coefficient) as the selection model. Diggle and Kenward [104] used a logit model to model the drop out process.

2. **Pattern-mixture models**: $f(Y, T) = f(Y|T)f(T)$, these models stratify the population of interest by the patterns of the drop-out mechanism and the unconditional model is the mixture (average) of the conditional ones (patterns). Li and Schluchter [15] used a pattern mixture model to adjust for non-ignorable drop-out in the MDRD study. Hedeker and Gibbons [105] used random effects pattern-mixture model to adjust for missing data in the longitudinal Inpatient Multidimensional Psychiatric Scale in a psychiatric clinical trial. Hogan and Laird [106] proposed a joint model using pattern-mixture approach to model CD4 counts and survival.

Further, with random-effects in the models, Little [102] called these models random coefficient selection model and random coefficient pattern mixture models, respectively. Little [102], Hogan and Laird [107], for example, provide a detailed review of the these models. Guo et al. [108] used a random pattern-mixture model to adjust for drop-out and analyzed a longitudinal continuous depression outcome. A detailed overview of selection
and pattern mixture models can be found, for example, in Verbeke and Molenberghs (ch. 16-18, [55]).

Note that, in addition to the two types of factorization described above, another third type of factorization of the joint distribution of \((Y, T)\) is called a shared parameter model. We discuss this approach in the next section.

4.2 Longitudinal and Time-To-Event Data

We now briefly review the joint modeling approach. There is an extensive literature on detail discussions on joint modeling. See, for example, Schluchter [109], Tsiatis and Davidian [110], and Diggle et al. [111] for historical overview of the joint modeling approaches. Simply, speaking the basis of the joint modeling approach is to specify the joint distribution, \(f(Y, T)\) of the longitudinal process, \(Y\) and time-to-event process, \(T\). That is, a joint model involves two linked (sub) models, one for the longitudinal model and the other for the time-to-event model. The most popular approach is to introduce a latent process, \(U\) in the two sub models, the the joint distribution \(f(Y, T)\) can be factorized into conditional distributions as follows:

\[
(4.1) \quad f(Y, T) = \int_U f_{Y|U}(Y) f_{T|U}(T) f_U(U) dU,
\]

where \(f_{Y|U}\) is a conditional distribution of the longitudinal process given the random effects \(U\), generally, a linear or non-linear mixed effect model to depict the longitudinal data; \(f_{T|U}\) is a conditional distribution of the time-to-event data given the random effects \(U\), generally, a proportional hazard model; and \(f_U\) is the distribution of random effects, generally a Gaussian. Note that, Song et al. [112] proposed a model with Gaussian assumption of \(U\) is relaxed. The specification (4.1) essentially states that the longitudinal
process and time-to-event data are conditionally independent. That is, given the latent process $U$ which measures the underlying unobserved process of a patient, for example, a patient’s health status, the distributions of longitudinal process and the time-to-event data are independent. In other words, in most of the cases, one can justify the conditional independence assumption by stating that, once the latent process is given, time-to-event is independent of any surrogate marker (longitudinal process). The two sub-models are linked through their shared dependence on the latent process $U$. The possible association between the longitudinal process and time-to-event is carried through the shared random effect $U$. Hence, these models are called shared parameter models.

Wulfsohn and Tsiatis [100] proposed a joint model where the longitudinal process was modeled through a linear mixed effects model (random intercept and random slope) and time-to-event data was modeled using a proportional hazards model with the random effects as covariates, and used the EM algorithm for parameter estimation. Pawitan and Self [95] proposed a joint model to model longitudinal $T4$ counts and time-to-infection and time-to-AIDS with (a fully parametric) Weibull regression model for time-to-event data with the distributions of these time-to-events are assumed to be independent given the baseline covariates. Further, the longitudinal process depended on the time-to-events. Henderson et al. [113] further extended the joint modeling formulation using two separate latent processes, one for each sub model (one for longitudinal and one for the time-to-event data) with these two latent processes are joined by a bivariate Gaussian. On the other hand, Schluchter [109] proposed a joint modeling approach where the longitudinal process and time-to-event data are linked through a trivariate Gaussian, where patient-specific random intercepts, random slopes and transformed survival times are linked through a trivariate normal distribution. He has mainly used log-normal transformation for survival time. He further shown that this trivariate Gaussian set up can be rewritten as selection or pattern mixture or shared parameter formulations. A detailed description and application of this approach in MDRD study is found in [114].
Vonesh et al. [115] used a shared parameter model to analyze the MDRD data using readily available software. Liu [116] jointly modeled 3 sub-models; a model for the odds of hospital visits, a model for medical costs, and a model for survival using a shared parameter model. Note that, here, the medical cost is a semi-continuous longitudinal data. Ratcliffe et al. [117] used a shared frailty model where the longitudinal process has a multi-level random effects (multiple hospitals as the cluster level, and patients within each hospital as the patient-level) and the longitudinal process and time-to-event are linked at the cluster level. Liu et al. [118] further generalized this shared random effect model where the longitudinal process and time-to-event are linked at both the cluster level and patient level. The joint modeling approach is not only being used in the field of medicine but also in other fields. Wintrebert et al. [119] used a joint shared frailty model to model the odds of breeding and survival of birds in an ecological study. They have used the Bayesian approach using the WinBUGS package [120] to estimate the regression parameters. Recently, there have been numerous joint modeling approaches that used readily available software for parameter estimation. For example, see Guo and Carlin [121] for detail description of the estimation of regression parameters of a joint modeling approach using SAS and WinBUGS.

4.2.1 Joint Modeling of Multivariate Longitudinal and Time-to-Event Data

It is now common to observe more than one longitudinal outcome. Hence, it is natural to consider joint modeling of multivariate longitudinal data and time-to-event data. All of the joint modeling approaches described in the case of univariate longitudinal data and time-to-event data can be extended to accommodate multivariate longitudinal data. The joint modeling of multivariate longitudinal data is the extension that is needed in this case. As described in the previous chapter, multiple longitudinal outcomes can be
linked together through correlated random effects and/or through the joint distribution of measurement errors. Lin et al. [122] extended Wulfsohn and Tsiatis’s model [100] to accommodate multivariate longitudinal outcomes. In their extension, Lin et al. [122] linked the multivariate longitudinal outcomes through the joint distribution of random effects and used a gamma frailty model with the conditional (on random effects) expectation of the longitudinal process (true process) as a covariate when modeling the time-to-event data. On the other hand, Thiebaut et al. [72] linked multivariate longitudinal outcomes and time-to-event through a joint distribution of the random effects and survival times. They, in fact used logarithmic transformation of survival time as the survival time variate and a multivariate normal as the joint distribution. This approach thus is an extension of the trivariate approach proposed by Schluchter [109].

4.2.2 Joint Modeling of Longitudinal Data and Competing Risks

Data

A natural extension in the joint modeling literature is to the case where a study involves multiple failure types and longitudinal process(es). To the best of our knowledge, only a few such extensions currently exist in the literature. The most common example of multiple failure types is competing risks where a patient is exposed to risk of more than one possible type of events. We now briefly summarize recent developments in joint modeling of longitudinal data and competing risks data.

Most of the work on this field is done by Elashoff et al. [123]. In their first paper, Elashoff et al.’s [123] main focus was the effect of a treatment on the longitudinal process of lung function %FVC and a time-to-event, treatment failure or death. However, they considered the dropouts due to adverse event or serious adverse event or worsening disease as informative, and treated it as a competing risk to the event of interest, treatment failure or death. Thus, their study has two competing risks and a longitudinal process.
They have used a mixture model of Larson and Dinse [44] to model the competing risks; a mixture of logistic model (for marginal distribution of failure type) and a proportional hazards model for each failure type. The mixture model and the linear mixed effects model was linked through the random effects. In 2008, Elashoff et al. [124] used the cause-specific hazards approach to model the competing risks. They have introduced frailty in the cause-specific hazards model of the competing risks and linked the frailty to the random effects from linear mixed-effects model for the longitudinal process through a joint distribution. That is, they have used proportional cause-specific hazards frailty submodel for the competing risks and a linear mixed-effects submodel for longitudinal process. The regression coefficients of the frailty term describe the associations among the competing risks, and the association between frailties and the random effects from the linear mixed-effects model describe the association between the competing risks and the longitudinal process. They have used the EM algorithm for parameter estimation. Williamson et al. [125] essentially followed a same approach as above, however, with one exception that the frailty term in the cause-specific hazards models are proportional to the random effects in the linear mixed effects model of the longitudinal process. Hu et al. [126] used a similar model as Elashoff et al. [124], and used a Bayesian approach for parameter estimation. Li et al. [127] proposed a robust joint modeling where they have used a model similar to Elashoff et al. [124] but assumed a $t-$distribution for the measurement error of the longitudinal process. In 2010, Li et al. [128] extended the longitudinal process to accommodate ordinal data.

### 4.3 Joint Modeling of Multivariate Longitudinal and Competing Risks Data

In this section, we describe the framework of our model. As described in the previous sections, for our motivating data, we fit the longitudinal measurements of liver function
(bilirubin) and renal function (GFR) using non-linear mixed effects models. Further, these multivariate responses are linked by random effects as described in the previous section. We use cause-specific hazards approach (Prentice et al. [36]) to model the competing risks time-to-event processes. Finally, following Williamson et al. [125], we link the multivariate longitudinal responses and competing risks time-to-event data through random effects.

We now describe a shared parameter joint model for multivariate longitudinal continuous responses and competing risks data. In the following, we describe the two submodels and the joining mechanism.

**Submodel 1: multivariate longitudinal process**

As described in section 3.5, let there be $K$ multivariate correlated longitudinal continuous responses $Y^k_i$ ($k = 1, \ldots, K$). Let $T^k_i = \{t^k_{i1}, t^k_{i2}, \ldots, t^k_{in^k_i}\}$ be the set of time points at with $n^k_i$ number longitudinal measurements for $k^{th}$ ($k = 1, \ldots, K$) response were observed for patient $i$ ($i = 1, \ldots, n$). Hence, for two longitudinal responses $j \neq k \in \{1, \ldots, K\}$, one can have three different scenarios: i). $T^k_i \equiv T^j_i$, all the time points at which longitudinal responses $j$ and $k$ are observed the same; ii). $T^j_i \cap T^k_i \neq \emptyset$, some (or all) of the time points at which longitudinal responses $j$ and $k$ are observed the same; iii). $T^j_i \cap T^k_i = \emptyset$, the time points at which the longitudinal response $j$ observed is completely different from the time points at which the longitudinal response $k$ observed. In general, scenario iii) is a general case. We define the submodel for multivariate longitudinal process that can handle all the scenarios.

The system of non-linear mixed effects model for the $K$ multivariate longitudinal process is given as follows:

$$
(4.2) \quad Y^k_i = X^k_{i0} \beta^k_o + \log \left( \sum_{l=1}^{L_k} \mu_l(X^k_{il}; \beta^k_l)T_l(t_i, \Theta^k_l) \right) + b^k_i + e^k_i, \quad k = 1, \ldots, K,
$$
where $\tilde{\log}(\cdot)$ is a vector as the same dimension as $Y^k_i$ ($k = 1, \ldots, K$). $X^k_{io}$, and $X^k_{il}$ ($l = 1, \ldots, L_k$) are overall and time phase-specific covariates related to longitudinal response $k$ ($k = 1, \ldots, K$), respectively. $X^k_{il}$ ($k = 1, \ldots, K$) may or may not have common elements. These multivariate responses are linked together by a joint distribution of the random effects $b_i = [b^1_i, \ldots, b^K_i]^\top$ with $b_i \sim N(0, G)$ where $G$ is the $K \times K$ variance-covariance matrix of the random effects. Off-diagonal elements of $G$ provide the measure of association among the $K$ longitudinal responses. We assume, random effects $b_i$ and measurement errors $\epsilon_i$ from the same subject $i$ are not correlated. That is, $\epsilon_i \perp b_i$, $\forall i \in \{1, \ldots, n\}$. Further, we assume the measurement errors from different subjects not correlated. That is, the distribution of the measurement errors $\epsilon_i^k|b_i \sim N(0, \sigma_{kk}I_{n_k})$ ($k = 1, \ldots, K$). In addition, we assume, conditional independence, a pure random effect model. That is, the measurement errors for different responses from the same subject are uncorrelated. That is, we have $\epsilon_i^j \perp \epsilon_i^k$, $\forall j \neq k \in \{1, \ldots, K\}$. Detailed description of the multivariate model is given in section 3.5.

**Submodel 2: competing risks time-to-event process**

As described in section 2.3, let us suppose that there are $D$ mutually exclusive competing events, denoted as $\{1, 2, \ldots, D\}$. We observe the following complete data $R = (T, C)$ with $C = 0$ if the censoring occurred and $C = j$ if event $j$ ($\in \{1, 2, \ldots, D\}$) is observed. When $C = j$, $T$ is the time at which event $j$ is observed and when $C = 0$, $T$ is the time when censoring occurred. We assume that the censoring mechanism is non-informative. That is, it is independent of longitudinal process and independent of covariates.

The system of cause-specific hazards models for the $D$ competing risks is given by,

\begin{equation}
\lambda_j(t; Z_i^j(t), W_i) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t, C = j| T \geq t, Z_i^j(t), W_i^j)}{\delta t} \right\}, \quad j = 1, \ldots, D.
\end{equation}

The function $\lambda_j(t; Z_i^j(t), W_i)$ gives the instantaneous event rate for event $j$ at time $t$ in
the presence of the other competing events; \( Z_j(t) \) is the covariate related to competing risks \( j \). The covariate matrices \( Z_j(t) \) (\( j = 1, \ldots, D \)) may or may not have common elements. Competing risks can be linked through common elements of \( Z_j \)'s (if any) and the random effect \( W_i = \left[ W_1^i, \ldots, W_D^i \right]^T \) (often called frailty term). Note that, it is often assumed the frailty terms are proportional. That is, \( W_j^i = a_j W_i, \ \forall j \), with \( a_1 = 1 \) for the purpose of identifiability.

**The joint model**

Joining longitudinal and time-to-event processes can be done in two ways: either through the joint distribution of two separate latent variables (random effects), one each from longitudinal and time-to-event processes (Henderson et al. [113]) or by considering the latent variable of the time-to-event as a function of latent variables of longitudinal response - it is often assumed that the random effect (latent variable) in the time-to-event process is taken as a linear function of random effects in the longitudinal process (Wulfsohn and Tsiatis [100]). Note that, Schluchter [109] linked the two processes through a trivariate Gaussian, where patient-specific random intercepts, random slopes from the longitudinal process and transformed survival times from the time-to-event process are linked through a trivariate normal distribution. For the joint modeling of competing risk process and single longitudinal process, while Elashoff et al. [124] linked the two processes through a joint distribution of two separate sets of latent variables, one set for each process, Williamson et al. [125] linked the two processes by assuming the latent variable in the competing risks process to be a linear function of latent variables of the longitudinal process.

As our major objective is to quantify the effect (or association) of the longitudinal process on the competing risks process, we follow the latter approach to link the multivariate longitudinal process and competing risks time-to-event process, where we link the two submodels by assuming \( W_j^i \) is a linear function of \( \left[ b_1^i, \ldots, b^K \right] \). That is, we assume \( W_j^i = \sum_{k=1}^{K} \gamma_j^k b_k^i \). In other words, the submodels are joined using a common
random-effects. That is, our joint model is a shared parameter model. Note that, \( W_i^j \) is a linear combination of normal random variables, hence, \( W_i^j \) is normal random variate. Thus, \( W_i = [W_i^1, \ldots, W_i^D] \top \) is a \( D \)-variate normal random variable. In general terminology, the submodel 1 referred to as \textit{measurement model} and the submodel 2 referred to as the \textit{intensity model}.

We now describe three different modeling approaches for the cause-specific hazards model (4.3).

**Description of Submodel 2:**

For submodel 2 (4.3), we use the following proportional and non-proportional hazards models,

\[
\begin{align*}
\text{Submodel 2: Case 1 - Weibull cause-specific hazards regression} \\
\lambda_j(t; Z_i^j(t), b_i) &= \kappa_j \eta_j t^{\kappa_j - 1} \exp \left\{ Z_i^j \phi^j + \sum_{k=1}^K \gamma_{jk}^k b_i^k \right\}, \quad j = 1, \ldots, D.
\end{align*}
\]

The regression coefficient \( \gamma_{jk}^k \) describes the association between \( k^{th} \) longitudinal response and \( j^{th} \) competing risk. Further, suppose both \( \gamma_{jk}^k \) and \( \gamma_{jk'}^k \) are positive, we can conclude that for the given longitudinal response \( k \), the competing risks \( j \) and \( j' \) are positively associated. On the other hand, suppose \( \gamma_{jk}^k \) is positive and \( \gamma_{jk'}^k \) is negative or vice versa, the competing risks \( j \) and \( j' \) are said to be negatively associated.

\[
\text{Submodel 2: Case 2 - Piecewise exponential baseline cause-specific hazards model}
\]
\( \lambda_j(t; \mathbf{Z}^j(t), \mathbf{b}_i) = \lambda_{j,0}(t) \exp \left\{ \mathbf{Z}^j_i \phi^j + \sum_{k=1}^{K} \gamma^k_j t^k_i \right\}, \ j = 1, \ldots, D. \)

where \( \lambda_{j,0}(t) \) is the baseline hazard assumed to be piecewise exponential. Let us suppose the time scale can be partitioned into \( P \) disjoint intervals, say, \( (t_0, t_1], (t_1, t_2], \ldots, (t_{P-1}, t_P], \) where \( t_0 \) could be 0 and \( t_P \) could be \( \infty \). The piecewise exponential hazards model is obtained by assuming the baseline hazard \( \lambda_{j,0}(t) \) constant within each interval but can vary between intervals. That is,

\[
(4.6) \quad \lambda_{j,0}(t) = \sum_{p=1}^{P} \lambda_{j,0p} I_{(t_{p-1}, t_p]},
\]

where \( I_{(t_{p-1}, t_p]} = 1 \) if \( t \in (t_{p-1}, t_p] \), and 0 otherwise. Note that, the partitioning process can be different for different competing risks. See for example, Vonesh et al. [115] for a detail application of this approach in a joint modeling set up. As they described, one advantage of this approach is that because the time scale is partitioned, it can accommodate time-varying covariates. Hence, this model can be either proportional or non-proportional hazard. The interpretation of \( \gamma^k_j \) is similar to Case 1 above.

**Submodel 2: Case 3 - Multi-phase cause-specific hazards model**
Under this case, we use the multi-phase hazards model proposed by Blackstone et al. [32]. This is a full parametric non-proportional hazard model which is based on decomposing the hazards (cumulative hazard) into multiple overlapping phases of risk which are additive. Each phase has its own set of covariates, which may or may not have common elements. Under this case, assuming \( C \) phases of hazards, we write their model as,
\[ \lambda_j(t; Z_{ij}^j(t), b_i) = \sum_{c=1}^{C} \exp \left\{ Z_{ij}^j \phi_c + \sum_{k=1}^{K} \gamma_{j,c}^k b_i^k \right\} g_c^j(t, \Theta_{c}^j), \quad j = 1, \ldots, D, \]

where \( g_c^j(t, \Theta_{c}^j) \) is a non-linear function of time with \( \Theta_{c}^j \) a vector of shaping parameters for the time function for phase \( j \) as described in Section 3.4.1. It can be clearly seen that the effect of \( Z_{ij}^j \) changes with time function \( g_c^j(t, \Theta_{c}^j) \). Further, \( Z_{ij}^j \) is not influenced by the time phase \( g_c^{j'}(t, \Theta_{c}^{j'}) \) for \( c \neq c' \). Hence, this is a nonproportional hazards model. Note that, when \( k=1 \), this model reduces to a proportional hazards model. Further note that, Weibull hazards described in case 1 is a special case of this multi-phase hazard model. Details of the time functions and estimation approach are given in Blackstone el al. [32] and in section 3.4.1. In general, the additive phases of the hazards decomposes the risks of death into, for example, early risks of deaths, and late late risks of deaths. Hence, the regression coefficient \( \gamma_{j,c}^k \) could describe, for example, the association between the \( k^{th} \) longitudinal response and, early risk of the \( j^{th} \) competing risk.

### 4.3.1 Estimation

We use maximum likelihood estimation to estimate the parameters in our shared parameter joint model. Let \( Y_i = [Y_i^1, \ldots, Y_i^K]^\top \) be the observed multivariate longitudinal data with dimension \( n_{ii} = \sum_{k=1}^{K} n_i^k \) and \( R_i \) is the observed competing risks data for the \( i^{th} \) subject. Then under the shared parameter model set up, the \( i^{th} \) subject’s contribution to the likelihood is given by,

\[ f(Y_i, R_i) = \int_{b_i} f_{Y|b_i}(Y_i) f_{R|b_i}(R_i) f_b(b_i) db_i, \]
where \( f_{Y|b_i}(Y_i) \) is the conditional multivariate function for the multivariate longitudinal process, \( f_{R|b_i}(R_i) \) is the conditional survivor function for the whole competing risks process and \( f_b(b_i) \) is the distribution of the random effects. That is, the joint distribution is conditionally independent given the latent process. Let \( \Psi = (\beta, \Theta, \Sigma) \) be the vector of parameters from the multivariate longitudinal process where \( \Sigma \) is a diagonal matrix with \( \sigma_{kk} \) as the \( k^{th} \) element, \( \Omega \) be the vector that includes \((\phi, \gamma)\) and other shaping parameters according the cause-specific hazards from the competing-risks process. We further know that, \( G \) is the variance-covariance of the random effects \( b_i \).

Suppose, there are \( n_i' \) longitudinal measurements for each of the \( K \) longitudinal measurements measured at the same time for subject \( i \). The likelihood of \((\Psi, \Omega, G)\), conditional on the observed data \((Y_i, R_i)\) for the \( i^{th} \) subject is given by,

\[
L_i(\Psi, \Omega, G; Y_i, R_i) = f(Y_i, R_i|\Psi, \Omega, G) \\
= \int_{b_i} f(Y_i, R_i|b_i, \Psi, \Omega, G) f(R_i|b_i, \Psi, \Omega, G) f(b_i|\Psi, \Omega, G) db_i \\
= \int_{b_i} f(Y_i|b_i, \Psi) f(R_i|b_i, \Omega) f(b_i|G) db_i \\
= \int_{b_i} \left[ \prod_{j=1}^{n_i'} \left\{ \frac{1}{(2\pi)^{K/2}|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (Y_{ij} - M(b_i))^\top \Sigma^{-1} (Y_{ij} - M(b_i))/2 \right\} \right\} \right] \times \exp \left[ -\int_0^{T_i} \sum_{d=1}^D \lambda_d(t; Z_i^j(t), b_i, \Omega) dt \right] \\
\times \frac{1}{(2\pi)^{K/2}|G|^{1/2}} \exp\{ -b_i^\top G^{-1} b_i/2 \} db_i, \\
= \int_{b_i} \left[ \prod_{k=1}^K \prod_{j=1}^{n_i^k} \left\{ \frac{1}{(2\pi)^{1/2}\sigma_{kk}^{1/2}} \exp\left\{ -\frac{1}{2\sigma_{kk}} (Y_{ij}^k - M^k(b_i))^2 \right\} \right\} \right] \times \exp \left[ -\int_0^{T_i} \sum_{d=1}^D \lambda_d(t; Z_i^j(t), b_i, \Omega) dt \right] \\
\times \frac{1}{(2\pi)^{K/2}|G|^{1/2}} \exp\{ -b_i^\top G^{-1} b_i/2 \} db_i,
\]
where the third equation on the right is obtained using the conditional independence of \( R_i \) and \( Y_i \) given \( b_i \), \( M(b_i) \) is the conditional mean response of the \( K \)-variate longitudinal process for the \( i^{th} \) subject, and \( I(C_i = d) = 1 \) if the \( d^{th} \) competing risk was observed, and 0 otherwise. Note that, because of the conditional independence assumption in the multivariate longitudinal process, the likelihood part of the multivariate longitudinal sub model process is further simplified into the product of \( K \) normal likelihoods in the fourth equation.

The complete likelihood is then given by,

\[
L(\Psi, \Omega, G; Y, R) = \prod_{j=1}^{n} L_j(\Psi, \Omega, G; Y_i, R_i).
\]

Since the random-effect enters the shared parameter model non-linearly, the maximum likelihood estimation involves an integral with no closed form. We use the Gaussian quadrature technique, as described in section 3.5.1, to approximate the integral. We then use quasi-Newton optimization method to estimate the parameters. This estimation approach can be conveniently adopted using PROC NLMIXED. See for example, Pinheiro and Bates [67] for details of the numerical integration.

### 4.4 Data Analysis

#### The Data

We now briefly describe the data. A detailed description of the data is given in Section 1.2. The motivating data are from Cleveland Clinic’s experience in using MCS as a bridge to heart transplant. Four hundred and thirty-nine patients who were transplant candidates (listed for transplant) and were bridged with MCS from December 1991 to July 2009 at the Cleveland Clinic are considered in this data analysis. Time zero for all the analyses is the time of initiation of MCS. Patients who at the outset were not
transplant candidates or were bridged to recovery were excluded. For simplicity, we have considered only the following 5 baseline variables for multivariable analyses: Age, pre-op Bilirubin, pre-op GFR, Gender (Male vs Female) and MCS device type: left ventricular assists device vs other device (LVAD vs. Other). Five patient who did not have any lab records on both bilirubin and GFR were excluded from this joint model analysis.

**Multivariate Longitudinal Data**

A total of 9620 lab records on post-MCS bilirubin and GFR are available for 434 patients. Both bilirubin and GFR are available in all 9620 records. Thus, 5 patients were excluded from the analysis. Among the 434 patients, 43 (10%) patients had only one record, 38 (9%) had only two records and the remaining had at least 3 records each. The mean follow-up time was 77 days (median=34 days, \( P_{25}=14 \) days, \( P_{75}=87 \) days).

**Competing risks Data**

We have 434 patients with 116 observed deaths on MCS and 288 had transplants. Hence as of 02/12/2010 (closing date - date of data extraction) there were still 30 patients alive, on MCS waiting for heart transplant. The Median follow-up time is 2.5 months. Among the survivors waiting for transplant, 25% were followed more than 1.4 years and 10% were followed 1.8 years or longer.

### 4.4.1 Application of submodel 2 for competing risks data

In section 3.4.3, we have demonstrated the appropriateness of a non-linear mixed effects model, as described in submodel 1 (4.2) to analyze the longitudinal responses described above. Hence, before we present the joint model analysis, we now demonstrate the appropriateness of the 3 cases of baseline hazards described in the submodel 2 (4.3) to analyze the two competing risks time-to-events described above.

**Submodel 2: Case 1 - Weibull baseline cause-specific hazards**

In this case, we have used the hazard of a Weibull time distribution with shape parameter
κ and scale parameter η as the baseline hazards as follows:

\[ \lambda_{0j}(t) = \kappa_j \eta_j t^{\kappa_j - 1}, \quad j = 1, 2. \]

The maximum likelihood estimates for the parameters of the two baseline cause-specific hazards are given in Table (4.1). The shapes of the baseline cause-specific hazards and parametric and non-parametric cumulative incidence functions are given in Figure 4.1. Note that the parametric cumulative incidence functions are estimated using estimated baseline cause-specific hazards and the equation 2.8. The estimated AICC is 478.2.

<table>
<thead>
<tr>
<th>Competing Risks</th>
<th>Parameter</th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death on MCS</td>
<td>κ</td>
<td>0.72 (0.054)</td>
</tr>
<tr>
<td></td>
<td>η</td>
<td>0.64 (0.066)</td>
</tr>
<tr>
<td>Transplant</td>
<td>κ</td>
<td>0.99 (0.042)</td>
</tr>
<tr>
<td></td>
<td>η</td>
<td>1.9 (0.12)</td>
</tr>
</tbody>
</table>

Table 4.1: Estimates of shape and scale parameters of Weibull baseline cause-specific hazard functions of the competing risks.

Cause-specific hazards of the competing risks yielded an early decreasing risk of
death on MCS and an almost constant risk of transplant. It can be seen from the plot of cumulative incidence functions that there appear to be slight deviations in the non-parametric and parametric estimates for the early events. One reason for this lack-of fit may be that the monotone Weibull hazard functions are not flexible enough to fit the different rate of risks at different time points in this data. However, overall, the fit appears to be adequate.

**Submodel 2: Case 2 - Piecewise exponential baseline cause-specific hazards model**

<table>
<thead>
<tr>
<th>Competing Risks</th>
<th>Time intervals (in months)</th>
<th>Parameter</th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death on MCS</td>
<td>(0, 0.125]</td>
<td>$\beta_{01}$</td>
<td>0.27 (0.41)</td>
</tr>
<tr>
<td></td>
<td>(0.125, 0.25]</td>
<td>$\beta_{02}$</td>
<td>0.32 (0.54)</td>
</tr>
<tr>
<td></td>
<td>(0.25, 0.5]</td>
<td>$\beta_{03}$</td>
<td>0.48 (0.47)</td>
</tr>
<tr>
<td></td>
<td>(0.5, 1]</td>
<td>$\beta_{04}$</td>
<td>0.029 (0.46)</td>
</tr>
<tr>
<td></td>
<td>(1, 3]</td>
<td>$\beta_{05}$</td>
<td>-0.58 (0.44)</td>
</tr>
<tr>
<td></td>
<td>&gt; 3</td>
<td>$\beta_{06}$</td>
<td>-1.1 (0.44)</td>
</tr>
<tr>
<td>Transplant</td>
<td>(0, 1]</td>
<td>$\beta_{01}$</td>
<td>0.24 (0.15)</td>
</tr>
<tr>
<td></td>
<td>(1, 2]</td>
<td>$\beta_{02}$</td>
<td>0.56 (0.20)</td>
</tr>
<tr>
<td></td>
<td>(2, 4]</td>
<td>$\beta_{03}$</td>
<td>0.89 (0.18)</td>
</tr>
<tr>
<td></td>
<td>(4,6]</td>
<td>$\beta_{04}$</td>
<td>0.70 (0.21)</td>
</tr>
<tr>
<td></td>
<td>(6,9]</td>
<td>$\beta_{05}$</td>
<td>-0.20 (0.27)</td>
</tr>
<tr>
<td></td>
<td>&gt; 9</td>
<td>$\beta_{06}$</td>
<td>-0.41 (0.27)</td>
</tr>
</tbody>
</table>

Table 4.2: Estimates of scale parameters of Piecewise baseline cause-specific hazard functions of the competing risks.

In this case, we use the piecewise exponential baseline hazards model obtained by assuming the baseline hazard $\lambda_{j,0}(t)$ is constant within each interval but can vary between intervals. Further that, the set of intervals can be different for different competing risks. That is,

$$\lambda_{j,0}(t) = \sum_{p=1}^{P} \lambda_{j,0p} I_{(t_{p-1}, t_p]}, \ j = 1, 2,$$

where $\lambda_{j,0p} = \exp(\beta_{0p})$. Based on the analysis results from the case 1 above and from
the Chapter 2, to improve the model fit of the higher early death on MCS events rate and to improve model fit of the the intermediate transplant events rate, we have created 2 different sets of 6 time intervals for each of the two competing events. The maximum likelihood estimates of the two baseline piecewise cause-specific hazards are given in Table (4.2).

The shapes of the baseline cause-specific hazards and parametric and non-parametric cumulative incidence functions are given in Figure 4.2. The estimated AICC is 421.9.

![Figure 4.2: Plot on the left depicts the shapes of the baseline cause-specific piecewise hazards of the competing risks, solid lines in the plot on the right depict cumulative incidence functions of the competing risks, and the symbols are the Aalen-Johansen non-parametric estimates of the cumulative incidence functions.](image)

There appear to be an early peaking risk followed by a constants risk for death on MCS and an intermediate peaking risk for transplant. It can be seen from the plot of cumulative incidence functions that there appear to be slight deviations in the non-parametric and parametric estimates for the late events. One reason for this lack-of fit may be the choice of the time intervals for the piecewise constant hazards. However, overall, the fit appears to be adequate.

**Submodel 2: Case 3 - Multi-phase baseline cause-specific hazards model**
### Table 4.3: Estimates of shape and scale parameters of baseline cause-specific multi-phase hazard functions of the competing risks.

<table>
<thead>
<tr>
<th>Competing Risks</th>
<th>Phases</th>
<th>Parameter</th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death on MCS</td>
<td>Early peaking</td>
<td>$\beta_0$</td>
<td>-2.3 (0.25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu$</td>
<td>-0.62 (0.13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$m$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{1/2}$</td>
<td>0.043</td>
</tr>
<tr>
<td>Constant</td>
<td>$\beta_0$</td>
<td>-0.67 (0.14)</td>
<td></td>
</tr>
<tr>
<td>Transplant</td>
<td>Intermediate peaking</td>
<td>$\beta_0$</td>
<td>0.68 (0.14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu$</td>
<td>-0.64 (0.038)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$m$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{1/2}$</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Under this case, we use the multi-phase hazards model proposed by Blackstone et al. [32]. Assuming $C_j$ phases of hazards for competing risks $j$, we can write the cause-specific multi-phase hazard models as,

$$
\lambda_{j0}(t) = \sum_{c=1}^{C_j} \exp(\beta_{j0c})g_{jc}^j(t, \Theta_{jc}^j), \ j = 1, 2,
$$

where $g_{jc}^j(t, \Theta_{jc}^j)$ is a non-linear function of time with $\Theta_{jc}^j$ a vector of shaping parameters for the time function for phase $j$ as described in Section 3.4.1. Data analysis using the multiphase hazards model yielded an early peaking and a late constant phase for death on MCS event, and an intermediate peaking phase of the transplant events. The maximum likelihood estimates of the scale and shaping parameters of the two baseline cause-specific hazards are given in Table (4.3).

The shapes of the baseline cause-specific hazards and parametric and non-parametric cumulative incidence functions are given in Figure 4.4. The estimated AICC is 397.8.

It can be seen from the plot of cumulative incidence functions that there is a very good fit achieved using the flexible multi-phase baseline hazards model. Further note that, the AICC is the smallest for this model than the other two baseline hazards approaches.
4.4.2 Data analysis using joint modeling approach

We now present the results of the joint modeling analyses in three sections, each with different baseline hazards for submodel 2. Further, we first present the results without the baseline covariates in the model. In each analysis, we focus on the following three types of models:

1. **Model 1**: Completely Independent: Here, we assume submodel 1 and submodel 2 are completely independent. That is, \( f(Y, R) = f_Y(Y) \times f_R(R) \), assumed to be Missing At Random (MAR - completely ignorable missing). Fitting the two submodels separately;

2. **Model 2**: Semi-Joint model: Here, we assume the submodel 1, the model for multivariate longitudinal process is completely independent. That is, \( f(Y_1, Y_2) = f_{Y_1}(Y_1) \times f_{Y_2}(Y_2) \). That is, we assume \( Cov(b^1, b^2) = 0 \). In other words, likelihood of the shared parameter model is written as,
\begin{align*}
L_i(\Psi, \Omega, G; Y_i, R_i) &= f(Y_i, R_i \mid \Psi, \Omega, G) \\
&= \int_{b_i} f(Y_i \mid R_i, b_i, \Psi, \Omega, G) f(R_i \mid b_i, \Psi, \Omega, G) f(b_i \mid \Psi, \Omega, G) db_i \\
&= \int_{b_i} f(Y_i \mid b_i, \Psi) f(R_i \mid b_i, \Omega) f(b_i \mid G) db_i \\
&= \int_{b_i} f(Y_i^1 \mid b_i^1, \Psi) f(Y_i^2 \mid b_i^2, \Psi) f(R_i \mid b_i, \Omega) f(b_i^1 \mid \sigma_{11}) f(b_i^2 \mid \sigma_{22}) db_i 
\end{align*}

(4.9)

The main difference between this likelihood formulation and the formulation in section 4.3.1 is that the joint distribution \( f(b_i \mid G) \) is the product of two marginals. Further, note that, the semi-joint model is not equivalent to fitting two separate joint models. That is, \( f(Y_1, Y_2, R) = f_{Y_1 \mid R}(Y_1 \mid R) \times f_{Y_2 \mid R}(Y_2 \mid R) \times f_R(R) \neq f(Y_1, R) \times f(Y_2, R) \).

3. **Model 3:** Full Joint Model: Here, submodel 1 and submodel 2 are modeled jointly.

**Joint Modeling using Weibull Baseline cause-specific hazards**

Maximum likelihood estimates based on the three different models are given in Table 4.4. It can be seen from AICC estimates that the full joint model (Model 3) is better than separate joint models (Model 2) and the ignorable missing mechanism models (Model 1). There is no noticeable change in the shaping and scaling parameters of temporal trends of the longitudinal responses. It is worth to noting that, the estimates of the temporal trends of the longitudinal responses from the MCAR model (Model 1) are very similar to those of the full joint model.

**Association parameter:** The association parameter \( \gamma_{kj} \) depicts the association between the \( k^{th} \) longitudinal response and the \( j^{th} \) competing risk. It is observed from the association parameters \( \gamma_{kj} \), that bilirubin and GFR are are significantly associated with competing risk death before transplant in both the semi-joint and full joint models.
Table 4.4: ML Estimates under completely ignorable missing (Model 1), semi-joint (Model 2) and non-ignorable (Model 3) mechanisms when using Weibull baseline cause-specific hazards.

\[ \gamma_{k,j}^a \] - the association between the \( k^{th} \) longitudinal response and the \( j^{th} \) competing risk
\[ b \] - marginal significant \( 0.05 < p < 0.1 \)
\[ c \] - \( p > 0.2 \)

with higher bilirubin and lower GFR are associated with higher risk of death on MCS. It is noted that in the semi-joint model, lower bilirubin and lower GFR are associated (marginally) with higher risk of transplant. However, this association becomes less significant in the full joint model.

**Joint Modeling using Piecewise Exponential Baseline cause-specific hazards**

Maximum likelihood estimates based on three different models are given in Table 4.5. In general, the results based on the piecewise exponential baseline cause-specific hazards as submodel 2 seems to be similar to that of submodel 2 based on Weibull baseline cause-specific hazards case.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ignorable Missing -Model 1-</th>
<th>Semi-Joint -Model 2-</th>
<th>Full Joint -Model 3-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal Process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Decreasing Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-1.1 (0.062)</td>
<td>-1.1 (0.054)</td>
<td>-1.1 (0.061)</td>
</tr>
<tr>
<td>( \nu )</td>
<td>-1 (0.024)</td>
<td>-1 (0.022)</td>
<td>-1 (0.024)</td>
</tr>
<tr>
<td>( m )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>0.084 (0.0077)</td>
<td>0.080 (0.0064)</td>
<td>0.083 (0.0077)</td>
</tr>
<tr>
<td>Late Increasing Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu )</td>
<td>0.93 (0.035)</td>
<td>0.95 (0.032)</td>
<td>0.93 (0.034)</td>
</tr>
<tr>
<td>( m )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>1.2 (0.026)</td>
<td>1.2 (0.025)</td>
<td>1.2 (0.026)</td>
</tr>
<tr>
<td>Log(Bilirubin)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Decreasing Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-0.22 (0.035)</td>
<td>-0.32 (0.041)</td>
<td>-0.22 (0.035)</td>
</tr>
<tr>
<td>( \nu )</td>
<td>-0.77 (0.0057)</td>
<td>-0.78 (0.0062)</td>
<td>-0.78 (0.0057)</td>
</tr>
<tr>
<td>( m )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>0.31 (0.012)</td>
<td>0.32 (0.014)</td>
<td>0.31 (0.012)</td>
</tr>
<tr>
<td>Late Increasing Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu )</td>
<td>0.64 (0.022)</td>
<td>0.69 (0.028)</td>
<td>0.64 (0.022)</td>
</tr>
<tr>
<td>( m )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>1.7 (0.054)</td>
<td>1.8 (0.071)</td>
<td>1.7 (0.054)</td>
</tr>
<tr>
<td>GFR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Peaking Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>-0.22 (0.035)</td>
<td>-0.32 (0.041)</td>
<td>-0.22 (0.035)</td>
</tr>
<tr>
<td>( \nu )</td>
<td>-0.77 (0.0057)</td>
<td>-0.78 (0.0062)</td>
<td>-0.78 (0.0057)</td>
</tr>
<tr>
<td>( m )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>0.31 (0.012)</td>
<td>0.32 (0.014)</td>
<td>0.31 (0.012)</td>
</tr>
<tr>
<td>Late Increasing Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu )</td>
<td>0.64 (0.022)</td>
<td>0.69 (0.028)</td>
<td>0.64 (0.022)</td>
</tr>
<tr>
<td>( m )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>1.7 (0.054)</td>
<td>1.8 (0.071)</td>
<td>1.7 (0.054)</td>
</tr>
<tr>
<td>V ar(( \epsilon ))</td>
<td>( \Sigma )</td>
<td>( \Sigma )</td>
<td>( \Sigma )</td>
</tr>
<tr>
<td>( \Sigma )</td>
<td>0.35 -0.21</td>
<td>0.38 -0.21</td>
<td>0.35 -0.21</td>
</tr>
<tr>
<td>V ar(( b ))</td>
<td>( G )</td>
<td>( G )</td>
<td>( G )</td>
</tr>
<tr>
<td>( G )</td>
<td>0.40 -0.41</td>
<td>0.40 -0.44</td>
<td>0.40 -0.41</td>
</tr>
<tr>
<td>Competing Risks Process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[0, 0.125] - ( \beta_{01} )</td>
<td>0.081 (0.39)</td>
<td>-0.00072 (0.39)</td>
<td>-0.069 (0.39)</td>
</tr>
<tr>
<td>[0.125, 0.25] - ( \beta_{02} )</td>
<td>0.22 (0.51)</td>
<td>0.16 (0.52)</td>
<td>0.14 (0.52)</td>
</tr>
<tr>
<td>[0.25, 0.5] - ( \beta_{03} )</td>
<td>0.25 (0.45)</td>
<td>0.24 (0.45)</td>
<td>0.24 (0.45)</td>
</tr>
<tr>
<td>[0.5, 1] - ( \beta_{04} )</td>
<td>-0.081 (0.44)</td>
<td>-0.071 (0.44)</td>
<td>-0.066 (0.44)</td>
</tr>
<tr>
<td>[1, 3] - ( \beta_{05} )</td>
<td>-0.54 (0.42)</td>
<td>-0.51 (0.42)</td>
<td>-0.51 (0.42)</td>
</tr>
<tr>
<td>Death before Transplant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[0, 0.125] - ( \beta_{01} )</td>
<td>-0.79 (0.42)</td>
<td>-0.75 (0.43)</td>
<td>-0.72 (0.065)</td>
</tr>
<tr>
<td>[0.125, 0.25] - ( \beta_{02} )</td>
<td>1.01 (0.12)</td>
<td>1.1 (0.13)</td>
<td>1.1 (0.13)</td>
</tr>
<tr>
<td>[0.25, 0.5] - ( \beta_{03} )</td>
<td>-0.55 (0.18)</td>
<td>-0.47 (0.17)</td>
<td>-0.47 (0.17)</td>
</tr>
<tr>
<td>Transplant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[0, 1] - ( \beta_{01} )</td>
<td>0.22 (0.16)</td>
<td>0.16 (0.16)</td>
<td>0.18 (0.16)</td>
</tr>
<tr>
<td>[1, 2] - ( \beta_{02} )</td>
<td>0.58 (0.21)</td>
<td>0.58 (0.21)</td>
<td>0.59 (0.21)</td>
</tr>
<tr>
<td>[2, 4] - ( \beta_{03} )</td>
<td>0.91 (0.19)</td>
<td>0.92 (0.19)</td>
<td>0.91 (0.19)</td>
</tr>
<tr>
<td>[4, 6] - ( \beta_{04} )</td>
<td>0.72 (0.21)</td>
<td>0.72 (0.21)</td>
<td>0.72 (0.21)</td>
</tr>
<tr>
<td>[6, 9] - ( \beta_{05} )</td>
<td>-0.0346 (0.27)</td>
<td>-0.0389 (0.27)</td>
<td>-0.0389 (0.27)</td>
</tr>
<tr>
<td>[9, 10] - ( \beta_{06} )</td>
<td>-0.39 (0.27)</td>
<td>-0.38 (0.27)</td>
<td>-0.39 (0.27)</td>
</tr>
<tr>
<td>Association Parameter- ( \gamma_{1} )</td>
<td>-0.21 (0.10)</td>
<td>-0.21 (0.10)</td>
<td>0.044 (0.10)</td>
</tr>
<tr>
<td>Association Parameter- ( \gamma_{2} )</td>
<td>-0.12 (0.10)</td>
<td>-0.12 (0.10)</td>
<td>0.044 (0.10)</td>
</tr>
<tr>
<td>AICC</td>
<td>38528</td>
<td>38808</td>
<td>38528</td>
</tr>
</tbody>
</table>

Table 4.5: ML Estimates under completely ignorable missing (Model 1), semi-joint (Model 2) and non-ignorable (Model 3) mechanisms when using Piecewise Exponential baseline cause-specific hazards.

\( a_{k}^{b} \) - the association between the \( k^{th} \) longitudinal response and the \( j^{th} \) competing risk

\( b_{p} > 0.2 \)

**Association parameter:** The association parameter \( \gamma_{k}^{b} \) depicts the association between the \( k^{th} \) longitudinal response and the \( j^{th} \) competing risk. It is noted that, unlike in the Weibull baseline cause-specific hazards, in the semi-joint model only lower bilirubin is associated with higher risk of transplant. However, similar to the Weibull baseline cause-specific hazards case, this association becomes less significant in the full joint model.

**Joint Modeling using Baseline cause-specific multi-phase hazards**

Maximum likelihood estimates based on three different models based on cause-specific
multi-phase hazards are given in Table 4.6. In general, the estimates of shaping and scaling parameters of the temporal trends are similar across the three models and further are also similar to the two baseline hazards cases discussed above. Based on AICC estimates, the full joint model appears to be better than the separate or semi-joint models. Further, based on the estimates of AICC, models (joint or separate) using multi-phase hazards are better than the models based on Weibull or Piecewise hazards.

**Association parameter:** The association parameter $\gamma_{j,c}^{k}$ depicts the association between the $k^{th}$ longitudinal response and the $c^{th}$ time phase of the $j^{th}$ competing risk. Estimates of association parameters reveals that higher bilirubin is associated with both early risk of death and late risk of death, however, the magnitude of association is much higher with early risk of death than the late risk of death. This is true for both semi-joint and full joint models. On the other hand, lower GFR is not significantly associated with early risk of death, but is associated with late risk of death. Neither bilirubin nor GFR is significantly associated with risk of transplant.

**Summary:** Table 4.7 summarizes the estimates of the association parameters obtained under 3 different cases of cause-specific hazards models for the full joint models. In summary, higher post-op bilirubin and lower post-op GFR are associated with higher risk of death on MCS. There appears a marginal or no effect of post-op longitudinal process on the risk of transplant. Further, based on the multiphase hazards model, we can infer that the magnitude of the effect of post-op bilirubin on death before transplant is higher on the early deaths than on the late deaths.

### 4.4.3 Multivariable analyses

As described in section 1.2, we have collected pre-MCS implant variables such as, demography, hemodynamics (for example, pulmonary artery pressure, cardiac index), cardiac related variables, markers of non-cardiac comorbidity (for example, levels of creatinine, total bilirubin, BUN, hematocrit), and index MCS device type. However, for illustra-
tion purposes, we now consider the following 5 variables: Age, Gender, MCS device type (LVAD vs other), pre-device implant bilirubin and GFR, and perform a multivariable analysis for all three joint models (with three different baseline cause-specific hazards). In both sub models, we focused only on the baseline variables that are, at least, marginally associated with the outcomes. Further, for simplicity, we report only the estimate of the regression coefficients related to the baseline covariates and the estimates of the association parameters, not the shaping parameters of the multi-phase or cause-specific hazard models.

**Joint model with Weibull cause-specific hazards model**

The regression estimates are given in Table 4.8. In the longitudinal process submodel, older age appeared to be associated with both post-op liver dysfunction (higher post-op bilirubin) and renal dysfunction (lower GFR); device type LVAD (versus other) appeared to be associated with early lower values of bilirubin and early higher values of GFR, that is, a protective effect on both renal and liver function early after the device implant. Further, it appears, LVAD also has a protective effect on the late GFR values, and this positive effect appears to be larger than that on early GFR values. Higher bilirubin is associated with higher post-op bilirubin. However, the magnitude of the influence is larger on the early post-op values of bilirubin than the late post-op values. A similar association is seen between pre-MCS GFR values and post-op GFR values.

In the competing risks submodel process, lower pre-MCS GFR values and devices other than LVAD appear to be risk factors for death before transplant. Older age (marginally) and Male gender are associated with higher risk of transplant.

The association parameter $\gamma^k_j$ depicts the association between the $k^{th}$ longitudinal response and the $j^{th}$ competing risk. It is observed from the association parameters $\gamma^k_j$, that post-op bilirubin and post-op GFR are significantly associated with competing risk- death before transplant, where higher bilirubin and lower GFR are associated with higher risk of death on MCS. However, only lower bilirubin is significantly associated
with higher risk of transplant. Note that, after adjusting for baseline variables, the association parameter $\gamma_2^1$ became significant (See Tables 4.4 and 4.8). Further note that the estimates of the shaping parameters have not changed significantly when compared with the estimates unadjusted for baseline variables (see Table 4.4).

**Joint model with Piecewise exponential cause-specific hazards model**

Under the assumption of piecewise exponential baseline hazards model, the regression estimates and the estimates of the association parameters have the same interpretation and inference as under the Weibull baseline cause-specific hazards model (Table 4.9). Note that, under the case of piecewise exponential cause-specific hazards assumption, the effect of age on the risk of transplant became highly non-significant.

**Joint model with multi-phase cause-specific hazards model**

Under the assumption of a multi-phase cause-specific hazards model, the estimates of regression parameters in the longitudinal process submodel are essentially similar to that of the previous two cases.

In the competing risks submodel, older age (marginally) and device type other than LVAD appear to be risks factors for early risk of death before transplant, and lower baseline GFR is associated with higher risk of late deaths. Similar to previous two cases, male gender appeares to be associated with higher risk of transplant.

The association parameter $\gamma_{j,c}^k$ depicts the association between $k^{th}$ longitudinal response and $c^{th}$ time phase of the $j^{th}$ competing risk. Estimates of association parameters reveal, as in the unadjusted case (Table 4.6), that higher post-op bilirubin is associated with both higher early risk of death and higher late risk of death. However, unlike the unadjusted case (Table 4.6), the magnitude of association now appears somewhat similar in both phases of risks of death (Table 4.10). Post-op GFR is not associated with either phase of the risks of death before transplant or transplant. Unlike the unadjusted case (Table 4.6), post-op lower bilirubin is significantly associated with risk of transplant.
4.5 Simulation Study

We now carry-out a series of simulation studies to assess performance of our full joint model under different scenarios of data setups. We particularly, are interested in assessing the performance of estimates of the association parameters. It should be noted that in this simulation study, we are not focusing on model building which is out of the scope of this analysis, rather, given a joint- model (with given phases), we would like to assess how well the parameters are estimated. We investigate the behavior of the association parameters under different scenarios.

We assess the performance of the full joint model based on simulated data using the following summary measures: suppose there are $B$ simulated data sets and $\beta$ is the true value and $\hat{\beta}_i$ is the estimate from the $i^{th}$ simulated data set; Percent Mean Bias: $\% \text{Bias} = 100 \times (\bar{\beta} - \bar{\hat{\beta}})/\beta$, where $\bar{\hat{\beta}} = \frac{1}{B} \sum_{i=1}^{B} \hat{\beta}_i/B$; Average within standard error, $\text{AvgSE} = \frac{1}{B} \sum_{i=1}^{B} SE(\hat{\beta}_i)/B$; Empirical standard error, $\text{EmpSE} = \sqrt{\frac{1}{B-1} \sum_{i=1}^{B} (\hat{\beta}_i - \bar{\hat{\beta}})^2}$; 95% Coverage Probability (CP),

4.5.1 Submodel 1: Multivariate Longitudinal process

We generate a bivariate continuous longitudinal response at the following 15 time points over a 5-year time period: 1-day, 2-day, 3-day, 1-week, 2-week, 1-month, 3-month, 6-month, 1-year, 1.5-years, 2-years, 2.5-years, and 3 time points at year 3 to year 5.

Following the phases and shapes of the bivariate longitudinal process from our motivating data, we consider the following bivariate longitudinal process for our simulation study.

- **Longitudinal Response** $Y^1$ with two phases, an early decreasing phase, $T^1_1(\Theta^1_1, t)$ with $\Theta^1_1 = (m = 0, \nu = -1, t_{1/2} = 0.25)$ and a late increasing phase, $T^1_2(\Theta^1_2, t)$ with $\Theta^1_2 = (m = 0, \nu = 1, t_{1/2} = 5)$ which plateau after 2 years. Note that, we have considered $g(\Theta, t)$ with the limmiting case of **Case 3** as $T^1_1(\Theta^1_1, t)$ and $h(\Theta, t)$
Figure 4.4: True shapes of the phases for temporal trends of the bivariate longitudinal processes.

(equation 3.9) with the limiting case of Case 1 as the time function \( T_2^1(\Theta_1^1, t) \).

See section 3.4 for detail description of the time functions. See plot on the left in Figure 4.4 for the shapes of the phases for the set values.

- **Longitudinal Response** \( Y^2 \) with two phases, an early peaking phase, \( T_2^2(\Theta_2^2, t) \) with \( \Theta_2^2 = (m = 0, \nu = -0.75, t_{1/2} = 0.5) \) and a late peaking phase, \( T_2^2(\Theta_2^2, t) \) with parameters \( \Theta_2^2 = (m = 0, \nu = 0.75, t_{1/2} = 4) \) which plateau after 2 years. The time functions are similar to that of used for longitudinal Response \( Y^1 \) as described above. See section 3.4 for detail description of the time functions. See plot on the right in Figure 4.4 for the shapes of the phases for the set values.

- **Random effects**: Unless otherwise stated, we use the following assumption for the joint distribution of the random effects of the bivariate longitudinal process,

\[
\begin{pmatrix}
    b^1 \\
    b^2
\end{pmatrix} \sim N \left(
    \begin{pmatrix}
        0 \\
        0
    \end{pmatrix},
    \begin{pmatrix}
        0.4 & -0.3 \\
        -0.3 & 0.35
    \end{pmatrix}
\right).
\]

That is, we assume a bivariate normal distribution for the random effects.

- **Measurement Errors**: We use the following distributional assumption for the
conditional distribution of the measurement errors of the bivariate longitudinal process,

\[
\begin{pmatrix}
\epsilon^1 \\
\epsilon^2
\end{pmatrix}
\sim N
\begin{pmatrix}
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
0.36 & 0 \\
0 & 0.49
\end{pmatrix}
\end{pmatrix}.
\]

That is, we assume independent normal distributions for the measurement errors given the random effects.

### 4.5.2 Submodel 2: Competing Risks process

We generate two competing risks with maximum follow-up of 5 years.

- **Baseline cause-specific hazards**: We assume constant baseline cause-specific hazards. That is, with exponential time distributions with with hazard rates of 1 and 0.75 for the two competing risks, respectively.

- **Censoring Distribution**: We assume an exponential time distribution with a low rate of 0.25 as the censoring rate. We further assume censoring is non-informative. That is, the censoring mechanism is independent of both competing risks and longitudinal processes. Note that, the specified baseline constant cause-specific rates and censoring rate, on average, yield 50% of the subjects with competing risk 1 and 38% with competing risk 2 and 12% were censored.

- The multivariate longitudinal processes of each subject were censored based on the censoring or event time.
4.5.3 *Association parameters*

Based on our data analysis, we keep association parameters as in Table 4.11. That is, while longitudinal process $Y^1$ is positively associated with competing risk 1, and negatively associated with competing risk 2, longitudinal process $Y^2$ is negatively associated with competing risk 1, and positively associated with competing risk 2.

**Remark:** Since the computing time increases considerably with increasing sample size and increasing number of covariates in the model, to limit the computing time, we have generated 250 Monte Carlo simulations with sample size of 400 (number of subjects) and have performed the simulation study as follows: We first perform a simulation study with two baseline covariates (a continuous and a binary), along with association variables, in the full joint model. We then perform a simulation study without the two baseline covariates in the joint-model. Finally, all the subsequent different simulation scenarios are compared with the joint-model without the two baseline covariates. Note that, our full-joint model with 2 covariates has 29 parameters to be estimated, and without the two covariates, the number of parameters to be estimated is reduced to 21.

4.5.4 *Simulation results*

**Case 1: With Baseline Covariates in the model**

We include two baseline covariates, a binary covariate $x_1$ with $\text{Binary}(p = 0.6)$ and a continuous covariate $x_2$ from a standardized normal distribution ($\mathcal{N}(0,1)$) in both submodels along with shaping, scaling and association parameters.

Summary measures based on the simulated data with covariates in the joint model are given in Table 4.12. It appears the joint model yields unbiased estimates of all the parameters and the average standard errors appear to be close to the empirical standard errors. The estimated coverage probabilities also appear to be close to the nominal value of 0.95. The estimates of association parameters also appear to be close to the nominal
values, although it appears the bias percentage is slightly higher for competing risk 2. However, it is noted that the other two summary measures, standard error and coverage probability for these parameters produced acceptable results.

**Case 2: Without Baseline Covariates in the model**

We now simulate 250 samples without the two covariates, $x_1$ and $x_2$. The main purpose of this exercise is to reduce the computing time of the simulation study. We then use results from this simulation study as the comparison for different scenarios of the model and parameter assumptions. That is, for the rest of section of simulation results comparisons and discussions, we use the model without covariates as the comparison model. Further, for the rest of this section, for all the other scenarios, except in one notable scenario, we present, compare and discuss only the results of the association parameters.

It can be noted from Table 4.13 that all the summary measures of the joint-model parameters produced acceptable values.

**Case 3: Varying Sample Size**

In this simulation scenario, we compare the influence of varying sample size on the parameter estimation of our full-joint model.

It can be noted from Table 4.14, as the sample size reduces there appears to be some increase in the bias of the association parameters. This is especially apparent in the estimation of association parameters, $\gamma_1^1$ and $\gamma_2^1$. One can speculate that since the number of parameters to be estimated for our joint full is relatively high (there are 21 parameters), as the sample size decreases, the efficiency may decrease. That is, the asymptotic properties of the estimates are less likely to hold.

**Case 4: Varying Censoring Rate**

In this simulation scenario, we compare the influence of different censoring rates on our
full joint model parameter estimation. We have used an exponential time distribution with constant rate $\lambda$ as the censoring distribution for the competing risks. Note that, we have assumed that the censoring distribution is independent of competing risk process and the longitudinal process. We have considered two censoring rates, a lower rate $\lambda = 0.25$, and a higher rate $\lambda = 1.25$. Note that, with the lower rate, approximately, 50% had experienced competing risk 1 and 38% experienced competing risk 2, and 12% were censored. With the higher rate, 33% had experienced competing risk 1 and 25% experienced competing risk 2, and 42% were censored.

It can be clearly noted from Table 4.15 that the bias increases slightly with increasing censoring rate. However, it appears that in both scenarios, the summary measures of the estimates of the association parameter produced acceptable values. Note that estimates of all the other parameters of the full joint model proceduced acceptable results, hence, not shown in the Table 4.15.

Case 5: Full-Joint Model versus Semi-Joint Model versus Separate Joint Model with Varying association between $Y^1$ and $Y^2$

In this case, given a bivariate longitudinal process and a competing risks process, we compare different joint-modeling scenarios under different assumptions of association between the bivariate longitudinal process. Suppose $(Y^1, Y^2)$ is the bivariate longitudinal process and $R$ is the competing risks process, we define the three joint-models as follow,

- Full Joint Model: $f((Y^1, Y^2), R)$;

- Semi-Joint Model: Under this scenario we assume the random effects from the longitudinal processes, $b^1$ and $b^2$ have independent normal distributions. That is, we assume, $\text{Corr}(b^1, b^2) = 0$. That is, the full joint model can be reduced to $f((Y^1, Y^2), R) = f(Y^1|R) \times f(Y^2|R) \times f(R)$;

- Separate Joint Models: Under this scenario we fit two separate joint models, one for joint modeling of longitudinal process $Y^1$ and competing risk process $R$. That
is, a joint model for \( f(Y^1, R) \). And the other one is a separate joint modeling of longitudinal process \( Y^2 \) and competing risk process \( R \). That is a joint model for \( f(Y^2, R) \).

We use

\[
\begin{pmatrix}
    b^1 \\
    b^2
\end{pmatrix}
\sim
N
\left(\begin{pmatrix}
    0 \\
    0
\end{pmatrix},
\begin{pmatrix}
    0.4 & -0.3 \\
    -0.3 & 0.35
\end{pmatrix}\right),
\]

which yields a higher correlation, \( Corr(b^1, b^2) = -0.8 \) and

\[
\begin{pmatrix}
    b^1 \\
    b^2
\end{pmatrix}
\sim
N
\left(\begin{pmatrix}
    0 \\
    0
\end{pmatrix},
\begin{pmatrix}
    0.4 & -0.05 \\
    -0.05 & 0.35
\end{pmatrix}\right),
\]

which yields a lower correlation, \( Corr(b^1, b^2) = -0.13 \). Summary measures based on the simulation are given in Table 4.16.

It can be seen from the summary measures in Table 4.16 that the strength of association between the longitudinal outcomes \( Y^1 \) and \( Y^2 \) has a clear impact on the choice of joint models to assess the association between the multivariate longitudinal process and the competing risk process. When the association between \( Y^1 \) and \( Y^2 \) is high, the full joint model clearly out performed the other two models, whereas separate joint models produced highly biased association estimates with very low confidence interval coverage probabilites. However, when the association between the longitudinal outcomes \( Y^1 \) and \( Y^2 \) is low, as expected, all 3 joint models proceduced acceptable summary measures.

**Case 6: Non-Normal Frailty**

In our full joint model, the multivariate longitudinal process and the competing risk process are linked by assuming the frailty term \( W^j \) for the \( j^{th} \) competing risk is a linear function of random effects in the multivariate longitudinal process. That is, \( W^j = \sum_{k=1}^{K} \gamma^j_kb^k \), where \( b^k \) is the random effect from \( k^{th} \) longitudinal outcome and \( \gamma^j_k \)
depicts the association between $j^{th}$ competing risk and $k^{th}$ longitudinal outcome. In this simulation scenario, we investigate the estimation when the frailty term is not a linear function of random effects in the multivariate longitudinal process. We also investigate performance when the distribution of frailty is non-normal. We simulate data by taking the $W^1 = (b^1)^2 - (b^2)^2$, this is clearly a non-normal frailty for competing risk 1 and keep the other frailty the same as case 2, that is a normal. We then use this simulated data to implement our full joint model where the frailty is assumed to be a linear function of random effects in the multivariate longitudinal process.

It can be seen from the summary measures in Table 4.17 that assuming the frailty is a linear function random effects in the multivariate longitudinal process, when in fact the frailty is a non-linear function produce a highly biased estimates of the association parameters with poor coverage probability.

**Case 7: Non-Normal Random Effects**

In our full joint model, we have assumed the random effects are jointly distributed as a multivariate normal. We now perform a simulation study by simulating the random effects of the bivariate longitudinal process from a bivariate $t$-distribution and estimating the parameters using our full joint model. That is, we perform data analysis by assuming the random effects of the bivariate longitudinal process is a bivariate normal variate. We use a bivariate generalization of Student’s $t$-distribution proposed by Dunnett and Sobel [129] to generate bivariate random effects. This algorithm is conveniently implemented using the mnormt package (R software [46]). In their notation, if $t_p(d, \Sigma')$ is a $p$-variate multivariate $t$-distribution with degrees of freedom, $df = d$ and scale distribution, $\Sigma'$ then when $d > 2$, $G = \Sigma' \times d/(d-2)$ can be considered as the variance-covariance matrix. In our simulations, we have kept the scale matrix $\Sigma'$, adjusted for degrees of freedom, such that the variance covariance matrix of the bivariate random effects is equivalent to our original specification (case 2). That is,
\[
\text{Var}([b^1, b^2]^{\top}) = \begin{pmatrix}
0.4 & -0.3 \\
-0.3 & 0.35
\end{pmatrix}.
\]

We have generated random effects from the bivariate-\(t\) under two scenarios, i). smaller \textit{degrees of freedom}, \(df = 3\) which is equivalent to having a heavy-tailed bivariate distribution, which may not be considered as bivariate-normal; ii). a larger \textit{degrees of freedom}, \(df = 10\) which may be closer to a bivariate normal distribution. We then use the simulated data to implement our full joint model to estimate the parameters under the assumption of bivariate normal random effects. Note that, since the joint distribution of the random effect is a bivariate \(t\)-distribution, the distribution of the frailty in the competing risk process, which is a linear combination of random effects in the multivariate longitudinal process, is no longer a normal random variate.

The results of the summary measures are given in Table 4.18. Note that, in addition to the summary measures of association parameters, because the results were interesting, we have also presented the summary measures of the parameters of the variance-covariance of the bivariate random effects \(b^1\) and \(b^2\).

The results of the summary measures shows that the estimates of the association parameters using our full joint model are closer to the results of the original simulation case (case 2, column 1 in Table 4.18) and also closer to the nominal values. However, when the bivariate \(t\)-distribution has a heavy-tail, \(df = 3\), parameter estimation of the variance-covariance matrix of the bivariate random effects performed very poorly with higher bias and poor coverage probability.
### Table 4.6: ML Estimates under ignorable missing (Model 1), semi-joint (Model 2) and non-ignorable (Model 3) mechanisms when using baseline cause-specific multi-phase hazards.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ignorable Missing -Model 1-</th>
<th>Semi-Joint -Model 2-</th>
<th>Full Joint -Model 3-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Longitudinal Process</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Decreasing Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1.1 (0.062)</td>
<td>-1.1 (0.054)</td>
<td>-1.1 (0.062)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>-1 (0.025)</td>
<td>-1 (0.022)</td>
<td>-1 (0.025)</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Log(Bilirubin) $t_{1/2}$</td>
<td>0.083 (0.0078)</td>
<td>0.081 (0.0065)</td>
<td>0.084 (0.0078)</td>
</tr>
<tr>
<td>Late Increasing Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.93 (0.035)</td>
<td>0.95 (0.032)</td>
<td>0.93 (0.035)</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>1.2 (0.026)</td>
<td>1.2 (0.025)</td>
<td>1.2 (0.026)</td>
</tr>
<tr>
<td>Early Peaking Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-0.22 (0.035)</td>
<td>-0.32 (0.041)</td>
<td>-0.22 (0.035)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>-0.77 (0.0057)</td>
<td>-0.78 (0.0062)</td>
<td>-0.78 (0.0057)</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GFR $t_{1/2}$</td>
<td>0.31 (0.012)</td>
<td>0.32 (0.014)</td>
<td>0.31 (0.012)</td>
</tr>
<tr>
<td>Late Increasing Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.64 (0.022)</td>
<td>0.69 (0.028)</td>
<td>0.64 (0.022)</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>1.7 (0.054)</td>
<td>1.8 (0.072)</td>
<td>1.7 (0.054)</td>
</tr>
<tr>
<td><strong>Var(\epsilon)</strong> $\Sigma$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.38 0</td>
<td>0.38 0</td>
<td>0.38 0</td>
</tr>
<tr>
<td>$\gamma_{1,1,c}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.40 0.41</td>
<td>0.40 0.44</td>
<td>0.40 0.41</td>
</tr>
<tr>
<td><strong>Var(b)</strong> $G$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.21 0.37</td>
<td>0.40 0.40</td>
<td>-0.21 0.37</td>
</tr>
</tbody>
</table>

### Table 4.7: ML Estimates of association parameters under 3 different cause-specific multi-phase hazard models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ignorable Missing -Model 1-</th>
<th>Semi-Joint -Model 2-</th>
<th>Full Joint -Model 3-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Competing Risks Process</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Peaking Phase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-2.2 (0.35)</td>
<td>-5.9 (1.8)</td>
<td>-6.03 (2.0)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>-0.68 (0.22)</td>
<td>-0.18 (0.071)</td>
<td>-0.18 (0.52)</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Death before Transplant $t_{1/2}$</td>
<td>0.046 (0.016)</td>
<td>0.059 (0.0047)</td>
<td>0.059 (0.0048)</td>
</tr>
<tr>
<td>Association Parameter - $\gamma_{1,1}$ &amp; $c$</td>
<td>- 3.1 (1.0)</td>
<td>3.2 (1.1)</td>
<td></td>
</tr>
<tr>
<td>Association Parameter - $\gamma_{1,1}$ &amp; $c$</td>
<td>-0.36 (0.68)</td>
<td>-0.12 (0.55)</td>
<td></td>
</tr>
<tr>
<td>Constant Phase $\beta_0$</td>
<td>-0.67 (0.18)</td>
<td>-0.48 (0.12)</td>
<td>-0.54 (0.13)</td>
</tr>
<tr>
<td>Association Parameter - $\gamma_{1,2}$ &amp; $c$</td>
<td>- 0.83 (0.16)</td>
<td>0.84 (0.15)</td>
<td></td>
</tr>
<tr>
<td>Association Parameter - $\gamma_{1,2}$ &amp; $c$</td>
<td>-0.61 (0.20)</td>
<td>-0.54 (0.20)</td>
<td></td>
</tr>
<tr>
<td>Intermediate Peaking Phase $\beta_0$</td>
<td>-69 (0.35)</td>
<td>-0.00072 (0.39)</td>
<td>0.63 (0.16)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>-0.63 (0.038)</td>
<td>-0.63 (0.038)</td>
<td>-0.63 (0.039)</td>
</tr>
<tr>
<td>$m$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transplant $t_{1/2}$</td>
<td>0.43 (0.041)</td>
<td>0.43 (0.041)</td>
<td>0.43 (0.042)</td>
</tr>
<tr>
<td>Association Parameter - $\gamma_{1,1}$ &amp; $c$</td>
<td>- 0.19 (0.11)</td>
<td>-0.13 (0.11)</td>
<td></td>
</tr>
<tr>
<td>Association Parameter - $\gamma_{1,1}$ &amp; $c$</td>
<td>-0.12 (0.11)</td>
<td>-0.03 (0.11)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{2,1}$ $b$</td>
<td>-0.19 (0.11)</td>
<td>-0.13 (0.11)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{2,1}$ $b$</td>
<td>-0.12 (0.11)</td>
<td>-0.03 (0.11)</td>
<td></td>
</tr>
</tbody>
</table>

\[ a_{p > 0.05} \quad b_{p < 0.01} \]

\[ \gamma_{k,c}^{l} \] - the association between \( k^{th} \) longitudinal response and \( c^{th} \) time phase of the \( j^{th} \) competing risk

\[ \gamma_{p,c}^{k} \] - the association between \( k^{th} \) longitudinal response and \( c^{th} \) time phase of the \( j^{th} \) competing risk

\[ \gamma_{p,c}^{k} \] - the association between \( k^{th} \) longitudinal response and \( c^{th} \) time phase of the \( j^{th} \) competing risk
### Longitudinal Process

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ± SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.18 ± 0.032</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Early Decreasing Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>-0.29 ± 0.024</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Baseline bilirubin</td>
<td>0.30 ± 0.0068</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Late Increasing Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline bilirubin</td>
<td>0.089 ± 0.015</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.72 ± 0.034</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Early Peaking Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>0.11 ± 0.019</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Baseline GFR</td>
<td>0.51 ± 0.017</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Late Increasing Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>0.60 ± 0.082</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Baseline GFR</td>
<td>0.13 ± 0.058</td>
<td>.03</td>
</tr>
</tbody>
</table>

### Competing Risks Process

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ± SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline GFR</td>
<td>-0.41 ± 0.24</td>
<td>.09</td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>-0.69 ± 0.20</td>
<td>.0007</td>
</tr>
<tr>
<td>Association Parameter $\gamma_1$</td>
<td>1.2 ± 0.13</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Association Parameter $\gamma_2$</td>
<td>-0.39 ± 0.20</td>
<td>.05</td>
</tr>
<tr>
<td>Age</td>
<td>0.41 ± 0.25</td>
<td>.1</td>
</tr>
<tr>
<td>Male</td>
<td>0.43 ± 0.17</td>
<td>.009</td>
</tr>
<tr>
<td>Association Parameter $\gamma_1$</td>
<td>-0.40 ± 0.12</td>
<td>.001</td>
</tr>
<tr>
<td>Association Parameter $\gamma_2$</td>
<td>-0.17 ± 0.13</td>
<td>.2</td>
</tr>
</tbody>
</table>

Table 4.8: ML Regression Estimates of baseline covariates and the association parameters in the joint-model (non-ignorable (Model 3) mechanism) when using Weibull baseline cause-specific hazards.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ± SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Longitudinal Process</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.14 ± 0.025</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Early Decreasing Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(Bilirubin) Device type: LVAD</td>
<td>-0.27 ± 0.022</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Baseline bilirubin</td>
<td>0.31 ± 0.0068</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Late Increasing Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline bilirubin</td>
<td>0.088 ± 0.015</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.76 ± 0.036</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Early Peaking Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFR Device type: LVAD</td>
<td>0.068 ± 0.020</td>
<td>.0009</td>
</tr>
<tr>
<td>Baseline GFR</td>
<td>0.62 ± 0.020</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td><strong>Late Increasing Phase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFR Device type: LVAD</td>
<td>0.61 ± 0.088</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Baseline GFR</td>
<td>0.25 ± 0.067</td>
<td>.0002</td>
</tr>
<tr>
<td><strong>Competing Risks Process</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline GFR</td>
<td>-0.47 ± 0.25</td>
<td>.06</td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>-0.68 ± 0.20</td>
<td>.0009</td>
</tr>
<tr>
<td>Association Parameter $\gamma_1$</td>
<td>1.1 ± 0.13</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Association Parameter $\gamma_2$</td>
<td>-0.39 ± 0.20</td>
<td>.05</td>
</tr>
<tr>
<td>Male</td>
<td>0.44 ± 0.17</td>
<td>.009</td>
</tr>
<tr>
<td>Association Parameter $\gamma_1$</td>
<td>-0.43 ± 0.12</td>
<td>.0005</td>
</tr>
<tr>
<td>Association Parameter $\gamma_2$</td>
<td>-0.16 ± 0.13</td>
<td>.2</td>
</tr>
</tbody>
</table>

Table 4.9: **ML Regression Estimates of baseline covariates and the association parameters in the joint-model (non-ignorable (Model 3) mechanism) when using piecewise exponential baseline cause-specific hazards.**
Table 4.10: ML Regression Estimates of baseline covariates and the association parameters in the joint-model (non-ignorable (Model 3) mechanism) when using multi-phase cause-specific hazards.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate ± SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Longitudinal Process</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>0.14 ± 0.026</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age</td>
<td>0.14 ± 0.026</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Early Decreasing Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>-0.27 ± 0.022</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Baseline bilirubin</td>
<td>0.31 ± 0.0068</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Late Increasing Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline bilirubin</td>
<td>0.088 ± 0.015</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>GFR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>-0.77 ± 0.036</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Early Peaking Phase</td>
<td>0.068 ± 0.020</td>
<td>.0009</td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>0.068 ± 0.020</td>
<td>.0009</td>
</tr>
<tr>
<td>Baseline GFR</td>
<td>0.62 ± 0.020</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Late Increasing Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>0.61 ± 0.088</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Baseline GFR</td>
<td>0.25 ± 0.067</td>
<td>.0002</td>
</tr>
<tr>
<td><strong>Competing Risks Process</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Peaking Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.7 ± 1.2</td>
<td>.14</td>
</tr>
<tr>
<td>Device type: LVAD</td>
<td>-2.3 ± 0.83</td>
<td>.006</td>
</tr>
<tr>
<td>Association Parameter- γ</td>
<td>0.91 ± 0.38</td>
<td>.02</td>
</tr>
<tr>
<td>Association Parameter- γ</td>
<td>-0.72 ± 0.76</td>
<td>.3</td>
</tr>
<tr>
<td>Death before Transplant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline GFR</td>
<td>-0.43 ± 0.29</td>
<td>.14</td>
</tr>
<tr>
<td>Association Parameter- γ</td>
<td>1.2 ± 0.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Association Parameter- γ</td>
<td>-0.35 ± 0.26</td>
<td>.17</td>
</tr>
<tr>
<td>Constant Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transplant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.51 ± 0.20</td>
<td>.01</td>
</tr>
<tr>
<td>Association Parameter- γ</td>
<td>-0.44 ± 0.14</td>
<td>.002</td>
</tr>
<tr>
<td>Association Parameter- γ</td>
<td>-0.17 ± 0.13</td>
<td>.2</td>
</tr>
</tbody>
</table>

Table 4.11: True value of association parameter.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>% Bias</th>
<th>AvgSE</th>
<th>EmpSE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Longitudinal Process: (Y^1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Early decreasing phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu)</td>
<td>-1</td>
<td>0.059</td>
<td>0.0117</td>
<td>0.0114</td>
<td>0.951</td>
</tr>
<tr>
<td>(t_{1/2})</td>
<td>0.25</td>
<td>-0.206</td>
<td>0.0154</td>
<td>0.0148</td>
<td>0.935</td>
</tr>
<tr>
<td>Binary-(x_1)</td>
<td>1.5</td>
<td>-0.129</td>
<td>0.0515</td>
<td>0.0490</td>
<td>0.967</td>
</tr>
<tr>
<td><strong>Late increasing phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu)</td>
<td>1</td>
<td>0.800</td>
<td>0.1946</td>
<td>0.1819</td>
<td>0.902</td>
</tr>
<tr>
<td>(t_{1/2})</td>
<td>5</td>
<td>-2.73</td>
<td>1.368</td>
<td>1.526</td>
<td>0.931</td>
</tr>
<tr>
<td>Continuous-(x_2)</td>
<td>-1</td>
<td>5.35</td>
<td>0.2910</td>
<td>0.3240</td>
<td>0.939</td>
</tr>
<tr>
<td><strong>Longitudinal Process: (Y^2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Early decreasing phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu)</td>
<td>-0.75</td>
<td>0.074</td>
<td>0.0071</td>
<td>0.0073</td>
<td>0.943</td>
</tr>
<tr>
<td>(t_{1/2})</td>
<td>0.5</td>
<td>-0.395</td>
<td>0.0460</td>
<td>0.0456</td>
<td>0.935</td>
</tr>
<tr>
<td>Continuous-(x_2)</td>
<td>1</td>
<td>-0.346</td>
<td>0.0254</td>
<td>0.0263</td>
<td>0.935</td>
</tr>
<tr>
<td><strong>Late increasing phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\nu)</td>
<td>0.75</td>
<td>-2.97</td>
<td>0.2264</td>
<td>0.2286</td>
<td>0.902</td>
</tr>
<tr>
<td>(t_{1/2})</td>
<td>4</td>
<td>-7.89</td>
<td>1.722</td>
<td>1.770</td>
<td>0.919</td>
</tr>
<tr>
<td>Binary-(x_1)</td>
<td>-1</td>
<td>6.49</td>
<td>0.7644</td>
<td>0.6841</td>
<td>0.947</td>
</tr>
<tr>
<td><strong>Measurement Error Variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma^2_{\epsilon_1})</td>
<td>0.36</td>
<td>0.172</td>
<td>0.0113</td>
<td>0.0111</td>
<td>0.951</td>
</tr>
<tr>
<td>(\sigma^2_{\epsilon_2})</td>
<td>0.49</td>
<td>0.454</td>
<td>0.0153</td>
<td>0.0153</td>
<td>0.939</td>
</tr>
<tr>
<td><strong>Random Effects Variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma^2_{b_1})</td>
<td>0.40</td>
<td>0.354</td>
<td>0.0335</td>
<td>0.0342</td>
<td>0.955</td>
</tr>
<tr>
<td>(\sigma^2_{b_2})</td>
<td>0.35</td>
<td>0.277</td>
<td>0.0317</td>
<td>0.0290</td>
<td>0.951</td>
</tr>
<tr>
<td>(Cov(b^1, b^2))</td>
<td>-0.3</td>
<td>-0.0005</td>
<td>0.0256</td>
<td>0.0262</td>
<td>0.967</td>
</tr>
<tr>
<td><strong>Competing Risk: 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binary-(x_1)</td>
<td>0.75</td>
<td>0.562</td>
<td>0.1591</td>
<td>0.1572</td>
<td>0.947</td>
</tr>
<tr>
<td>Continuous-(x_2)</td>
<td>1</td>
<td>-0.900</td>
<td>0.0839</td>
<td>0.0887</td>
<td>0.951</td>
</tr>
<tr>
<td><strong>Competing Risk: 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binary-(x_1)</td>
<td>-1</td>
<td>2.23</td>
<td>0.1968</td>
<td>0.1872</td>
<td>0.951</td>
</tr>
<tr>
<td>Continuous-(x_2)</td>
<td>-1</td>
<td>1.19</td>
<td>0.1177</td>
<td>0.1147</td>
<td>0.967</td>
</tr>
<tr>
<td><strong>Association Parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\gamma^1_1)</td>
<td>1.0</td>
<td>1.24</td>
<td>0.2649</td>
<td>0.2882</td>
<td>0.939</td>
</tr>
<tr>
<td>(\gamma^2_1)</td>
<td>-1</td>
<td>1.91</td>
<td>0.2944</td>
<td>0.3123</td>
<td>0.947</td>
</tr>
<tr>
<td>(\gamma^1_2)</td>
<td>-0.5</td>
<td>-6.87</td>
<td>0.3446</td>
<td>0.3593</td>
<td>0.951</td>
</tr>
<tr>
<td>(\gamma^2_2)</td>
<td>1</td>
<td>-5.45</td>
<td>0.3882</td>
<td>0.3816</td>
<td>0.959</td>
</tr>
</tbody>
</table>

Table 4.12: **Case 1**: Summary measures of the parameter estimates based on the 250 Monte Carlo Simulations with two covariates in the joint model.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>% Bias</th>
<th>AvgSE</th>
<th>EmpSE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Longitudinal Process: ( Y^1 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early decreasing phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu )</td>
<td>0.134</td>
<td>0.0138</td>
<td>0.0138</td>
<td>0.963</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>2.11</td>
<td>0.0227</td>
<td>0.0256</td>
<td>0.935</td>
</tr>
<tr>
<td>Late increasing phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu )</td>
<td>-3.04</td>
<td>0.1965</td>
<td>0.2134</td>
<td>0.894</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>0.090</td>
<td>0.9354</td>
<td>0.9151</td>
<td>0.922</td>
</tr>
<tr>
<td><strong>Longitudinal Process: ( Y^2 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early decreasing phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu )</td>
<td>0.054</td>
<td>0.0076</td>
<td>0.0074</td>
<td>0.963</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>0.308</td>
<td>0.0596</td>
<td>0.0631</td>
<td>0.943</td>
</tr>
<tr>
<td>Late increasing phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu )</td>
<td>-2.45</td>
<td>0.1700</td>
<td>0.1768</td>
<td>0.918</td>
</tr>
<tr>
<td>( t_{1/2} )</td>
<td>-1.92</td>
<td>0.8067</td>
<td>0.7421</td>
<td>0.935</td>
</tr>
<tr>
<td><strong>Measurement Error Variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma^2_{\epsilon_1} )</td>
<td>-0.276</td>
<td>0.0110</td>
<td>0.0116</td>
<td>0.947</td>
</tr>
<tr>
<td>( \sigma^2_{\epsilon_2} )</td>
<td>0.286</td>
<td>0.0148</td>
<td>0.0144</td>
<td>0.951</td>
</tr>
<tr>
<td><strong>Random Effects Variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma^2_{b_1} )</td>
<td>0.569</td>
<td>0.0330</td>
<td>0.0339</td>
<td>0.931</td>
</tr>
<tr>
<td>( \sigma^2_{b_2} )</td>
<td>0.313</td>
<td>0.0311</td>
<td>0.0324</td>
<td>0.931</td>
</tr>
<tr>
<td>( Cov(b_1, b_2) )</td>
<td>0.196</td>
<td>0.0272</td>
<td>0.0275</td>
<td>0.939</td>
</tr>
<tr>
<td><strong>Association Parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma^1_1 )</td>
<td>-2.55</td>
<td>0.2859</td>
<td>0.2891</td>
<td>0.943</td>
</tr>
<tr>
<td>( \gamma^2_1 )</td>
<td>2.18</td>
<td>0.3197</td>
<td>0.3329</td>
<td>0.951</td>
</tr>
<tr>
<td>( \gamma^1_2 )</td>
<td>1.65</td>
<td>0.3029</td>
<td>0.3089</td>
<td>0.951</td>
</tr>
<tr>
<td>( \gamma^2_2 )</td>
<td>-1.29</td>
<td>0.3409</td>
<td>0.3405</td>
<td>0.971</td>
</tr>
</tbody>
</table>

Table 4.13: Case 2: Summary measures of the parameter estimates based on the 250 Monte Carlo Simulations without covariates in the joint model.
<table>
<thead>
<tr>
<th>Association Parameters</th>
<th>Statistic</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n = 400</td>
</tr>
<tr>
<td>$\gamma_1^1$</td>
<td>% Bias</td>
<td>-2.55</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.2859</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.2891</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.943</td>
</tr>
<tr>
<td>$\gamma_1^2$</td>
<td>% Bias</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.3197</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.3329</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.951</td>
</tr>
<tr>
<td>$\gamma_1^2$</td>
<td>% Bias</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.3029</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.3089</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.951</td>
</tr>
<tr>
<td>$\gamma_2^1$</td>
<td>% Bias</td>
<td>-1.29</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.3409</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.3405</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.971</td>
</tr>
</tbody>
</table>

Table 4.14: **Case 3:** Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on varying sample sizes.

<table>
<thead>
<tr>
<th>Association Parameters</th>
<th>Statistic</th>
<th>Censoring distribution: $\exp(\lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\lambda = 0.25$</td>
</tr>
<tr>
<td>$\gamma_1^1$</td>
<td>% Bias</td>
<td>-2.55</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.2859</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.2891</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.943</td>
</tr>
<tr>
<td>$\gamma_1^2$</td>
<td>% Bias</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.3197</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.3329</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.951</td>
</tr>
<tr>
<td>$\gamma_1^2$</td>
<td>% Bias</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.3029</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.3089</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.951</td>
</tr>
<tr>
<td>$\gamma_2^1$</td>
<td>% Bias</td>
<td>-1.29</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.3409</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.3405</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.971</td>
</tr>
</tbody>
</table>

Table 4.15: **Case 4:** Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on varying censoring rates.
<table>
<thead>
<tr>
<th>Association Parameters</th>
<th>Statistic</th>
<th>High Correlation: Corr($b^1, b^2$) = -0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ̂_1</td>
<td>% Bias</td>
<td>Full Joint Model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2.55</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.2859</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.2891</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.943</td>
</tr>
<tr>
<td>γ̂_2</td>
<td>% Bias</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.3197</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.3329</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.951</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Association Parameters</th>
<th>Statistic</th>
<th>Low Correlation: Corr($b^1, b^2$) = -0.13</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ̂_1</td>
<td>% Bias</td>
<td>Full Joint Model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1.30</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.1418</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.1442</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.952</td>
</tr>
<tr>
<td>γ̂_2</td>
<td>% Bias</td>
<td>-1.19</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.1638</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.1720</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.952</td>
</tr>
<tr>
<td>γ̂_1</td>
<td>% Bias</td>
<td>-3.19</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.1539</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.1501</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.956</td>
</tr>
<tr>
<td>γ̂_2</td>
<td>% Bias</td>
<td>-2.42</td>
</tr>
<tr>
<td></td>
<td>AvgSE</td>
<td>0.1800</td>
</tr>
<tr>
<td></td>
<td>EmpSE</td>
<td>0.2006</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.931</td>
</tr>
</tbody>
</table>

Table 4.16: Case 5: Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on varying association between random effects $b^1$ and $b^2$. 
\[
W^1 = b^1 - b^2 \quad W^2 = (b^1)^2 - (b^2)^2
\]

<table>
<thead>
<tr>
<th>Association Parameters</th>
<th>Statistic</th>
<th>Frailty:W</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1^1$</td>
<td>$W^2 = -0.5b^1 + b^2$</td>
<td>$W^2 = -0.5b^1 + b^2$</td>
</tr>
<tr>
<td>% Bias</td>
<td>-2.55</td>
<td>-68.0</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.2859</td>
<td>0.2856</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.2891</td>
<td>0.3064</td>
</tr>
<tr>
<td>CP</td>
<td>0.943</td>
<td>0.351</td>
</tr>
<tr>
<td>$\gamma_1^2$</td>
<td>$W^2 = -0.5b^1 + b^2$</td>
<td>$W^2 = -0.5b^1 + b^2$</td>
</tr>
<tr>
<td>% Bias</td>
<td>2.18</td>
<td>138</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.3197</td>
<td>0.2853</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.3329</td>
<td>0.3145</td>
</tr>
<tr>
<td>CP</td>
<td>0.951</td>
<td>0.024</td>
</tr>
<tr>
<td>$\gamma_2^1$</td>
<td>$W^2 = -0.5b^1 + b^2$</td>
<td>$W^2 = -0.5b^1 + b^2$</td>
</tr>
<tr>
<td>% Bias</td>
<td>1.65</td>
<td>5.33</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.3029</td>
<td>0.3227</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.3089</td>
<td>0.2919</td>
</tr>
<tr>
<td>CP</td>
<td>0.951</td>
<td>0.967</td>
</tr>
<tr>
<td>$\gamma_2^2$</td>
<td>$W^2 = -0.5b^1 + b^2$</td>
<td>$W^2 = -0.5b^1 + b^2$</td>
</tr>
<tr>
<td>% Bias</td>
<td>-1.29</td>
<td>-4.00</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.3409</td>
<td>0.3341</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.3405</td>
<td>0.2976</td>
</tr>
<tr>
<td>CP</td>
<td>0.971</td>
<td>0.976</td>
</tr>
</tbody>
</table>

Table 4.17: **Case 6:** Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on non-normal frailty.
<table>
<thead>
<tr>
<th>Association Parameters</th>
<th>Statistic</th>
<th>Bivariate Normal</th>
<th>Bivariate $t$-distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma_1^1$</td>
<td>$N_2(0, \Sigma)$</td>
<td>$t_2(\Sigma', \text{df}=3)$</td>
</tr>
<tr>
<td>% Bias</td>
<td>-2.55</td>
<td>2.67</td>
<td>-0.131</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.2859</td>
<td>0.3576</td>
<td>0.2906</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.2891</td>
<td>0.3858</td>
<td>0.3233</td>
</tr>
<tr>
<td>CP</td>
<td>0.943</td>
<td>0.931</td>
<td>0.923</td>
</tr>
<tr>
<td>$\gamma_1^2$</td>
<td>2.18</td>
<td>-6.07</td>
<td>-4.71</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.3197</td>
<td>0.4000</td>
<td>0.3252</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.3329</td>
<td>0.4363</td>
<td>0.3812</td>
</tr>
<tr>
<td>CP</td>
<td>0.951</td>
<td>0.935</td>
<td>0.919</td>
</tr>
<tr>
<td>$\gamma_2^1$</td>
<td>1.65</td>
<td>-0.397</td>
<td>-6.11</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.3029</td>
<td>0.3749</td>
<td>0.3054</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.3089</td>
<td>0.4066</td>
<td>0.2801</td>
</tr>
<tr>
<td>CP</td>
<td>0.951</td>
<td>0.935</td>
<td>0.955</td>
</tr>
<tr>
<td>$\gamma_2^2$</td>
<td>-1.29</td>
<td>-4.27</td>
<td>-1.56</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.3409</td>
<td>0.4252</td>
<td>0.3463</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.3405</td>
<td>0.4620</td>
<td>0.3280</td>
</tr>
<tr>
<td>CP</td>
<td>0.971</td>
<td>0.940</td>
<td>0.964</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Var({b_1^1, b_2^2}^T)$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$var(b_1^1)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias</td>
<td>0.569</td>
<td>13.9</td>
<td>0.398</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.0329</td>
<td>0.0305</td>
<td>0.0333</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.0339</td>
<td>0.2390</td>
<td>0.0372</td>
</tr>
<tr>
<td>CP</td>
<td>0.931</td>
<td>0.274</td>
<td>0.915</td>
</tr>
<tr>
<td>$var(b_2^2)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias</td>
<td>0.313</td>
<td>13.3</td>
<td>0.717</td>
</tr>
<tr>
<td>AvgSE</td>
<td>0.0311</td>
<td>0.0293</td>
<td>0.0312</td>
</tr>
<tr>
<td>EmpSE</td>
<td>0.0325</td>
<td>0.2318</td>
<td>0.0354</td>
</tr>
<tr>
<td>CP</td>
<td>0.931</td>
<td>0.322</td>
<td>0.923</td>
</tr>
</tbody>
</table>

| $cov(b_1^1, b_2^2)$       |   |   |   |
| % Bias                    | 0.196 | 11.8 | 0.138 |
| AvgSE                     | 0.0272 | 0.0252 | 0.0274 |
| EmpSE                     | 0.0275 | 0.2327 | 0.0314 |
| CP                        | 0.939 | 0.274 | 0.927 |

Table 4.18: **Case 7:** Summary measures of the association parameter estimates based on the 250 Monte Carlo Simulations based on non-normal random effect.
References

[1] Heart Disease and Stroke Statistics - 2010 update at a glance, American Heart Association, Dallas Texas. 2


[6] Shreenivas SS, Rame JE, Jessup M, Mechanical Circulatory Support as a Bridge to Transplant or for Destination Therapy, Current Heart Failure Report 2010; 7: 159-166. 3, 4


146
[34] Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence Biometrika: 1978; 65, 141-151. 23


[38] Pepe M, Mori M. Kaplan-Meier marginal or conditional probability curves in summarizing competing risks failure data? Statistics in Medicine: 1993; 12, 737-751. 28


[44] Larson MG, Dinse GE, A Mixture model for the regression analysis of competing risks data Applied Statistics: 1985; 34, 201-211. 40, 103


148
[52] O’Brian PC, Procedures for comparing samples with multiple endpoints, Biometrics, 1984; 40: 1079-87. 50


[54] Patterson HD, Thompson R Recovery of inter-block information when block sizes are unequal, Biometrika, 1971; 58: 545-54. 52


[57] Wedderburn RWM, Quasi-likelihood functions, generalized linear models and Gaussian method, Biometrika, 1974; 61: 439-47. 52

[58] Laird NM, Ware JH, Random-Effects Models for Longitudinal Data, Biometrics, 1982; 38: 963-74. 52, 54


[81] Wu L, A Joint Model for Nonlinear Mixed-Effects Models with Censoring and Co-
variates Measured with Error, With Application to AIDS Studies, J Am Stat Assoc,
2002;97: 955-964. 62, 66

[82] Hazelrig, J.B., Turner, M.E., Jr., Blackstone E. H., Parametric Survival Analysis
Combining Longitudinal and Cross-sectional Censored and Interval-Censored Data
with Concomitant information, Biometrics, 1982; 39:1-5. 62, 66

[83] Odell PM, Anderson KM, D’Agostino RB., Maximum Likelihood Estimation for
interval-censored Data Using Weibull-Based Accelerated Failure Time Model, Bio-
metrics, 1992; 48:951-959. 66

[84] Vonesh E.F., Chincill V.M., Pu K., Goodness-of-Fit in Generalized Nonlinear
Mixed-Effects Models, Biometrics, 1996; 52:572-587. 74, 75

[85] Lin L.I., A Concordance Correlation Coefficient to Evaluate Reproducibility, Bio-
metrics, 1989; 45:255-268. 74

[86] Beckett LA., Tancredi DJ., Wilson RS., Multivariate longitudinal models for com-
plex change processes, Statistics in Medicine, 2004; 23: 231-239. 84

[87] Faes C., Aerts M., Molenberghs G., Geys H., Teuns G., Bijnens L., A high-
dimensional joint model for longitudinal outcomes of different nature, Statistics
in Medicine, 2008; 27: 4408-4427. 84, 87

[88] Abramowitz M, Stegun I., Handbook of Mathematical Functions Dover, New York,
1972. 85

[89] Sheiner LB, Beal SL, Pharmacokinetic parameter estimates from several least
squares procedures: superiority of extended least squares, Journal of Pharmacoki-
netics and Biopharmaceutics, 1985; 13: 185-201. 86
[90] Beal SL, Sheiner LB. *Heteroscedastic Nonlinear Regression*, Technometrics, 1988; **30**: 327-338. 86


[114] Schluchter MD, Greene T, Beck GJ, *Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progression of renal disease*. Statistics in Medicine, 2001; **20**: 989-1007. 100


[118] Liu L, Ma JZ, O’Quigley J, Joint analysis of multi-level repeated measures data and survival: an application to the end stage renal disease (ESRD) data. Statistics in Medicine, 2008; 27: 5679-5691. 101


[126] Hu W, Li G, Li N, A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. Statistics in Medicine, 2009; 28: 1601-1619. 103


[128] Li N, Elashoff RM, Li G, Saver J, Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. Statistics in Medicine, 2010; 29: 546-557. 103

[129] Dunnett CW, Sobel M, A Bivariate Generalization of Student’s t-Distribution, with Tables for Certain Special Cases, Biometrika, 1954; 41: 153-169. 131