ABSTRACT

Primary Biliary Cirrhosis (PBC) is a relatively rare chronic liver disease that mainly affects women. When someone's immune system attacks the liver, the bile ducts are damaged and accumulated in the liver. Over time, it will lead to fibrosis and cirrhosis of the liver. PBC progresses differently among patients and its severity is indicated by repeated measurements of longitudinal biomarkers. In practice, insights on how biomarkers associate with death risk contribute to better adjustment of personal care and improvement of treatment regimen generally.

In this project, we are interested in the association between the biomarker serum bilirubin and overall survival of PBC patients. When the liver fails to excrete bilirubin, high levels of this serum can cause jaundice of the skin, which is a common symptom of cirrhosis. This association is investigated with three different statistical approaches: Cox Proportional Hazards Model, Time-Dependent Cox Model, and Joint Model for Longitudinal and Time-To-Event Data. For each of the three models, the following procedure is applied: univariate analysis, variable selection, and multivariate analysis. The study data comes from a PBC clinical trial conducted by the Mayo Clinic over 10 years from 1974 to 1984. The hazard ratios estimated from these three models are compared.

Intuitively, the difference in the estimated hazard ratios can be explained by the different levels of information considered. The Cox Proportional Hazards model uses the baseline values of bilirubin. The Time-Dependent Cox model uses the current values of bilirubin by accounting for the changes of bilirubin over time. The Joint Model captures the internal progression of bilirubin and measurement errors. For applications where sample size is large and computational resources are available, Joint Models should be used because they reduce potential bias in parameter estimation relative to the other models in survival analysis.

Statistical Analysis of the Association between Bilirubin and Survival in Primary Biliary Cirrhosis

Amelia Huong Tran

Advisor
Marie Ozanne

A thesis submitted to the Department of Mathematics and Statistics in partial fulfillment of the requirements for the degree of Bachelor of Arts with Honors

Department of Mathematics and Statistics
Mount Holyoke College
South Hadley, MA 01075
May 2021

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor, Professor Marie Ozanne, for giving me the opportunity to continue my summer project and to further explore my passion for survival analysis. I am grateful for her dedication and encouragement during such challenging and uncertain times. My scientific writing has greatly improved under her guidance.

Professor Margaret Robinson has been a wonderful academic advisor. I am grateful for her insights and patience, and for allowing me to make mistakes and to grow during the past four years. I would like to extend my thanks to Professor Audrey St. John for her valuable advice and perspectives, which I will carry with me in my future endeavors. Many thanks to Professor Samuel Mitchell and Professor Amy Nussbaum for kindly accepting my invitation to be the members of my thesis committee.

I would like to express my gratitude to Dr. Audrey Mauguen for her guidance and mentorship during my internship at Memorial Sloan Kettering Cancer Center, where I was first introduced to survival analysis. I owe my special thanks to Dr. Kay See Tan and Dr. Margaret Du for pushing for the QSURE program to be remote, against all the challenges. Kay See has always been there to support me in every possible way that she can, which I am incredibly grateful for.

I would like to thank Professor Evan Ray for his guidance on the copula project, my first statistical research. I appreciate him for his patience and support in all statistical matters. Under his mentorship, my computations have immensely advanced, and proved very valuable in my studies.

I would like to thank Professor Daniel Kelleher, my Probability instructor. His class brought me so much joy and sparked my small passion for mathematics and statistics. I thank him for his fabulous advice and encouragement, and for believing in me, even when I did not.

Last but not least, I am thankful to my parents and my friends for their unconditional love and constant support. Thank you for always having my back in every step of the way. I feel lucky to have you all in life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Primary Biliary Cirrhosis (PBC)

Primary Biliary Cirrhosis, also known as *Primary Biliary Cholangitis*, is a relatively rare chronic liver disease that mainly affects women (Kaplan, 1996). PBC is considered an autoimmune disease, which means that the immune system mistakenly destroys healthy cells and tissues. When the immune system attacks the liver tissues, it causes slow and progressive damage to the bile ducts in the liver, which aid with digestion and help the body get rid of cholesterol, toxins and worn-out red blood cells (Mayo Clinic, 2021). The damaged cells from bile ducts and other toxins build up in the liver and can sometimes lead to irreversible scarring of liver tissues, as shown in Figure 1.1. Many research experts think that PBC is triggered by genetic and environmental factors such as bacterial infections or toxic chemicals (Canadian Liver Foundation, 2021).

More than half of PBC patients do not experience noticeable symptoms. Usually the disease is detected in blood tests which are administered for other reasons (Mayo Clinic, 2021). PBC patients often experience common early symptoms such as itching, eye and mouth drying, and fatigue. In

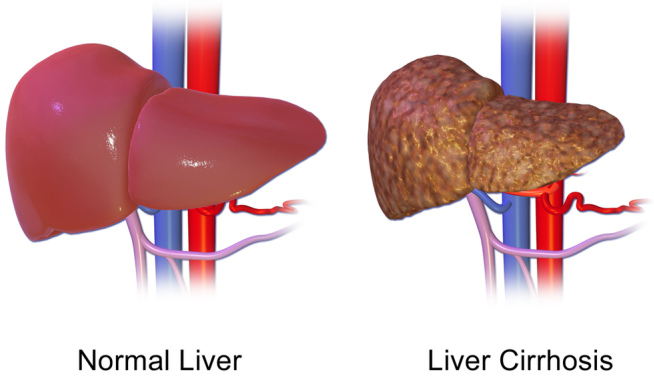

Figure 1.1: Liver Cirrhosis (Bruce Blaus, Wikimedia Commons)

advanced stages of PBC, later signs and symptoms may include high cholesterol, diarrhea, weight loss, abdominal pain, and yellowing of the skin. As liver damage worsens, PBC patients suffer from serious complications such as enlarged veins, swelling of the body, or liver cancer. Without early diagnosis and proper treatments, PBC will lead to fibrosis and eventually cirrhosis of the liver (Canadian Liver Foundation, 2021). Even though there is no cure for PBC, researchers have studied possible treatments such as medications and liver transplantation to slow the disease progression and to prevent complications. Doctors and healthcare workers have been analyzing PBC biomarkers, the biological measurements of the state of diseases, to further their understanding of this chronic health condition. These liver biomarkers offer valuable information on the disease prognosis and progression. In practice, meaningful insights on how liver biomarkers are associated with patients' survival contribute to better adjustments of personal care, and improve PBC treatment regimen, and allocate healthcare resources efficiently.

1.2 Thesis Objectives

In this project, we are interested in the association between liver biomarker serum bilirubin and overall survival of PBC patients. When the liver fails to excrete bilirubin, high levels of this serum can cause yellowing of the skin, a common symptom of cirrhosis (Schuppan and Afdhal, 2008). Insights on how serum bilirubin is associated with the death risk of patients are valuable for the PBC prognosis and progression. Researchers can personalize patient care and monitor treatment efficacy accordingly.

This association is investigated with three different statistical approaches: the Cox Proportional Hazards Model, the Time-Dependent Cox Model, and the Joint Model for Longitudinal and Time-To-Event Data (Joint Model). The difference among these models is the levels of information considered. The Cox Proportional Hazards model uses the baseline values of bilirubin measured at the initiation of the study. As a result, it does not take into consideration the dynamic nature of bilirubin when this biomarker changes over time. The Time-Dependent Cox model uses the current values of bilirubin by accounting for the longitudinal progression of bilirubin over time. However, it assumes that the level of bilirubin remains constant in the time interval between doctor visits. In practice, bilirubin dynamically generates itself and has a distinct longitudinal trajectory for each patient. Thus, this assumption is not realistic because it does not consider the measurement errors induced by biological variations of PBC patients. The Joint Model eliminates these shortcomings by capturing the longitudinal history of bilirubin while accounting for the measurement errors. Estimated hazard ratios from these statistical models are compared to assess the model performances. The difference in these estimates indicates potential

bias in the parameter estimation if some assumptions are overlooked and the most appropriate model is not used.

Bias reduction usually comes with model complexity, in forms of required sample size and computational cost. In this project, even though the Joint Model appears as the most optimal approach because it accommodates the limitations of other statistical models for censored data, the model has many parameters to estimate and requires a large sample size to avoid convergence issues in the estimation process. When the conditions are not properly satisfied, it is not ideal to use the Joint Model since the estimates might be even more biased than other statistical models. In addition to comparing the estimated hazard ratios from the Cox Proportional Hazards model, the Time-Dependent Cox model, and the Joint Model for Longitudinal and Time-To-Event Data, the thesis will evaluate the benefits and drawbacks of the Joint Model when its requirements are not fully met, and what is recommended to do in those cases.

Chapter 2

Background

2.1 Survival Analysis

Survival analysis is a collection of statistical procedures that aim to study the *time until an event of interest occurs* (Kleinbaum and Klein, 2010). In biomedical research, *time* can be years, months, or age from the initiation of a study until an event occurs; an *event* can be death, progression of diseases, appearance of tumors, or any designated experience of interest. The response variable *time* is a continuous outcome which is usually referred to as *survival time* or *failure time*. The variable *event* is a binary outcome which takes on value 1 if the event of interest occurs and value 0 otherwise. An example of survival analysis in health problems includes studies of the association between serum bilirubin and overall survival in Primary Biliary Cirrhosis (PBC). The event of interest is life failure and the outcome is time until death occurs.

In survival analysis, the response variable *time* is always positive, which means that its distribution is skewed right. This feature can be solved by data transformations such as logarithmic or square root if the normality condition is

required, for linear regression models for instance. Otherwise, nonparametric Wilcoxon's test can also be used because it requires no distributional assumption. However, another distinguishing feature of survival data is *censoring*, which occurs when the event of interest is not observed. Censoring may happen because: (1) event does not occur during the observation period; (2) subject is lost to follow-up during the study period; (3) subject withdraws from the study for either unknown or known reasons (Kleinbaum and Klein, 2010). Therefore, it is inappropriate to use a t-test, Wilcoxon's test, or logistic regression because they do not account for censoring. This motivates an advanced statistical model, the Cox regression model, to handle censored data.

In particular, there are different three types of censoring: (1) right censoring - event of interest occurs after a certain time point (i.e., a PBC patient was alive at the study termination or lost to follow-up during the study), (2) left censoring - event of interest occurs before a certain time point (i.e., a person was followed up until they became HIV positive; the exact time of their first exposure to the virus is unknown; it might have happened before their first recorded positive test), and (3) interval censoring - event of interest occurs between a known time interval (i.e., an HIV patient tested positive for AIDS; the patient might have developed the disease at some point between their pre-last and last doctor visits).

Censoring mechanisms can also be classified based on censoring information. Censoring is noninformative when the subject withdraws for reasons unrelated to the prognosis while informative censoring means that the subject withdraws for reasons related to the expected failure time. In survival analysis, censoring is often assumed to be independent, random, and non-informative (Kleinbaum and Klein, 2010). To describe censored data, *Kaplan-Meier curves*

use information on censoring and event times to visualize the survival probability and number of subjects at risk.

There has been extensive research to study Primary Biliary Cirrhosis (PBC) with statistical models in survival analysis. Murtaugh developed an updated Mayo model to predict short-term survival at any time in the course of the disease for PBC using the values of prognostic variables measured at the latest patient visit (Murtaugh et al., 1994). The predicted survivals showed that the updated Mayo model was superior to the original one in terms of accuracy. The New England Journal of Medicine published a research study on the efficacy of liver transplantation in PBC patients with the Mayo model (Markus et al., 1989). Liver transplantation was found to be an efficacious treatment for PBC patients, based on the estimated survival probability. In addition to studies on the survival time and treatment effect of PBC, there have been research on the longitudinal measurements of PBC biomarkers because of their sources of important information about the disease.

2.2 Longitudinal Studies

Longitudinal analysis involves data sets that consist of repeated measurements of covariates for each of the subjects. The defining characteristic of a longitudinal study is that subjects are followed up over time (Diggle et al., 2002). A longitudinal study can be designed either prospectively or retrospectively. Such data are common in health sciences as they offer important insights into the development of diseases and therefore improve the construction of patient care. For example, repeated measurements of serum bilirubin can be used to monitor the progression of PBC. In longitudinal analysis, repeated measurements of covariates can be visualized with *spaghetti plots* in which each path represents a subject trajectory in the cohort.

The common objectives often arising in the study of longitudinal effects are to investigate whether treatment means differ over time and to investigate individual changes over time (Verbeke, 1997). In a longitudinal setting, there is an obvious strong correlation between the repeated measurements of covariates taken on the same subject. Thus, it is inappropriate to use t-tests or linear regression because the assumption of independent observations is violated. Since covariates for each subject have a unique trajectory path, subjects in the cohort have different intercepts and slopes. This leads to linear mixed-effects models, which offer flexibility for these differences by having subject-specific random effects in addition to fixed effects for all subjects in the cohort.

In randomized clinical trials, patients are divided into placebo and treatment groups. Each patient is followed up and the outcome covariate of interest is recorded at predetermined times. However, in reality, these measurements are usually not fully observed because of dropouts or loss to follow-ups before

the completion of treatment studies. There are three different types of missing data mechanisms: Missing Completely At Random (MCAR), Missing At Random (MAR), and Missing Not At Random (MNAR). Data is MCAR when the missingness does not depend on observed or unobserved data, MAR when the missingness conditions on unobserved data, and MNAR depends on both observed and unobserved data (Rubin, 1976). In a longitudinal study, linear mixed-effects models are used to handle incomplete continuous data with repeated measurements. These models work under the assumption that data is missing at random.

There have been studies that employed statistical models to investigate PBC treatment effects more closely by taking into consideration the disease progression over time. Neuberger investigated the effect of D-penicillin in PBC patients in double blind controlled trials implementing the occurrence rate ratio, but was unable to establish any therapeutic benefit of D-penicillin (Neuberger et al., 1985). Lombard studied the treatment effect of Cyclosporine in a PBC clinical trial followed up for six years with Cox univariate and multivariate analyses. The researchers concluded that given that blood pressure and renal function are examined, Cyclosporin A has some therapeutic potential in PBC (Lombard et al., 1993).

In addition to survival analysis and longitudinal studies, there has been statistical research on PBC genetics. A genome-wide association study identified 12 new susceptibility loci for PBC to investigate genetic architecture of the disease (Mells et al., 2011). Other researchers focus more on measurable substances in human bodies called biomarkers to analyze PBC. An international follow-up study performed a meta-analysis to conclude that PBC biomarker alkaline and bilirubin are surrogate endpoints of patient outcomes (Lammers

et al., 2014).

In clinical trials, PBC patients are often examined and followed up for a certain period of time during which their body substances are closely monitored and mortality rates are registered. In these studies, time to death of each subject is divided into small time intervals corresponding to each doctor's visit. For each of these time intervals, events of interest and measurements of biomarkers are recorded. It is appealing to conduct statistical analysis using all this information by combining statistical models involving time-to-event data and repeated measurements. This motivates a joint analysis using statistical models from survival analysis and longitudinal studies.

2.3 Motivation for Joint Analysis

In biomedical research, there are different types of outcomes of interest recorded over time. Two of these outcomes are longitudinally measured responses and the time until an event of interest occurs. It is relevant to investigate their association structure to identify potential longitudinal prognostics for the time-to-event outcome. Specifically, biomarkers are generate differently among subjects and dynamically over time even on the same subject. Their dynamic changes may result from the health conditions of subjects. A thorough understanding of this nature can help researchers personalize patient care more efficiently. Examples of this type of investigation are the association between CD4 cell counts and time to AIDS, prostate-specific antigen levels and the time to the development of prostate cancer, and serum bilirubin and time to death in liver cirrhosis (Rizopoulos, 2012).

In this project, we are interested in measuring the association between serum bilirubin and overall survival of PBC patients. The PBC dataset includes recorded outcomes such as information on survival and longitudinal measurements on PBC biomarkers, including serum bilirubin. There are different approaches to study this relationship. Some physicians are interested in investigating how strongly related the baseline level of bilirubin is to overall survival of PBC patients. Others might be interested in the dynamic progression of bilirubin with respect to death risk since bilirubin is a time-dependent covariate. Therefore, it is comprehensive to study this association with different approaches using either the baseline measurements or longitudinal measurements of bilirubin to compare the results and decide the best statistical estimator.

For the purpose of this thesis, the association between serum bilirubin and overall survival is studied with three different statistical models: Cox Proportional Hazards Model, Time-Dependent Cox Model, and Joint Model for Longitudinal and Time-To-Event Data. Intuitively, the Cox Proportional Hazards model uses the baseline level of bilirubin. The Time-Dependent Cox model uses the current level of bilirubin with the assumption that the progression of bilirubin is unaffected by failure time of patients. The Joint Model uses the current level of bilirubin on the condition that its progression is dependent on the health condition of patients.

Chapter 3

Statistical Methods

3.1 Cox Proportional Hazards (Cox PH)

Sometimes called *Cox regression model*, the Cox PH is the most common statistical model in survival analysis (Therneau and Grambsch, 2000). It assesses the effects of multiple covariates simultaneously by the hazard function:

$$\begin{aligned} h_i(t \mid w_i) &= h_0(t) \times \exp(\gamma^T w_i) \\ &= h_0(t) \times \exp(\gamma_1 w_{i1} + \gamma_2 w_{i2} + ... + \gamma_p w_{ip}) \end{aligned} \tag{3.1}$$

where $w_i^T = [w_{i1}, w_{i2}, ..., w_{ip}]$ denotes the vector of covariates and γ^T denotes the vector of corresponding regression coefficients. The Cox PH works under the assumption that all subjects share the same *baseline risk* function $h_0(t)$ which only depends on time t.

The baseline risk function $h_0(t)$ is unspecified to avoid mis-specifying the distribution of survival time. The individual multiplier $\exp(\gamma^T w_i)$ is a constant, time-independent exponential function of linear regression of covariates

unique to each subject. The Cox PH model assumes multiplicative effects of the covariates on the hazard for an event. In other words, for Subject i, each unit increase in γ_j associates with a multiplicative change in the hazard by $\exp(w_{ij})$.

From this setup, the hazard ratio (HR), which is an estimate of the hazard rate between groups, is assumed to be constant over time:

$$\begin{aligned} \frac{h_i(t \mid w_i)}{h_k(t \mid w_k)} &= \frac{h_0(t)\exp(\gamma^T w_i)}{h_0(t)\exp(\gamma^T w_k)} \\ &= \exp\{\gamma^T(w_i - w_k)\} \end{aligned} \tag{3.2}$$

If HR > 1, the hazard rate increases, meaning that the hazard risk for the first group $h_i(t \mid w_i)$ increases by $\exp\{\gamma^T(w_i - w_k)\}$ compared to the hazard risk for the second group $h_k(t \mid w_k)$. If HR < 1, the opposite is true. The HR is constant over time because it does not involve time t. This explains the proportionality assumption in the Cox PH regression that the hazard for one subject is proportional to any other subject at any time t.

Since the Cox PH regression model assumes that the hazard rate for each subject is time-independent, it uses the baseline values of the covariates. Specifically, the model does not take into consideration the progression of covariates γ_j over time. Therefore, the Cox PH is appropriate to study prognostic factors such as sex or randomized treatments whose values stay constant over time. This is unrealistic in practice to study biomarkers because longitudinal measured covariates such as biomarker serum bilirubin are expected to vary during follow-ups. This motivates the next model, the Time-Dependent Cox model, to investigate how the time-dependent covariates associate with patient outcome.

3.2 Time-Dependent Cox

Before explaining the mathematics of the Time-Dependent Cox model, it is essential to distinguish different types of time-dependent covariates. There are two common types: *exogenous* and *endogenous*. An exogenous covariate, sometimes called an external covariate, inherits a predictable pattern of change and is unaffected by the true failure time (Kalbfleisch and Prentice, 2011). In other words, the value of an exogenous covariate is known at any time t. For instance, age is an obvious exogenous covariate whose pattern of change is predictable. The age in the next five years is predicted easily if the current age is provided and its future path is independent of failure time. Other examples include environmental factors such as air pollution or predetermined covariates such as treatment strategies according to specific criteria.

On the other hand, endogenous covariates are time-dependent but measured on the subjects in the study. As a result, they may be affected by the health conditions of the subjects. In biomedical research, they are biomarkers or clinical parameters with repeated measurements, including serum bilirubin levels. For other diseases, endogenous covariates can be CD4 cell counts which shows the robustness of the immune system; CD4 cell counts are repeatedly measured to study the progression of human immunodeficiency virus (HIV). Generally, it is difficult to analyze data involving endogenous covariates because of their complicated and dynamic features. In particular, endogenous covariates generate themselves and are directly affected by true failure time, and often require the survival of the subjects.

Compared to the Cox PH regression, the Time-Dependent Cox model, sometimes referred to as the *Extended Cox*, allows more flexibility by taking

into account the progression of covariates with repeated measurements over time (Kleinbaum and Klein, 2010). The mathematical formula for its hazard function is given in Equation 3.3:

$$\begin{aligned} h_i(t \mid Y_i(t), w_i) &= h_0(t) \times \exp(\gamma^T w_i + \alpha y_i(t)), \qquad t > 0 \\ &= h_0(t) \times \exp(\gamma_1 w_{i1} + \gamma_2 w_{i2} + ... + \gamma_p w_{ip} + \alpha y_i(t)) \end{aligned} \tag{3.3}$$

where the interpretation of w_i and γ^T are the same as in the Cox PH model. In particular, w_i denotes the vector of baseline covariates while γ denotes the vector of corresponding regression coefficients. The new term $y_i(t)$ in Equation 3.3 captures the longitudinal measurement history $Y_i(t)$ of a time-dependent covariate y_i up to time t. The corresponding regression coefficient quantifies the association between this time-dependent covariate y_i and overall survival, after fixing other covariates with baseline values in the model. It also has a multiplicative effect: a one unit increase $y_i(t)$ results in a multiplicative change of $\exp(\alpha)$ in the risk for an event. The interpretations for HR of the Time-Dependent Cox model are similar to one of the Cox PH. The baseline risk function $h_0(t)$ is also unspecified for the same reasons as the Cox PH.

The major computational difference between the Cox PH and the Time-Dependent Cox models is that the Cox PH uses the short data format while the Time-Dependent Cox uses the Counting Process format. In the short format, the last time point of follow-ups and the corresponding event status are recorded. Each subject only has one row for baseline values of covariates which are recorded at the initiation of the study. On the other hand, in the Counting Process data format, there are multiple rows holding information for the same subject to divide the total follow-up time into small time intervals with

START and END points for time-dependent covariates. In addition, these rows also hold information for repeated measurements of longitudinal biomarkers on each time interval, which has event status 1 if the event of interest occurs and 0 otherwise.

Since the Time-Dependent Cox model assumes time-varying covariates to be exogenous with predetermined patterns of change, it does not take into consideration the measurement errors induced by biological variation from the subjects. In other words, the Time-Dependent Cox assumes time-varying covariates remain constant during the time intervals between doctor visits and are unaffected regardless of true failure time. In practice, the model is usually used to study association involving prognostic covariates such as age or environmental impacts. Unfortunately, it is not ideal to study cirrhosis and bilirubin with this model because the measurements of serum bilirubin are expected to fluctuate constantly[1] between visits and are strongly associated with the disease progression. Thus, HR estimates from the Time-Dependent Cox can potentially be less accurate. This motivates the Joint Model for Longitudinal and Survival Data (Joint Model), which takes into consideration the internal progression of bilirubin by accounting for the measurement errors.

[1]We can imagine that the bilirubin generation is similar to the heartbeat, which does not stay the same necessarily even in people with no health problems, though repeatedly measured minutes apart. In the case when someone just finished exercising, their heartbeat might be higher than usual. Alternatively, if the liver is not in a good condition, bilirubin level might be higher than usual.

3.3 Joint Model for Longitudinal and Time-To-Event Data

The Joint Model handles the special features of endogenous covariates when measuring the association between longitudinal biomarkers and overall survival (Rizopoulos, 2012). It accomplishes this by using the estimated *true* and *unobserved* value $m_i(t)$ of the longitudinal biomarkers at time t. The Joint Model consists of two components: a *survival submodel* and a *longitudinal submodel*. The longitudinal submodel estimates the true and unobserved value $m_i(t)$ of biomarkers by including measurement errors, which is then used in the survival submodel to correctly quantify the association of interest. These two submodels are modeled jointly, which explains the name Joint Model.

3.3.1 Survival submodel - Cox PH regression model

The hazard function for the survival submodel of is given in 3.4:

$$\begin{aligned} h_i(t \mid M_i(t), w_i) &= h_0(t) \times \exp(\gamma^T w_i + \alpha m_i(t)), \qquad t > 0 \\ &= h_0(t) \times \exp(\gamma_1 w_{i1} + \gamma_2 w_{i2} + ... + \gamma_p w_{ip} + \alpha m_i(t)) \end{aligned} \tag{3.4}$$

where w_i denotes the vector of baseline covariates and γ^T denotes the vector of corresponding regression coefficients. The new term $m_i(t)$ in Equation 3.4 is the true and unobserved value of longitudinal covariates at time t, as opposed to the observed value $y_i(t)$ in Equation 3.3, which is subject to measurement errors. The corresponding regression coefficient quantifies the effect of the underlying longitudinal endogenous covariate $m_i(t)$ on overall survival, after accounting for the impact of other covariates at baseline measurements in the model. Unlike the other two Cox models, the baseline risk function $h_0(t)$ is now specified explicitly. The possible distributions for the baseline hazard function $h_0(t)$ are:

- Weibull, log-normal, and gamma;
- Step-functions and linear-splines;
- B-splines approximation;
- Restricted cubic splines;
- Piecewise-constant and regression splines.

All these specifications of the baseline function $h_0(t)$ share a common objective: to allow more flexibility for non-linear patterns of the longitudinal trajectories. In this project, the baseline hazard function is a piecewise-constant spline with six knots placed at equally spaced percentiles of the observed event times.

3.3.2 Longitudinal submodel - Linear Mixed-Effects

As mentioned above, the Joint Model surpasses the other two Cox models by using the estimated true and unobserved longitudinal value $m_i(t)$ of the markers to take into account measurement errors. However, values $m_i(t)$ of covariates are never observed. The longitudinal submodel of the Joint Model estimates $m_i(t)$ by modeling the longitudinal history $M_i(t)$ of the covariates, which contains the true values of m_i at time t. The longitudinal submodel constructs $M_i(t)$ with the Linear Mixed-Effects model to describe the subject-specific time evolution of covariates.

In the Linear Mixed-Effects model (LME), the longitudinal outcomes are normally distributed:

$$\begin{aligned} y_i(t) &= m_i(t) + \epsilon_i(t) \\ m_i(t) &= x_i^T(t)\beta + z_i^T(t)b_i \end{aligned} \tag{3.5}$$

where $\epsilon_i(t) \overset{iid}{\sim} N(0, \sigma^2)$. The term $x_i^T(t)\beta$ describes the fixed effects and $z_i^T(t)b_i$ describes the random effects which are unique to each Subject i. Specifically, the design matrix $x_i^T(t)$ is for the vector of fixed effects β; the design matrix $z_i^T(t)$ is for the vector of random effects b_i; and the error terms $\epsilon_i(t)$ are time-dependent in Equation 3.5. LME accounts for measurement errors by the additive terms $\epsilon_i(t)$, which are assumed to be mutually independent and independent of random effects. The trajectory for each patient over time is dictated by time-independent random effects $b_i \sim N(0, D)$ where D is the variance-covariance matrix. The random effects b_i capture subject-specific deviations from the population mean estimates.

3.3.3 Transformation in LME

In the Linear Mixed-Effects of the longitudinal submodel, the response $m_i(t)$ is assumed to be normally distributed. Therefore, serum bilirubin is transformed with the Box-Cox transformation to get the normal distribution. However, the Box-Cox transformed serum bilirubin contains many very small values, which is incomparable to the values of the other biomarkers in the data. This results in a convergence issue since the estimate is unable to obtain. Thus, in this project, all the other longitudinal biomarkers are log-transformed so that their values are on a more comparable scale. Since the ultimate goal of the project is to make predictions on the association between serum bilirubin and overall survival, transformations do not have an impact on the estimates.

3.3.4 Missing Data - Last Observation Carried Forward

Both the Time-Dependent Cox model and the Joint Model use the repeated measurements of the covariates of interest. However, not all measurements during visits are available due to patients' failure to return for follow-ups. In the longitudinal submodel of the Joint Model, the LME assumes that the missing data mechanism is missing at random (MAR). It corresponds to the non-informative censoring when the reasons why subjects withdraw from the study are independent of their disease prognosis in the two Cox models. In other words, missing responses are unrelated to the observed ones. In this way, it is valid to consider observed data a random sample of complete data.

Theoretically, it is reasonable to ignore the missing values under the MAR mechanism. In practice, the LME model in the Joint Model works on the condition that no missingness is present in the data. In this project, instead of removing the observations with missingness, the *last observation carried forward* approach is employed to impute the missing data in the hope that any potential bias in parameter estimation is reduced and that the estimation is converged. This naive approach is expected to result in a more stable estimation despite non-monotone missingness mechanisms where patients' follow-ups are irregular.

3.4 Evaluation of Proportionality Assumption

The Cox PH models work under the assumption that the hazard ratio is proportional over time. In practice, this assumption is not always satisfied. Thus, it is crucial to assess the proportionality because a violation may affect the accuracy of the estimates and undermine the validity of the results. There are three common ways to assess the proportional hazards assumption (Kleinbaum and Klein, 2010):

- Graphical approach with survival curves
- Goodness-of-fit test with Schoenfeld residuals
- Time-covariates interaction terms

In this project, the goodness-of-fit (GOF) test is used to assess the appropriateness of the Cox PH models. The testing approach uses *Schoenfeld residuals*, which are defined for every subject who has an event, one for each of predictors in the model (Kleinbaum and Klein, 2010). For example, if a Cox PH model has n covariates, there are n Schoenfeld residuals for each subject that has an event. The Schoenfeld residual for suspiciously time-dependent covariate A that has the event is calculated:

$$\textit{Schoenfeld residual} = \textit{Observed } A \textit{ - Weighted Average } A \tag{3.6}$$

where the weighted average of A is the hazard of the other subjects still at risk at time t. The correlation between Schoenfeld residuals and failure times is then tested. If Schoenfeld residuals are correlated with failure times, the proportionality assumption is violated. Otherwise, the assumption is satisfied. In the GOF test, the null hypothesis states that there is not a correlation

between them. Thus, an insignificant p-value from this test indicates that the proportional hazard is met. Each of the p-values checks the assumption for its corresponding covariate under the assumption that the proportionality is satisfied for the other covariates in the model (Kleinbaum and Klein, 2010).

In the project, the proportionality is examined to assess the appropriateness of the Cox PH model and the survival submodel in the Joint Model. This test ensures the validity and accuracy of the estimates from these statistical models.

3.5 Variable Selection - A Backward Approach

For the two Cox model, it starts with the univariate analysis to get an overview of the data using functions from the package *gtsummary* (Sjoberg et al., 2020). Covariates are selected at the significance level $p_1 = 0.1$ from the univariate analysis to include in the backward elimination with the threshold of $p_2 = 0.05$. Finally, multivariate Cox models are fit with the selected statistically significant covariates from the variable selection. This is a common variable selection procedure in survival modeling. Examples of this approach can be found in the studies of overall survivals in colorectal carcinoma (Schmitz-Moormann et al., 1987), gastric cancer (Maruyama, 1987), advanced non-small-cell lung cancer (Paesmans et al., 1995), and supraglottic carcinoma (Nicolai et al., 1997).

As mentioned, backward elimination is used to select a subset of informative covariates from the pool of candidate covariates. The goal is to exclude any irrelevant covariates since the presence of extra variables increases bias in the estimation. This selection procedure can also detect any multi-collinearity in the model. As a result, model performance will be greatly improved by avoiding any covariate redundancy. Forward selection with the same significance level is also employed to verify that both backward and forward approaches arrive at the same subset of covariates.

In the project, backward elimination and forward selection are carried out with the functions *add1* and *drop1* supported by R, respectively. The parameter estimation is based on the maximum partial likelihood. Any two nested models with different subsets of covariates are compared by the χ^2 likelihood ratio test performed in the functions. Particularly, the likelihood

ratio test examines whether the inclusion or exclusion of covariates gives a statistically significant improvement to the model fit.

For the Joint Model, backward elimination is used as the variable selection approach for the LME in the longitudinal submodel, after the univariate analysis. The obtained subset of statistically significant covariates is fixed for the LME in the Joint Model, where serum bilirubin is the response. Given the fixed longitudinal submodel, backward elimination is employed again to select the statistically significant covariates for the Cox PH model in the survival submodel. Since the functions *add1* and *drop1* do not support the Joint Model, the function *anova* is used instead. The variable selection procedure is similar to the one for the Cox models: the function *anova* also uses the likelihood ratio test to compare two nested Joint Models (Kleinbaum and Klein, 2010). The summary of this approach is presented by steps below:

1. Univariate analysis with the threshold 0.1
2. Backward elimination with the threshold 0.05
3. Forward selection with the threshold 0.05 to compare results with (2)
4. Multivariate analysis with chosen covariates from (2)

3.6 Fitting and Estimation Methods in R

3.6.1 Cox PH and Time-Dependent Cox models

In the project, the R package for the Cox PH and the Time-Dependent Cox models is the *survival* package (Therneau and Lumley, 2015). In this package, parameters are estimated based on the maximum likelihood method which involves the two key functions in survival analysis: the survival and hazard functions. The survival function of the Cox PH is defined as:

$$S(t) = Pr(T^* > t) = \int_t^\infty p(s)ds \tag{3.7}$$

The survival function $S(t)$ describes the survival probability past a certain time point t (Kleinbaum and Klein, 2010). It is a non-increasing function as t ranges from 0 to ∞ with $t = 0$ corresponding to $S(t) = 1$. The function gives the probability that the random variable of failure times T^* exceeds the specified time t. T^* is assumed to be continuous and $p(s)$ is the probability density function. The survival function is closely related to the instantaneous risk function, which is defined as:

$$h_i(t \mid w_i) = lim_{dt \to 0} \frac{Pr(t \leq T^* \leq t + dt \mid T^* \geq t, w_i)}{dt} \tag{3.8}$$

Let T^* denote the random variable of the failure time under the study, $h(t)$ is a non-negative function that describes the instantaneous risk for occurrence of an event over the time interval $[t, t+dt)$, on the condition that the individual has survived up to time t. The relationship between survival function and

instantaneous risk function of the Cox PH is:

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t h(s)ds\right\} \tag{3.9}$$

where $H(t)$ is the cumulative hazard function that describes the accumulated risk up to time t.

The survival function $S(t)$ is used to construct the likelihood for survival data with censoring. Let $p(t;\theta)$ denote the probability density function of a random sample from a distribution function P parameterized by θ. If an event of interest is observed at time T_i, Subject i contributes to $p(T_i;\theta)$ in the likelihood function; otherwise, Subject i contributed to $S_i(T_i,\theta)$ when censoring happens. Thus, a log-likelihood function of the Cox PH accounting for both censored and uncensored observations is given by:

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{n} \delta_i \times \log\{p(T_i;\theta)\} + (1-\delta_i) \times \log\{S_i(T_i,\theta)\} \\ &= \sum_{i=1}^{n} \delta_i \times \log\{h_i(T_i;\theta)\} - \int_0^{T_i} h_i(s;\theta)ds\end{aligned} \tag{3.10}$$

where δ_i provides survival information in the random sample. In the estimation of parameters of interest, γ, regardless of the unspecified baseline hazard function $h_0(t)$, the partial log-likelihood function of the Cox PH model is:

$$p\ell(\gamma) = \sum_{i=1}^{n} \delta_i \left[\gamma^T w_i - \log\Big\{\sum_{T_j \geq T_i} \exp(\gamma^T w_i)\Big\}\right] \tag{3.11}$$

The estimators are calculated by taking the derivative of the partial log-

likelihood with respect to γ^T:

$$\frac{\partial p\ell(\gamma)}{\partial \gamma^T} = \sum_{i=1}^{n} \delta_i \left\{ w_i - \frac{\sum_{T_j \geq T_i} w_j \exp(\gamma^T w_j)}{\sum_{T_j \geq T_i} \exp(\gamma^T w_j)} \right\} = 0 \tag{3.12}$$

The solution to this equation is $\hat{\gamma}$, the estimated coefficient vector that is consistent and asymptotically normally distributed.

The survival function of the Time-Dependent Cox is given in 3.13:

$$S_i(t|Y_i(t)) = \exp\left\{ - \int_0^t h_i(s \mid Y_i(s)) ds \right\} \tag{3.13}$$

The partial likelihood function of the Time-Dependent Cox with respect to γ and α is:

$$\begin{aligned} p\ell(\gamma) = \sum_{i=1}^{n} \int_0^{\infty} \Bigg\{ & R_i(t) \times \exp\{\gamma^T w_i + \alpha^T y_i(t)\} \\ & - \log\Big[\sum_j R_i(t) \times \exp\{\gamma^T w_i + \alpha^T y_i(t)\} \Big] \Bigg\} dN_i(t) \end{aligned} \tag{3.14}$$

where $N_i(t)$ denotes the number of events for Subject i and $R_i(t)$ is an indicator variable. $R_i(t) = 1$ if Subject i is at risk at time t and 0 otherwise. The solution to Equation 3.14 are $\hat{\gamma}$ and $\hat{\alpha}$, the estimated coefficient vector and the estimate of the association of interest, respectively. Both of them are consistent and asymptotically normally distributed.

3.6.2 Joint Model

The Joint Model is fit using the *JM* package, which is dependent on the *nlme* and *survival* packages (Rizopoulos, 2010). The model consists of longitudinal and survival components. The longitudinal submodel, which is the Linear Mixed-Effects model, is fitted using package *nlme* version 3.1 (Pinheiro et al., 2007); the survival submodel, which is the Cox PH model, uses the package *survival* version 3.2-7 (Therneau and Lumley, 2015).

There have been many proposed estimation methods for Joint Models such as semi maximum likelihood, Bayesian estimation using Monte Carlo Markov Chain (MCMC) techniques, and a conditional score approach. In the *JM* package, parameter estimation is based on the full maximum likelihood approach. This traditional estimation method works under two assumptions. The first assumption is that the vector of time-independent random effects b_i underlies both longitudinal and survival processes (Rizopoulos, 2012). Since the vector random effects b_i accounts for the association between the longitudinal responses and the event outcomes, in addition to the correlation between the repeated measurements in the longitudinal process, it can be written as:

$$\begin{aligned} p(T_i, \delta_i, y_i \mid b_i; \theta) &= p(T_i, \delta_i \mid b_i; \theta) \times p(y_i \mid b_i; \theta) \\ p(y_i \mid b_i; \theta) &= \prod_j p(y_i(t_{ij}) \mid b_i; \theta) \end{aligned} \tag{3.15}$$

where $\theta = (\theta_t^\top, \theta_y^\top, \theta_b^\top)^\top$ denotes the full vector of parameters for the event outcome, longitudinal outcome and random-effects covariance matrix respectively. The other assumption is that the censoring mechanism and visiting process are dependent on the observed past longitudinal history but independent of un-

derlying, latent subject characteristics associated with prognosis (Rizopoulos, 2012). In other words, the censoring is assumed non-informative.

The log-likelihood corresponding to the joint distribution of observed outcomes T_i, δ_i, y_i for Subject i is:

$$\begin{aligned}\log p(T_i, \delta_i, y_i; \theta) &= \log \int p(T_i, \delta_i, y_i, b_i; \theta) db_i \\ &= \log \int p(T_i, \delta_i \mid b_i; \theta_t, \beta) \times \Big[\prod_j p(y_i(t_{ij}) \mid b_i; \theta_y)\Big] \times p(b_i; \theta_b) db_i\end{aligned} \tag{3.16}$$

where the conditional density for the survival part is $p(T_i, \delta_i \mid b_i; \theta_t, \beta)$ and the joint density for longitudinal responses with random effects is $\prod_j p(y_i(t_{ij}) \mid b_i; \theta_y) \times p(b_i; \theta_b)$.

The log-likelihood function $\ell(\theta) = \sum_i \log p(T_i, \delta_i, y_i; \theta)$ is optimized using a hybrid optimization of *Expectation-Maximization* (EM) and quasi-Newton algorithms. Particularly, the estimation starts with the EM algorithm with a fixed number of iterations and if convergence is not obtained, a *quasi-Newton* algorithm is used (Rizopoulos, 2012). The initial values for the optimization are from the fitted linear mixed-effects and survival submodels.

Chapter 4

Data Description

4.1 The Data

The dataset is from a Primary Biliary Cirrhosis (PBC) clinical trial with 312 patients conducted by the Mayo Clinic for a 10-year period from 1975 to 1984 (Murtaugh et al., 1994). Among these patients, 154 of them were randomly put in the placebo group and the rest were in the treatment group. By the end of the follow-up, which was extended to year 1988, 140 of the patients died, 29 had received orthotopic liver transplantation and 143 were still alive. After accounting for death and censoring, there were 1945 patient visits, with repeated measurements capturing the progression of PBC longitudinal biomarkers such as serum bilirubin, albumin, prothrombin time, the presence of spiders - blood vessel malformation in the skin, etc. These biomarkers were measured at specific visits at six months, one year, and annually thereafter. In addition to these biomarkers, there were baseline covariates such as age, drug, and gender. Below is the summary of the PBC data with a brief explanation of biomedical terminology (Mayo Clinic, 2021):

Categorical covariates:

- drug : placebo or D-penicillin groups
- sex : male or female groups
- ascites : swelling of abdomen from fluid accumulation (Yes/No)
- edema : swelling of leg, ankle and feet caused by excess fluid trapped in body's tissues (3 levels: No edema/Edema no diuretic[2]/Edema despite diuretic)
- hepatomegaly: enlarged liver (Yes/No)
- spiders : blood vessel malformations in the skin (Yes/No)
- histologic: histologic stage for chronic liver disease based on liver biopsy of the disease (4 levels: 1-4)
- status: alive/transplanted or deceased

[2]A type of drug that helps kidneys make more urine to help the body get rid of extra fluid and salt

Continuous covariates:

- age (years): time since the registration
- albumin (g/dl): blood protein
- alkaline (U/l): alkaline phosphatase (ALP) is an enzyme in the liver; higher levels of ALP may indicate liver damage
- SGOT (U/ml): serum glutamic oxaloacetic transaminase (AST) is an enzyme in the liver released into blood when the liver is damaged; elevated blood SGOT levels may indicate liver damage
- platelets (per cubic ml/1000): colorless blood cells that help blood clot; decreased white blood cells and platelets can be a sign of cirrhosis
- prothrombin (seconds): prothrombin time is the time it takes blood to clot; increased PT may indicate liver damage
- year: number of years between enrollment and this visit date, remaining values on the line of data refer to this visit
- years: number of years between enrollment and the earlier of death, transplantation, or study analysis time
- bilirubin (mg/dl): substance produced during the normal breakdown of red blood cells and passes through the liver; elevated levels of bilirubin cause yellowing of the skin and may indicate liver damage

The primary objective of this randomized placebo controlled trial was to investigate the treatment effect of the drug D-penicillin on overall survival of the PBC patients. In this project, the main goal is to study how serum bilirubin is in association with overall survival of PBC patients. In the study, though there were 312 patients followed up for 10 years with specified visits at six months, one year and annually thereafter, there were only 1945 observations in the dataset due to censoring and termination.

4.2 Data Exploration

Table 4.1 shows the statistics summary of the PBC clinical data. Serum bilirubin is the only biomarker discussed in this section because the project focuses on the association between bilirubin and overall survival. Please refer to Appendix A for more details of statistics summary of other biomarkers.

In Table 4.1, 312 patients were randomized almost equally with 51% of the patients in the treatment group and 49% in the placebo group. Among 312 patients in the study, 88% of them were females and only 12% were males. At the end of the study, 45% of the patients died; 55% of the patients were alive or liver-transplanted. The median number of years between registration and the earlier of death, transplantation, or study time was 6.3 years; the median age was 50.

In general, patients were equally randomized among the placebo and the treatment groups between the alive/transplanted and the deceased groups (Table 4.1). The percentage of male patients in the deceased group outnumbered the alive group. Overall, the statistics summary suggests that the level of serum bilirubin was lower in the alive/transplanted group. The median age for the alive/transplanted group was lower than for the deceased group, while the study time for the alive/transplanted group was more than doubled for the deceased group.

Statistics Summary			
Characteristics	**N = 312**	**Alive = 172**	**Deceased = 140**
drug: placebo	154 (49%)	85 (49%)	69 (49%)
D-penicillin	158 (51%)	87 (51%)	71 (51%)
sex: male	36 (12%)	10 (5.8%)	26 (19%)
female	276 (88%)	162 (94.2%)	114 (81%)
years	6.3 (3.7, 8.9)	7.8 (5.7, 9.9)	3.7 (2.1, 6.7)
age	50 (42, 57)	47 (40, 55)	53 (46, 61)
bilirubin	1.4 (0.8, 3.4)	1.0 (0.6, 1.8)	3.0 (1.3, 6.4)

Table 4.1: Statistics Summary of PBC Clinical Data

4.3 Data Visualization

There are many biomarkers in the data set but the visualization will mainly represent the exploratory insights of biomarker bilirubin because of the primary goal of the project. For further details of the data exploration with other biomarkers, please refer to Appendix A: Exploratory Analysis.

Histogram

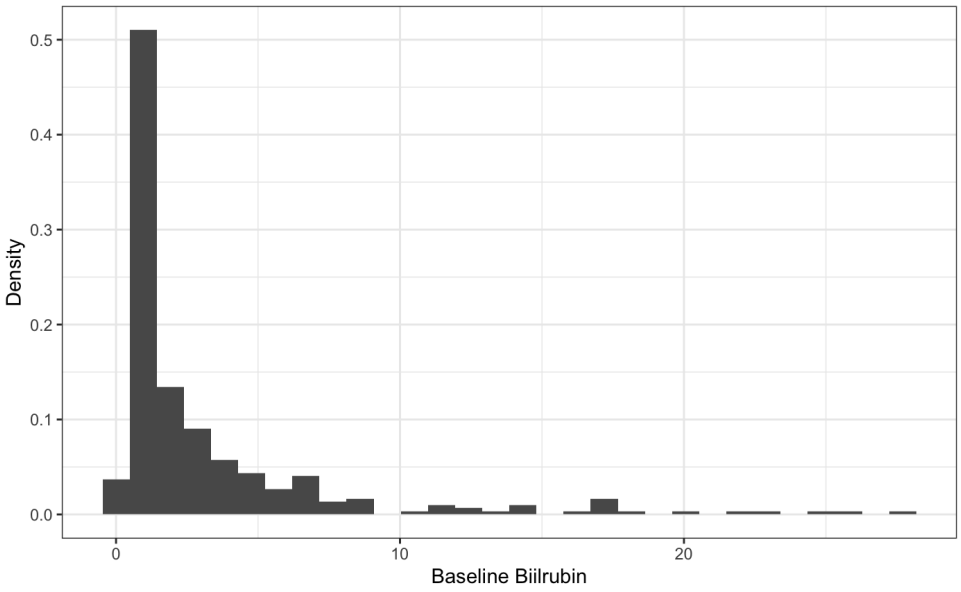

Figure 4.1: Baseline bilirubin measured at the study registration

Figure 4.1 illustrates the histogram distribution of the baseline values of serum bilirubin. In Figure 4.1, the distribution is right-skewed, suggesting that the liver conditions varied widely among 312 patients.

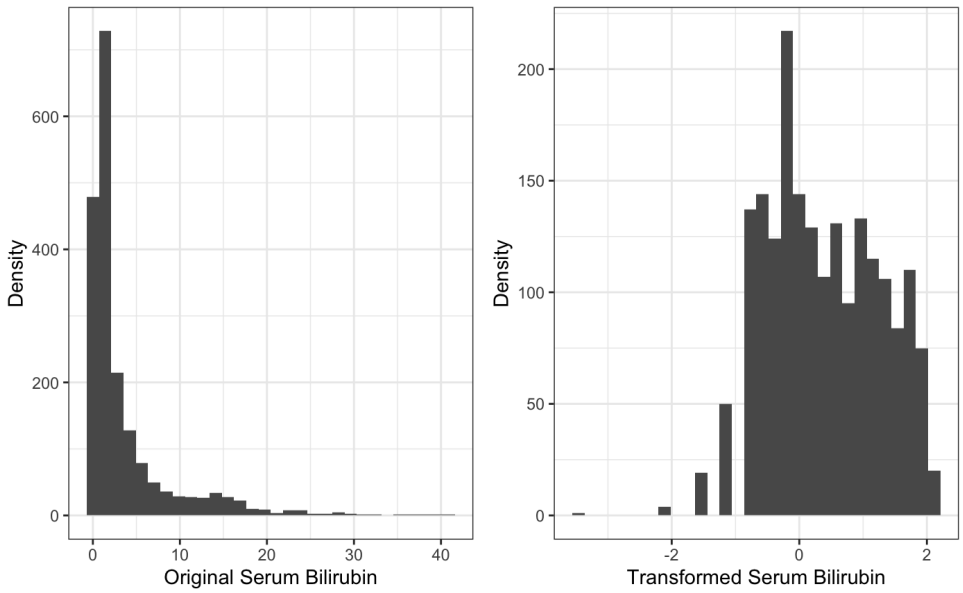

Figure 4.2: Bilirubin before and after the Box-Cox transformation

Figure 4.2 shows that the distribution of longitudinal values of bilirubin is right-skewed. Since these longitudinal measurements were used as the response in the Linear Mixed-Effects model of the Joint Model, the Box-Cox transformation is performed to satisfy the normality condition. The distributions before and after the Box-Cox transformation are represented in Figure 4.2. For more details on Box-Cox transformation, please refer to Appendix C: Mathematical Formulae.

Boxplot

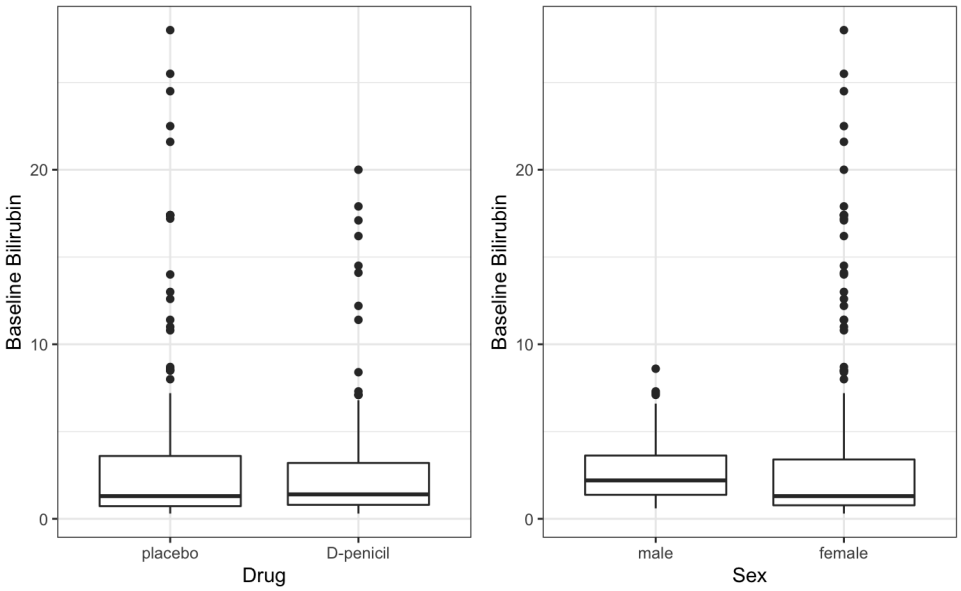

Figure 4.3: Baseline bilirubin between treatment groups and sex groups

Figure 4.3 shows the boxplots of the baseline bilirubin between treatment groups and sex groups. In Figure 4.3, the placebo and the treatment group have many outliers although the average baseline bilirubin levels are roughly similar. In the second boxplot, females have more outliers and lower bilirubin than males.

Spaghetti Plot

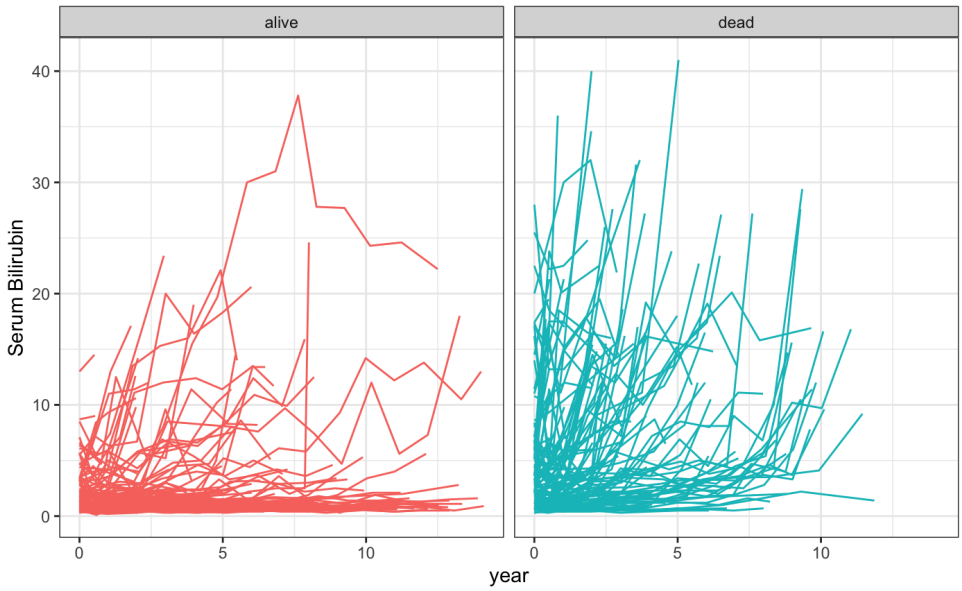

Figure 4.4: Spaghetti plot of bilirubin by outcomes. Each path represents different trajectories of the bilirubin for each patient. The y-axis is the longitudinal measurements of bilirubin recorded at time t and the x-axis is the time point t

Figure 4.4 shows the spaghetti plot of the serum bilirubin progression overtime between the alive/transplanted and deceased patients. In Figure 4.4, the alive/transplanted patients were more likely to have a lower level of bilirubin than the deceased patients.

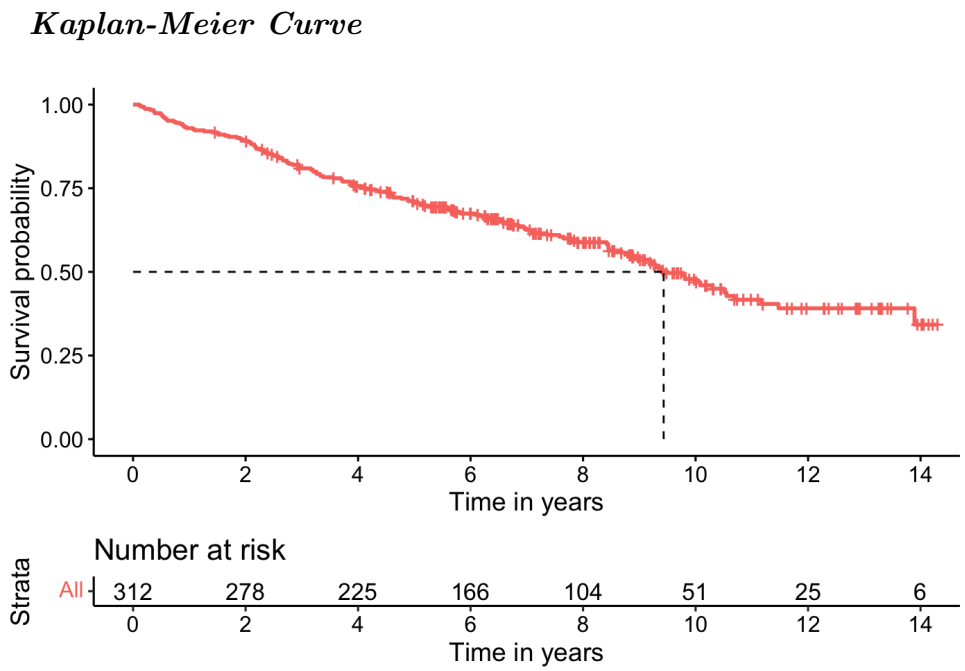

Figure 4.5: Overall survival probability. The survival probability starts at 1 when $t = 0$ and decreases over time. The tick marks illustrate censoring at specific time points. The time at which the survival probability reaches 0.50 is the median survival time

Kaplan-Meier survival curve is used to visualize the overall survival probability of 312 patients over the study time. In Figure 4.5, the median survival time was around 9.4 years. At the beginning of the study, there were 312 patients at risk. After 6 years since the initiation of the study, there were 166 patients at risk. For more details on Kaplan-Meier survival curve and log-rank test, please refer to Appendix C: Mathematical Formulae.

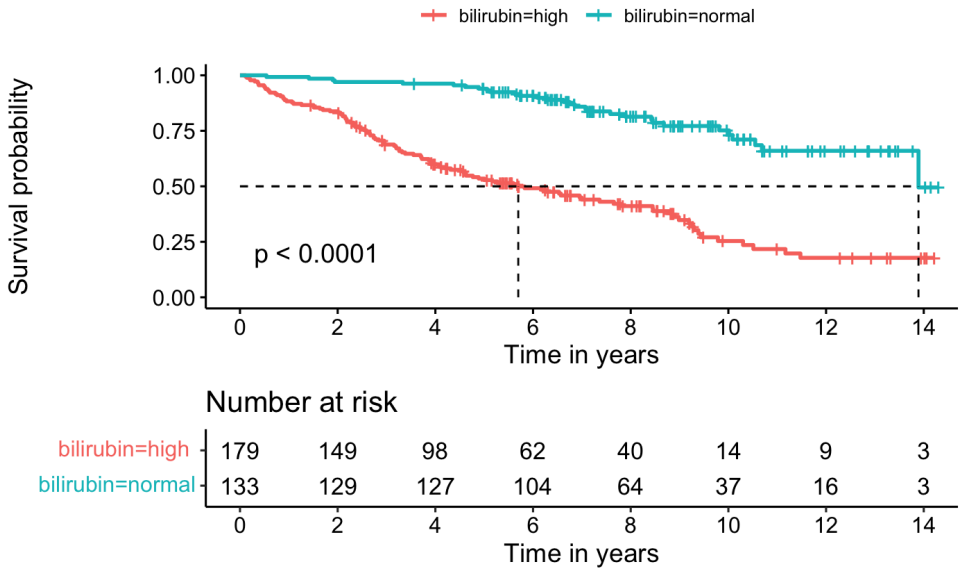

Figure 4.6: Survival probability with respect to baseline bilirubin

Figure 4.6 shows the survival probabilities with respect to the baseline bilirubin for which the clinical cut-off 1.2 mg/dl is used to classify the high and normal levels of baseline bilirubin. The statistically significant log-rank p-value indicates that there is a difference in the survival probabilities between the two groups. The median survival time was around 6 years for the high bilirubin and 14 years for the normal bilirubin.

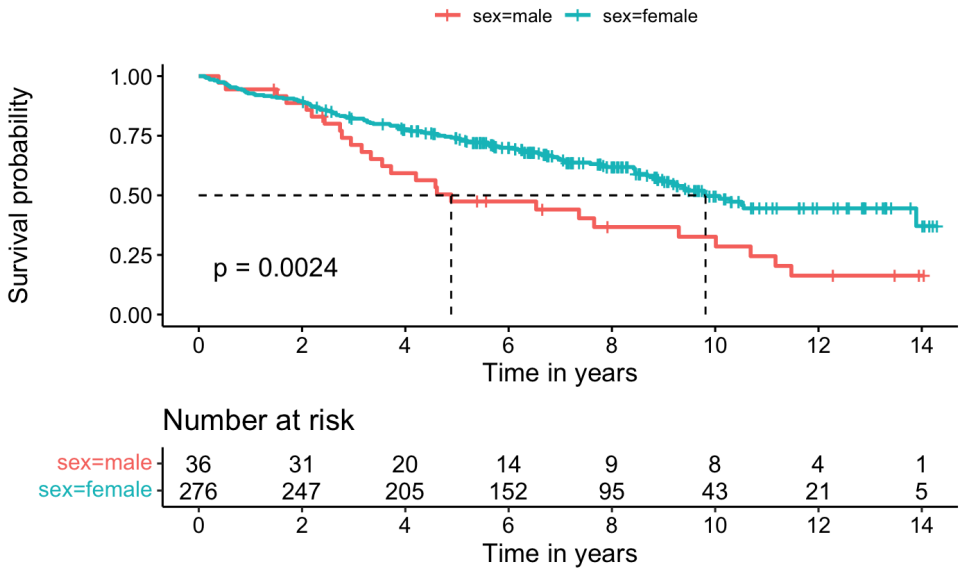

Figure 4.7: Survival probability in different sex groups

Figure 4.7 explores the survival of the patients in different sex groups. In general, females had a higher probability survival than males. The log-rank p-value of 0.0024 confirms that sex is statistically significant. The median survival time was around 5 years for male patients while around 10 years for female patients. The number of patients at risk for females was more than 10 times larger than the one for males after 6 years since the registration. This huge difference between the two groups can be partly explained by the distribution of patients between the groups. From the beginning of the study, only 12% of the total patients were males while the rest were females (Table 4.1).

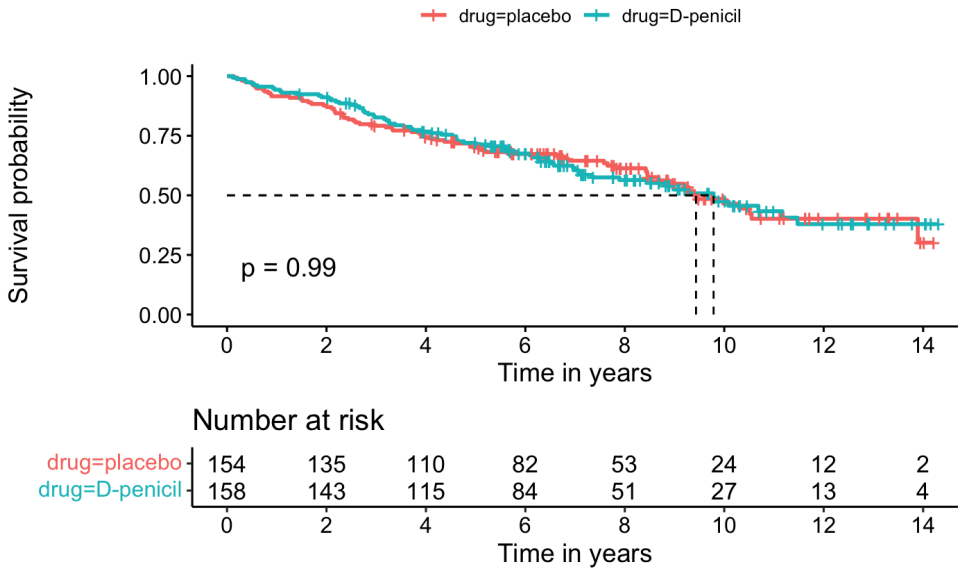

Figure 4.8: Survival probability in different treatment groups

Figure 4.8 represents the survival of the patients in different treatment groups. Interestingly, the survival probabilities in the placebo and the treatment groups were roughly similar. Drug D-penicillin is not significantly significant from the log-rank test. The median survival times were around 9.5 and 9.7 years for the placebo group and the treatment group, respectively.

Chapter 5

Statistical Results

5.1 Univariate Analysis

Clinical Factors

Table 5.1 shows the results for clinical factors such as *Age*, *Drug*, and *Sex* from the univariate analysis for the Cox PH and the Time-Dependent Cox models. The Cox models give the same results for the hazard ratios, the 95% confidence intervals and the p-values for the clinical factors. *Age* and *Sex* are both statistically significant at the threshold of 0.10. In contrast, *Drug* is statistically insignificant with a p-value larger than 0.9. Consequently, *Age* and *Sex* are included in the variable selection for the multivariate analysis.

Univariate Analysis - Clinical Factors			
Clinical Factor	**Hazard Ratio**	**95% CI**	**p-value**
Drug: placebo	*ref*		
D-penicillin	1.00	(0.72 - 1.39)	>0.9
Age	1.05	(1.03 - 1.06)	<0.001
Sex: male	*ref*		
female	0.52	(0.34 - 0.80)	0.005

Table 5.1: Results for Clinical Factors from Univariate Analysis

Longitudinal Biomarkers

Table 5.2 shows the results for biomarkers other than bilirubin from the univariate analysis for the Cox PH and the Time-Dependent Cox models. Since the focus lies in the association between longitudinal measurements of serum bilirubin and overall survival, the baseline measurements for the other biomarkers in the models are used in the Time-Dependent Cox. Thus, the two Cox models only differ in the the information of bilirubin being considered. Specifically, the Cox PH model uses the baseline measurements of bilirubin while the Time-Dependent Cox uses the current measurements of bilirubin by taking into account its longitudinal history.

The estimation from the univariate analysis for all biomarkers in the Cox PH and Time-Dependent Cox models are exactly the same (Table 5.2). In other words, the two Cox models have the same output for the hazard ratios, the 95% confidence intervals and the p-values. All the biomarkers are statistically significant at the threshold of 0.10. Thus, they are included in the variable selection for the multivariate analysis.

Univariate Analysis - Longitudinal Biomarkers			
Biomarker	**Hazard Ratio**	**95% CI**	**p-value**
Albumin	0.19	(0.13 – 0.28)	<0.001
Alkaline	1.00	(1.00 – 1.00)	0.094
SGOT	1.01	(1.00 – 1.01)	<0.001
Platelets	1.00	(0.99 – 1.00)	<0.001
Prothrombin	2.12	(1.81 – 2.48)	<0.001
Ascites: No	*ref*		<0.001
Yes	7.58	(4.78 - 12.0)	
Hepatomegaly: No	*ref*		<0.001
Yes	3.06	(2.14 - 4.38)	
Spiders: No	*ref*		<0.001
Yes	2.42	(1.72 - 3.42)	
Edema: No edema	*ref*		<0.001
Edema no diuretics	1.63	(1.04 - 2.55)	
Edema diuretics	10.9	(6.61 - 18.0)	
Histologic: 1	*ref*		<0.001
2	6.39	(0.86 - 47.5)	
3	9.66	(1.33 - 70.1)	
4	24.0	(3.33 - 174)	

Table 5.2: Results for Biomarkers from Univariate Analysis

Serum Bilirubin

Table 5.3 shows the results from the univariate analysis for the Cox PH, the Time-Dependent Cox, and the Joint Model with bilirubin as the only covariate. The Cox PH and the Time-Dependent Cox models have the same point estimates for the hazard ratios and p-values. However, the 95% confidence interval for the Time-Dependent Cox is smaller than for the Cox PH model. This is reasonable because in the Time-Dependent Cox, repeated measurements of bilirubin are used; therefore, more information on bilirubin level reduces the uncertainty of the estimated hazard ratio.

The survival submodel of the Joint Model is handled by the Cox PH model with bilirubin measured from the longitudinal submodel as the only explanatory covariate. The longitudinal submodel includes the set of statistically significant covariates in the LME model from the variable selection. In the LME model, repeated measurements of bilirubin are used as the response variable while the baseline measurements for other biomarkers are used as the explanatory covariates.

The interpretation of the hazard ratios for the Cox PH is that a one unit increase in the *baseline bilirubin* is associated with a 16% increase in the death risk. For the Time-Dependent Cox, a one unit increase in the *longitudinal bilirubin* is associated with a 16% increase in the death risk. For the Joint Model, a one unit increase in the longitudinal bilirubin is associated with an 83% increase in the death risk, after accounting for *measurement errors*.

Univariate Analysis - Serum Bilirubin			
Model	**Hazard Ratio**	**95% CI**	**p-value**
Model 1: Cox PH	1.16	(1.13 - 1.19)	<0.001
Model 2: Time-Dependent Cox	1.16	(1.14 - 1.18)	<0.001
Model 3: Joint Model	1.83	(1.66 - 2.02)	<0.0001

Table 5.3: Results for Serum Bilirubin from Univariate Analysis

5.2 Model Adjustment

After the variable selection, the Cox PH, Time-Dependent Cox, and Joint Model have different adjusting covariates corresponding to Model 1, 2, and 3:

Model 1: bilirubin, albumin, age, edema, histologic, SGOT, and PT[3]

Model 2: bilirubin, albumin, age, edema, histologic

Model 3: bilirubin, albumin, age, edema

At the threshold level of 0.05, the sets of statistically significant covariates for the three models from the variable selection are different. In order to examine the consistency of the results, despite different model adjustments, the multivariate analysis is re-conducted with the union of those sets of covariates for each of all three models. For example, since *bilirubin*, *albumin*, *age*, *edema*, *histologic*, *SGOT*, and *prothrombin* are statistically significant for the Cox PH model, a multivariate analysis including these covariates is carried out for the Time-Dependent Cox model and the Joint Model as well. This approach is applied to the other two sets of covariates.

Results of the comparison among the estimated hazard ratios from three models with three unions of those set of covariates are available in Appendix B: Supplementary Analysis. The comparison shows that the results are fairly consistent since the differences among estimated hazard ratios for each model with three different sets of covariates are negligible. The statistically insignificant covariates for each model from the variable selection do not add much information to the investigated association.

[3]Prothrombin

5.3 Multivariate Analysis

Table 5.4 shows the results from the multivariate analysis for the Cox PH, the Time-Dependent Cox, and the Joint Model with bilirubin and other covariates selected from the variable selection corresponding to each model. Bilirubin is statistically significant in three models. The hazard ratios for serum bilirubin are 1.11, 1.20, and 1.82 for the Cox PH, the Time-Dependent Cox, and the Joint Model, respectively. Compared to the univariate analysis, the hazard ratio decreases by 0.5 for the Cox PH model, and increases by 0.4 for the Time-Dependent Cox. For the Joint Model, it is roughly similar, increasing by 0.1 (Table 5.4).

The interpretation of the hazard ratio after accounting for other covariates in each model is straightforward: For the Cox PH model, a one unit increase in the *baseline bilirubin* is associated with a 1.11-fold increase in the death risk; for the Time-Dependent Cox, a one unit increase in the *longitudinal bilirubin* is associated with a 1.20-fold increase in the death risk; for the Joint Model, a one unit increase in the longitudinal bilirubin is associated with a 1.82-fold increase in the death risk with *measurement errors* taken into consideration. In other words, for one mg/dl increase in the level of serum bilirubin, the death risk increases by 11%, 20%, and 82% for the Cox PH model, the Time-Dependent Cox, and the Joint Model, respectively.

Compared to the univariate analysis, 95% confidence intervals from the multivariate analysis are smaller, indicating that the uncertainty about the estimates is reduced. It is reasonable because more statistically significant covariates are expected to add more information to the bilirubin levels and the investigated association generally.

Multivariate Analysis - Serum Bilirubin			
Model	**Hazard Ratio**	**95% CI**	**p-value**
Model 1: Cox PH	1.11	(1.06 - 1.15)	<0.001
Model 2: Time-Dependent Cox	1.20	(1.17 - 1.22)	<0.001
Model 3: Joint Model	1.82	(1.64 - 2.03)	<0.0001

Table 5.4: Results for Serum Bilirubin from Multivariate Analysis

5.4 Proportionality Assumption

Cox PH Model

Table 5.5 shows that from the goodness-of-fit test, the hazards are not proportional for bilirubin over time in the multivariate analysis of the Cox PH model. This is within expectation because bilirubin dynamically generates itself for each subject over time. Thus, bilirubin is expected to have time-varying effect on the hazard of subjects, which is captured in the Time-Dependent Cox and the Joint Model. The results from the multivariate analysis of the Cox PH model are still recorded for comparison, despite a violation of the proportionality assumption.

Proportionality Assumption in Cox PH Model			
Characteristics	**Chisq**	**df**	**p-value**
Bilirubin	6.305	1	0.012
Albumin	3.043	1	0.081
Age	0.195	1	0.659
Edema	4.208	2	0.122
Histologic	6.603	3	0.086
SGOT	0.405	1	0.525
Prothrombin	1.482	1	0.223
GLOBAL	20.873	10	0.022

Table 5.5: Evaluation of Proportionality Assumption in Cox PH Model

Survival Submodel of Joint Model

At the significance level of 0.05, the survival submodel of the Joint Model satisfies the proportionality assumption (Table 5.6). Thus, the estimates from the Joint Model are likely to be valid and reliable.

Proportionality in Survival Submodel of Joint Model			
Characteristics	**Chisq**	**df**	**p-value**
Albumin	2.279	1	0.131
Age	0.204	1	0.651
Edema	4.848	2	0.089
GLOBAL	6.394	4	0.172

Table 5.6: Evaluation of Proportionality Assumption for Survival Submodel

5.5 Diagnostic Plots for Linear Mixed-Effects

In Figure 5.1, the Q-Q plot of the residuals from the LME. In the Q-Q plot, the majority of the points match the straight line. There seem to be no serious outliers or influential observations that violate the normality assumption. The distribution of the residuals has a bell-shaped curve. Thus, the normality condition is likely satisfied. The scatterplot of the residuals and the fitted values from the LME shows that there is no specific pattern in the graph and the data points spread above and below the zero line. In general, the diagnostic plots indicate that the estimate from LME is likely accurate since the underlying conditions are properly met.

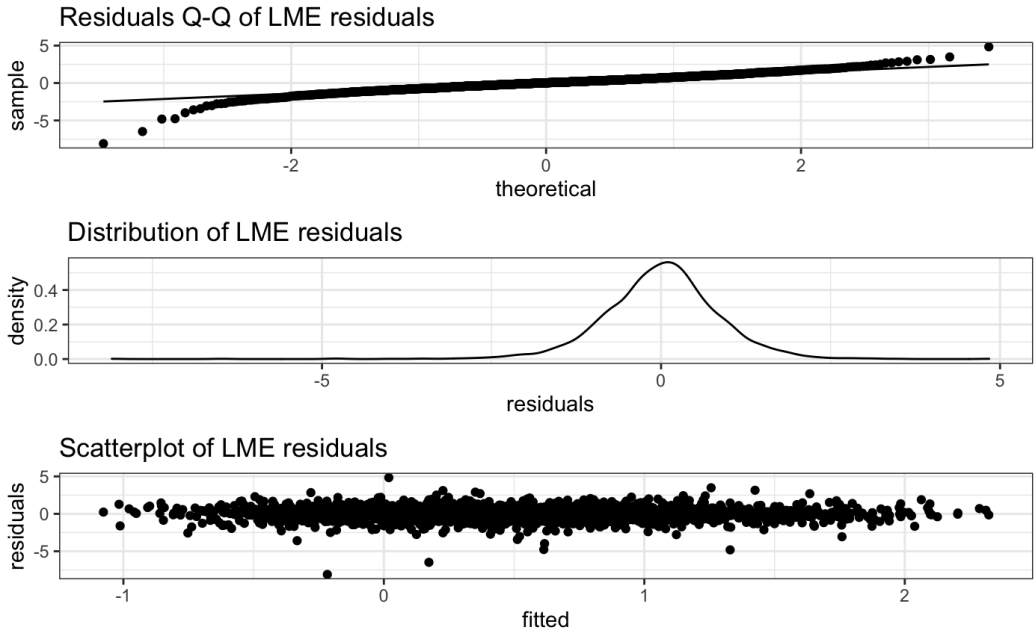

Figure 5.1: Diagnostic plots for Linear Mixed-Effects Model

Chapter 6

Conclusion

6.1 Summary of Findings

The association between serum bilirubin and survival of Primary Biliary Cirrhosis is investigated using different levels of the information of covariates. The Cox PH model uses the baseline measurements of all covariates. The Time-Dependent Cox model accounts for the progression of bilirubin while fixing the measurements at the baseline level for the other covariates. For the Joint Model, the Cox PH and LME in the two submodels uses the adjusting covariates from the variable selection. The LME model in the longitudinal submodel uses the longitudinal measurement for bilirubin and baseline measurements for other covariates.

In Table 6.1, for the Cox PH model, one unit increase in the *baseline bilirubin* is associated with a 1.11-fold increase in the death risk. For the Time-Dependent Cox, one unit increase in the *longitudinal bilirubin* is associated with a 1.20-fold increase in the death risk. For the Joint Model, a one unit increase in the longitudinal bilirubin is associated with a 1.82-fold increase in

the death risk with the *measurement errors* taken into consideration. In other words, one unit increase in the level of serum bilirubin increases the death risk by 11%, 20%, and 82% for the Cox PH model, the Time-Dependent Cox, and the Joint Model respectively.

Multivariate Analysis - Serum Bilirubin			
Model	**Hazard Ratio**	**95% CI**	**p-value**
Model 1: Cox PH	1.11	(1.06 - 1.15)	<0.001
Model 2: Time-Dependent Cox	1.20	(1.17 - 1.22)	<0.001
Model 3: Joint Model	1.82	(1.64 - 2.03)	<0.0001

Table 6.1: Results for Serum Bilirubin from Multivariate Analysis

6.2 Discussion

From the summary of findings, with the progression of bilirubin and measurement errors taken into consideration, the estimated hazard ratio from the Joint Model is much larger than the ones from the Cox PH and Time-Dependent Cox models. It seems that there is a non-negligible difference in the parameter estimates after accounting for both progression of bilirubin and measurement errors in the longitudinal history.

Despite the comparison of the hazard ratios from three models, it is essential to be aware that the adjusting covariates for the Cox PH, the Time-Dependent Cox, and the Joint Model from the variable selection are different, corresponding to sets 1, 2, and 3 respectively:

Set 1: bilirubin, albumin, age, edema, histologic, SGOT, and prothrombin

Set 2: bilirubin, albumin, age, edema, histologic

Set 3: bilirubin, albumin, age, edema

In the Cox PH model, SGOT and prothrombin are statistically significant at the baseline measurements. SGOT, or Aspartate Transaminase (AST), is an enzyme normally present in blood at low levels; a high level of SGOT may indicate liver damage (Mayo Clinic, 2019). Prothrombin is a protein produced by the liver. In the PBC dataset, covariate prothrombin is Prothrombin Time (PT) which is the time it takes blood to clot; increased PT may indicate liver damage (Mayo Clinic, 2019). High levels of both SGOT and prothrombin indicate potential liver damage, while elevated levels of bilirubin are also associated with liver cirrhosis. Therefore, baseline measurements of SGOT and prothrombin may not add more information to the generating process of bilirubin and therefore, the investigated association between longitudinal bilirubin

and overall survival. This might explain why SGOT and prothrombin are not statistically significant in the Time-Dependent Cox model.

Though baseline SGOT and prothrombin are not statistically significant in the survival submodel of the Joint Model, they are statistically significant in the longitudinal submodel LME where longitudinal bilirubin is the response. Based on the findings, it can be inferred that even though SGOT and prothrombin with baseline measurements do not add valuable information to the studied association involving longitudinal bilirubin, their baseline measurements are informative on how bilirubin generates itself over time, given the measurement errors.

A multivariate analysis is re-conducted with the union of those sets of covariates for each of all three models (see Appendix B: Supplementary Analysis for more details). Though the estimates for each of the three statistical models with the same set of adjusting covariates are not exactly the same, the estimated hazard ratios for each model are not radically different across different covariate sets. In fact, the Time-Dependent Cox model has fairly consistent estimates (Table 6.2).

Hazard Ratio - Bilirubin Analysis			
Model	**Set 1**	**Set 2**	**Set 3**
Model 1: Cox PH	**1.11**	1.15	1.14
Model 2: Time-Dependent Cox	1.20	**1.20**	1.19
Model 3: Joint Model	1.81	1.84	**1.82**

Table 6.2: Comparisons of Hazard Ratios for Bilirubin

The PBC data is from a clinical trial conducted by Mayo Clinic over 10 years. Therefore, it is not possible to evaluate the model performances with the true hazard ratio because it is not available. Nevertheless, there have been simulation studies to evaluate the model performances of Cox PH, Time-Dependent Cox, and Joint Model in literature. The results showed that Joint Models lead to an estimate with a smaller SE and less bias (Ibrahim et al., 2010). In other simulation studies, Arisido showed the robustness of Joint Models in providing a more accurate and precise estimate of the hazard ratio (Arisido et al., 2019). Thus, in this PBC study, the Joint Model is expected to reduce bias in the estimate compared to the other two Cox models.

Even though both the Time-Dependent Cox model and the Joint Model account for the progression of bilirubin, there are notable differences in the assumptions behind these two models. The Time-Dependent Cox assumes bilirubin to remain constant during the time interval between the visits. In practice, this assumption is not realistic, because bilirubin generates itself and continually fluctuates. Graphically, the Time-Dependent Cox has a step function for a longitudinal trajectory. This may result in very biased estimates since the Time-Dependent Cox might overestimate or underestimate the true and unobserved value of the biomarkers. On the other hand, the Joint Model estimates the true value of bilirubin by including measurement errors in addition to its longitudinal progression. In Figure 6.1, the Joint Model smooths the longitudinal trajectory, thereby reducing potential bias and providing a more reliable estimate.

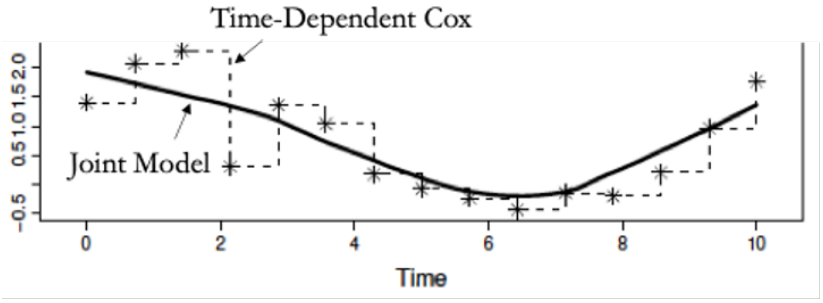

Figure 6.1: Measurements in Time-Dependent Cox and Joint Model
Source: Joint Models (Rizopoulos, 2012)

Though theoretically, the Joint Model appears as the most optimal approach because it increases estimation accuracy by addressing the limitations of the two Cox models, it also has disadvantages as a trade-off. The Joint Model has many parameters to estimate from the survival and longitudinal submodels. Consequently, it requires a large sample size to avoid convergence issues in the estimation process. Moreover, it is very computationally expensive to implement the estimation for the Joint Model because of its model complexity. When the sample size and computational resources are not fully met, the resulting bias in the estimation from the Joint Model can surpass one from other statistical models for censored data. Thus, when the sample size is not large enough or the computation is not available, it is recommended to use the Time-Dependent Cox model with caution. The trade-off is that hazard ratios estimates can be potentially less accurate, depending on the data.

As mentioned earlier, there are different possible specifications for the baseline hazard function $h_0(t)$ for the Joint Model with the same objective to capture the non-linearity in the longitudinal trajectories. In this project, the baseline hazard function is the piecewise-constant spline with six knots at equally spaced percentiles of the observed event times. Theoretically, the

increasing number of internal knots allows more flexibility and thus captures various shapes of the $h_0(t)$. However, given the computational demand of the Joint Model, the idea of increasing the number of knots to introduce more flexibility is infeasible. Similarly, though theoretically it is possible to have interaction or polynomial terms in the Cox PH model in survival submodel or the LME in the longitudinal submodel, these options are not feasible for the same reason.

6.3 Future Work

As mentioned earlier, it is not possible to evaluate the model performances of the Cox PH, the Time-Dependent Cox, and the Joint Model with the true hazard ratio because it is not available. One possible direction to expand the work is to use simulation studies with the similar setting to the PBC data. Particularly, it is possible to simulate data with predetermined censoring rate and hazard ratio. In that way, we can use simulated data to evaluate the model performances by measuring the amount of bias present for each model, and therefore draw conclusions on the PBC data analyses.

To avoid mis-specifying the distribution of the survival time, the baseline hazard functions in the Cox PH and the Time-Dependent Cox models are unspecified. However, it is possible to give the baseline function a specific form such as Weibull, exponential, or lognormal distribution though generally, even if the correct distribution can be parametrically specified, the Cox models typically give results approximately close to those obtained from the parametric model for censored data (Kleinbaum and Klein, 2010). For the Joint Model, the baseline hazard function is specified as the piecewise-constant spline with six knots at equally spaced percentiles of the observed event times. However, there are other options with the same purpose to introduce more flexibility for the longitudinal trajectories of the markers such as linear splines or restricted cubic splines.

For variable selection, the χ^2 likelihood ratio test is used for nested models in each of the three statistical approaches. Other traditional variable selection criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be extended to survival analysis, where the penalty term is

defined in terms of the uncensored events, instead of the number of observations (Fan et al., 2005). Elastic net or LASSO regularization is not appropriate for this PBC clinical data set because this is not a high dimensional setting where elastic net or LASSO is used to obtain a parsimonious model. Usually, elastic net or LASSO regularization is utilized to penalize the redundant covariates in statistical genetics where the genomic data is of high-dimension.

In this project, the missing data is imputed using the last value carried forward, which is a common approach when the data is missing at random. Though this assumption fits in with the PBC data in the project, it is also possible to try multiple imputation methods for missing data in the longitudinal measurements. The procedure is to impute the missing data multiple times by injecting appropriate random variability, perform desired data analysis on the complete imputed data, and average the parameter estimates across samples to obtain a single point estimate. Multiple imputation methods are more complicated and sophisticated, but the computation is often expensive and requires a large sample size.

For simplicity, the alive and transplanted patients are combined into one group. The future direction for this project is to use competing risk models such as the cause-specific hazard model or the sub-distribution model (Fine-Gray model) to correctly separate these two events. Another way to extend the analysis is to use the Joint Model that uses more than one longitudinal biomarker. It could be interesting to investigate the association between serum bilirubin and overall survival in PBC, given the presence of another longitudinal biomarker such as albumin or prothrombin.

Appendix A

Exploratory Analysis

This section includes the exploratory plots for the data exploration for the other explanatory covariates in the PBC clinical data besides bilirubin such as SGOT, prothrombin, and albumin. In particular, they include the histogram and boxplots for baseline measurements; and spaghetti plots for longitudinal measurements of biomarkers. In addition, Kaplan-Meier curves are used to visualize the survival probability of 312 patients across different groups for categorical covariates.

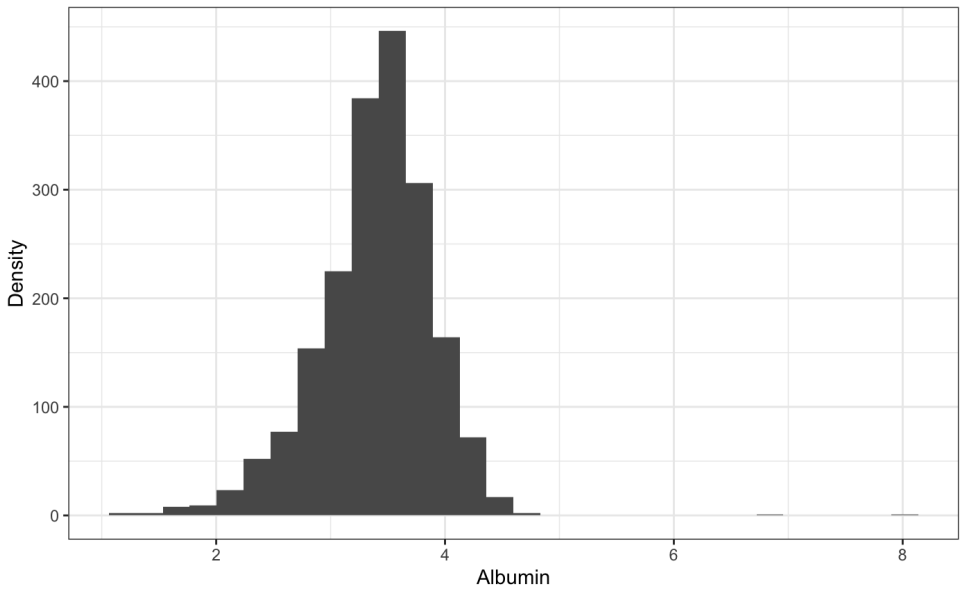

Figure A.1: Distribution of albumin

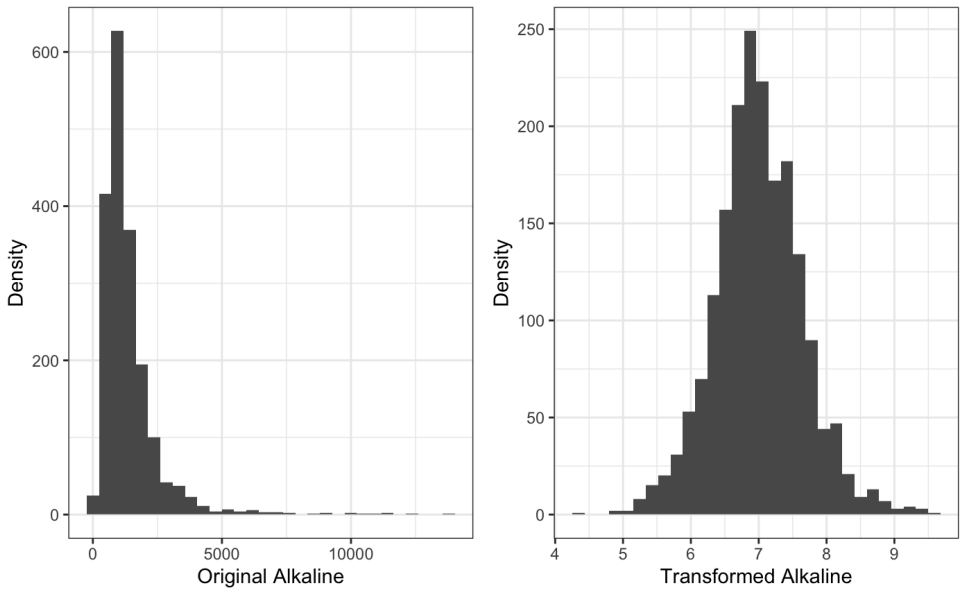

Figure A.2: Distribution of alkaline before and after log transformation

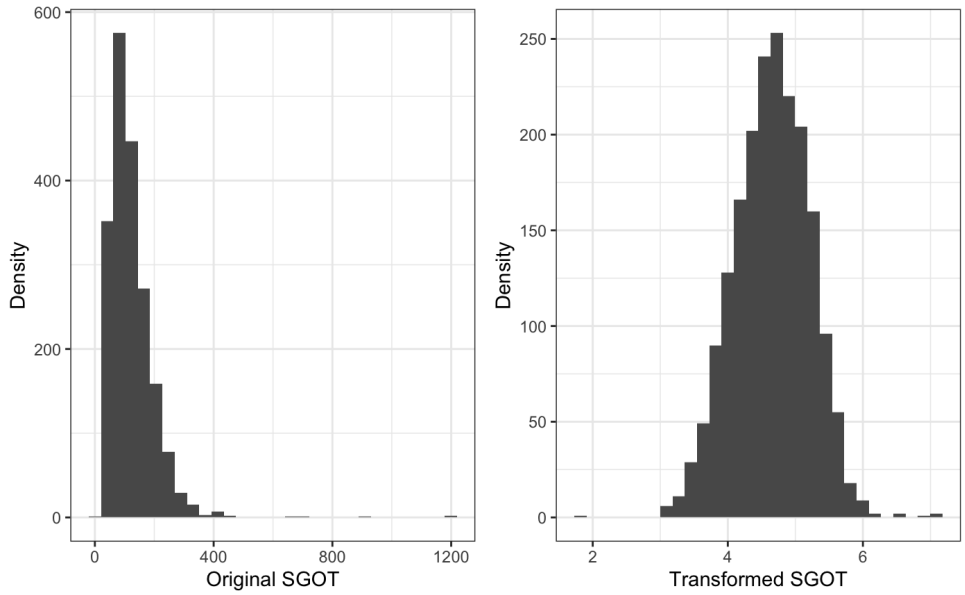

Figure A.3: Distribution of SGOT before and after log transformation

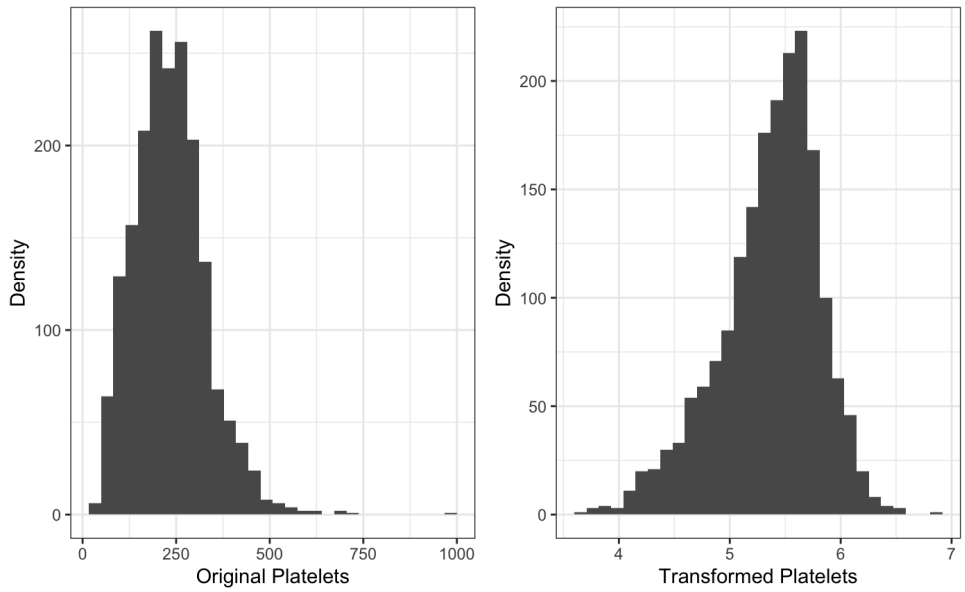

Figure A.4: Distribution of platelets before and after log transformation

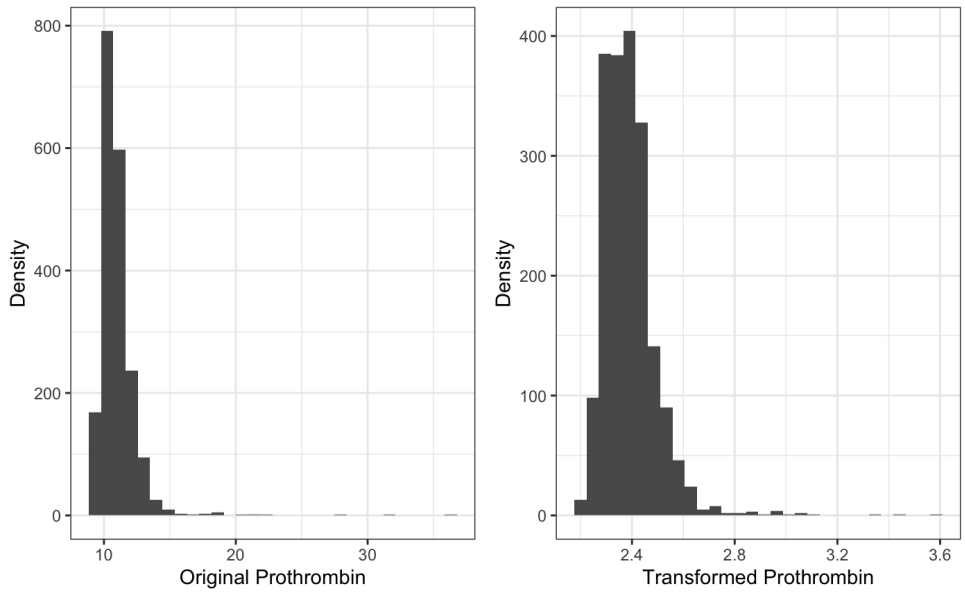

Figure A.5: Distribution of prothrombin before and after log transformation

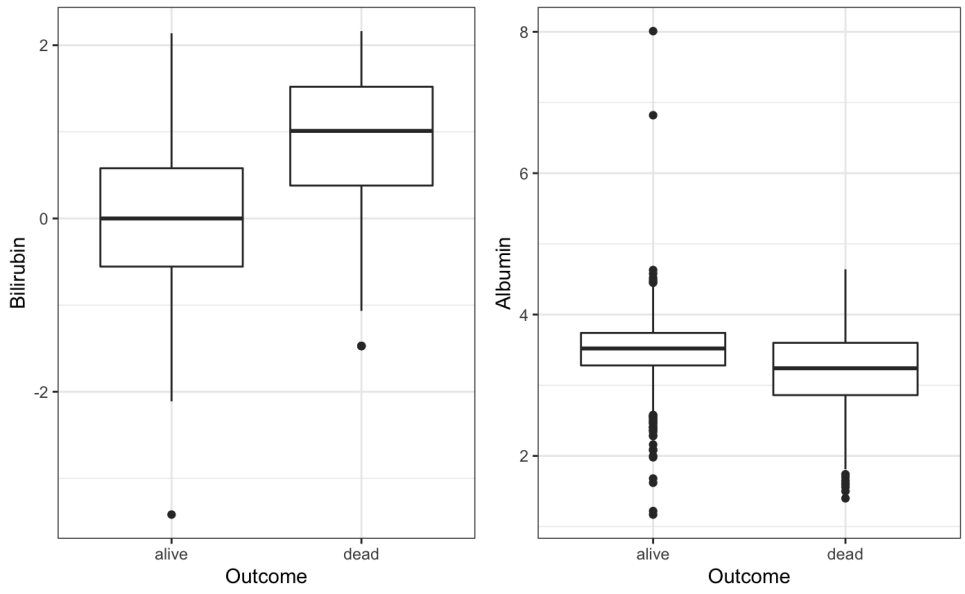

Figure A.6: Boxplots of longitudinal bilirubin and albumin by outcomes

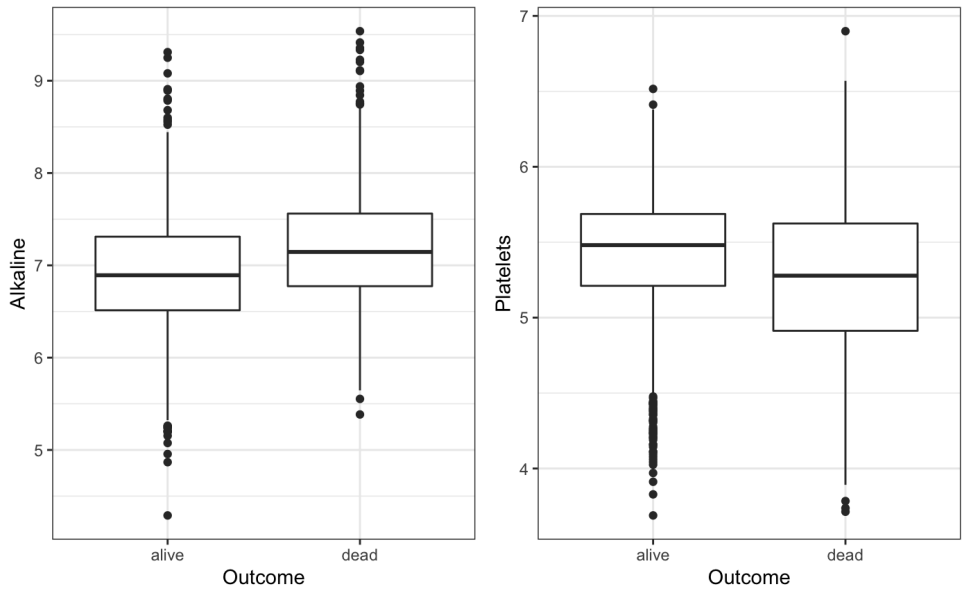

Figure A.7: Boxplots of longitudinal alkaline and platelets by outcomes

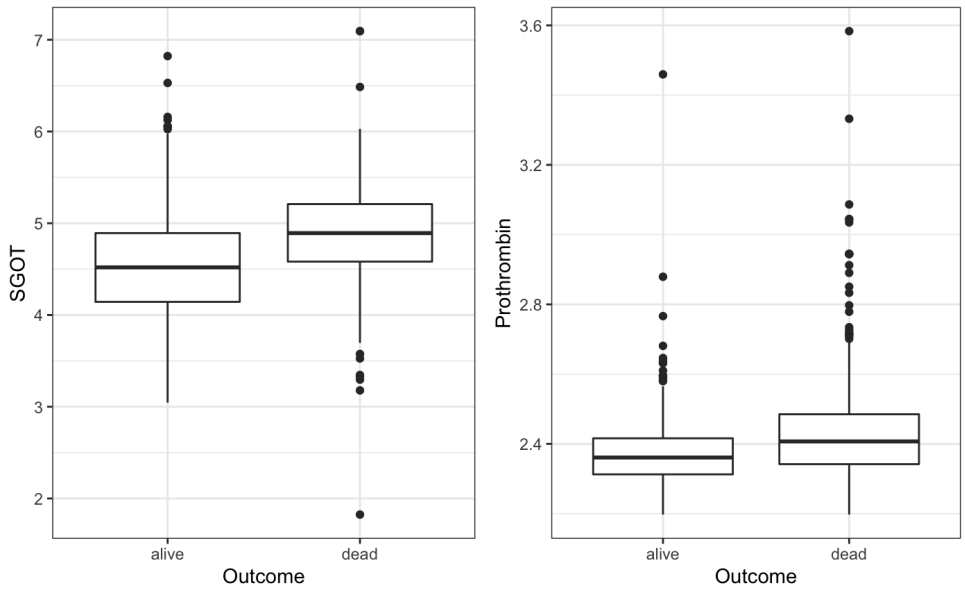

Figure A.8: Boxplots of longitudinal SGOT and prothrombin by outcomes

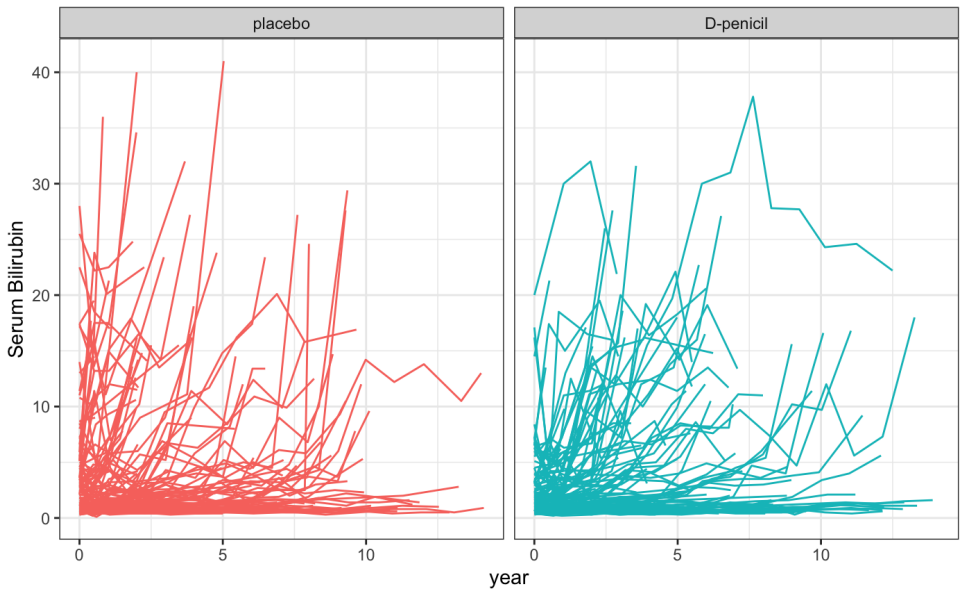

Figure A.9: Spaghetti plot of bilirubin by treatment groupss

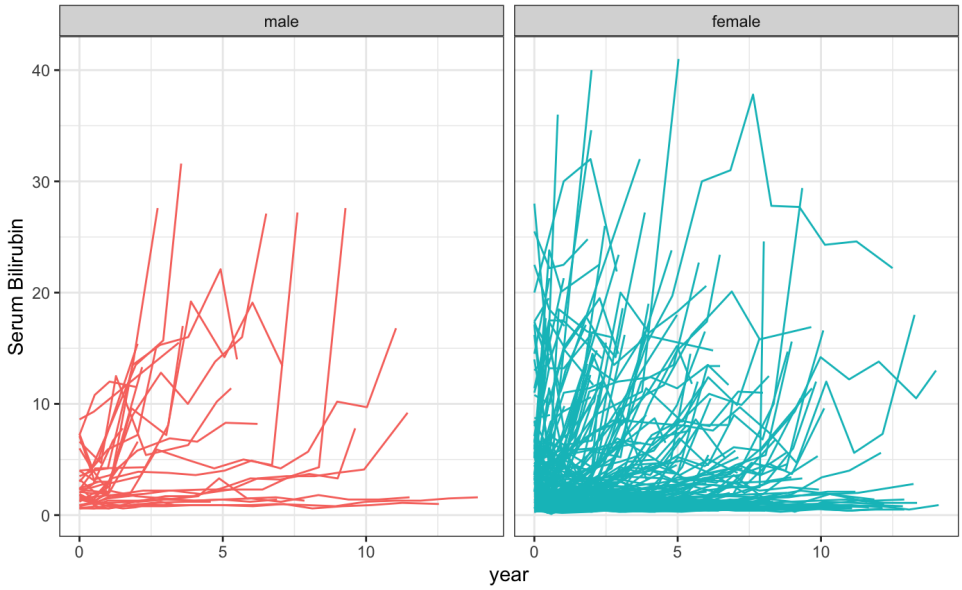

Figure A.10: Spaghetti plot of bilirubin by sex groups

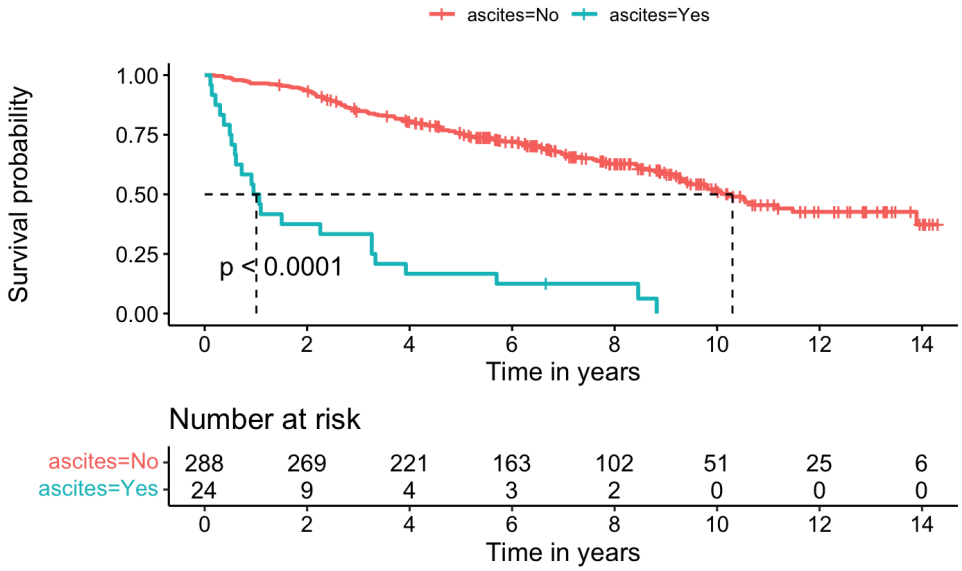

Figure A.11: Survival probability in different ascites groups

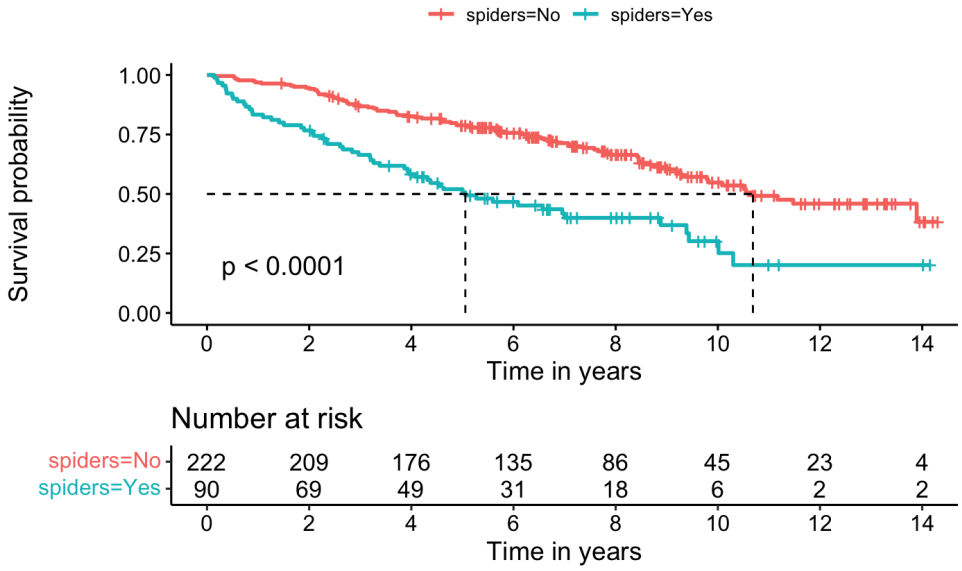

Figure A.12: Survival probability in different spiders groups

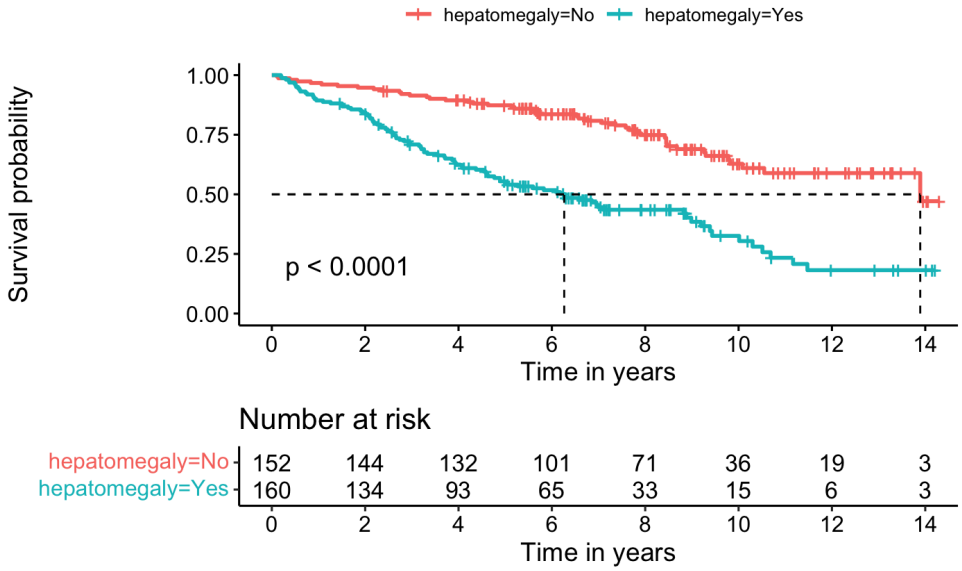

Figure A.13: Survival probability in different hepatomegaly groups

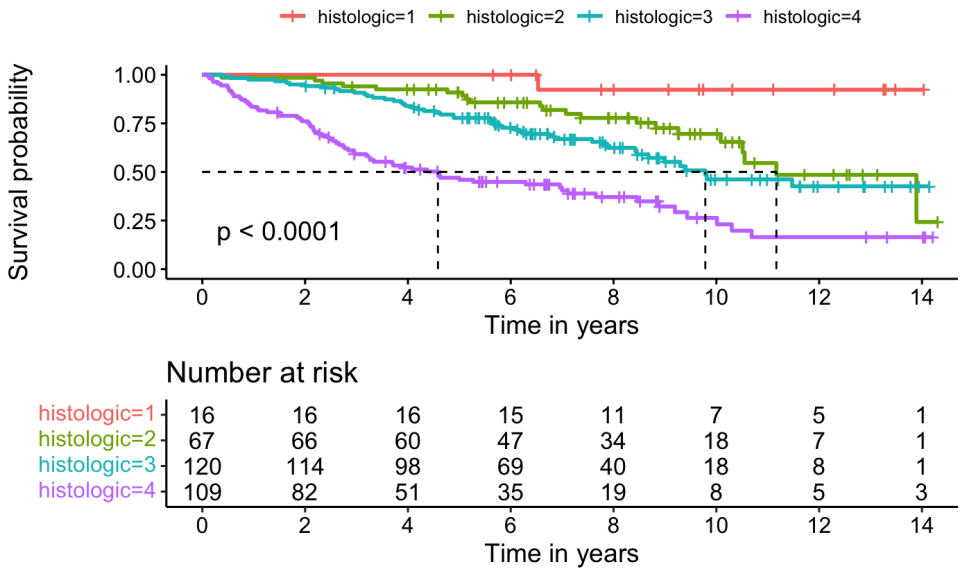

Figure A.14: Survival probability in different histologic groups

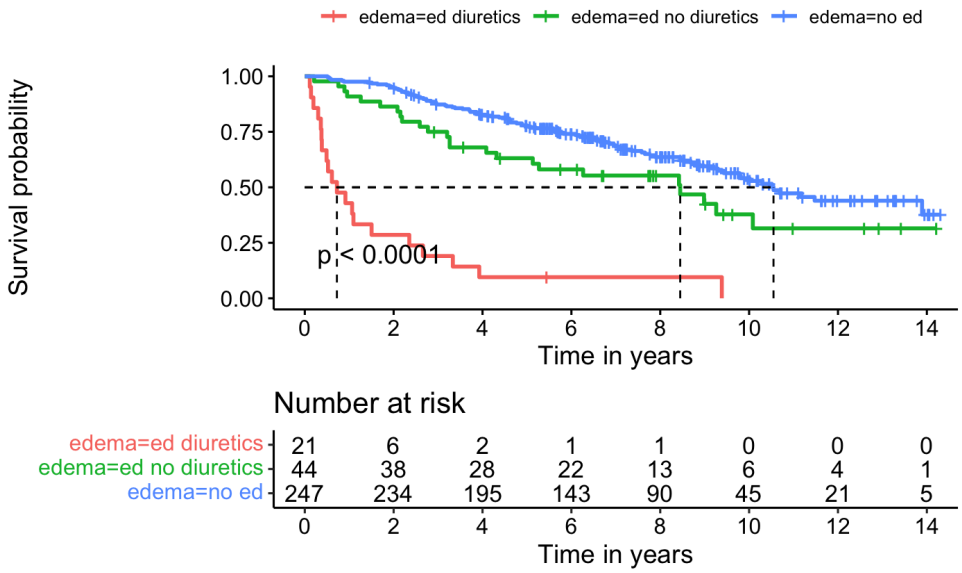

Figure A.15: Survival probability in different edema groups

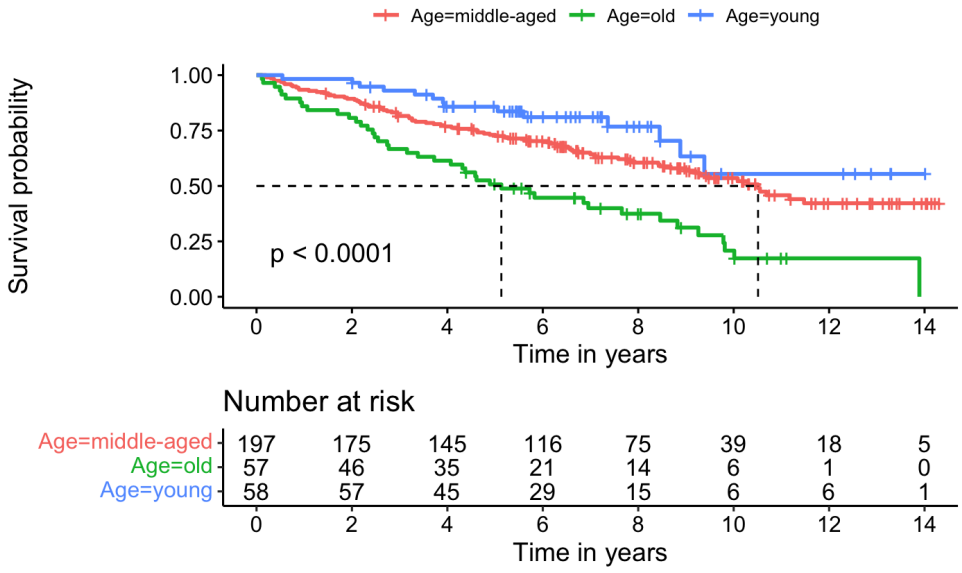

Figure A.16: Survival probability in different age groups. Age is categorized into three groups: young, middle-aged, and high based on the histogram distribution

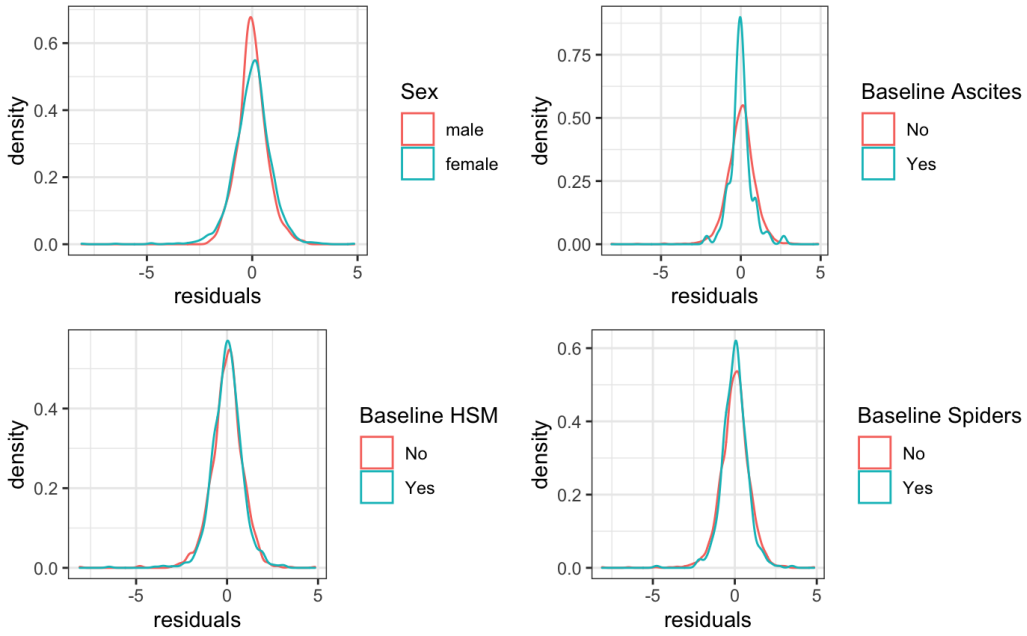

Figure A.17: Distributions of residuals from Linear Mixed-Effects Longitudinal Submodel of the Joint Model

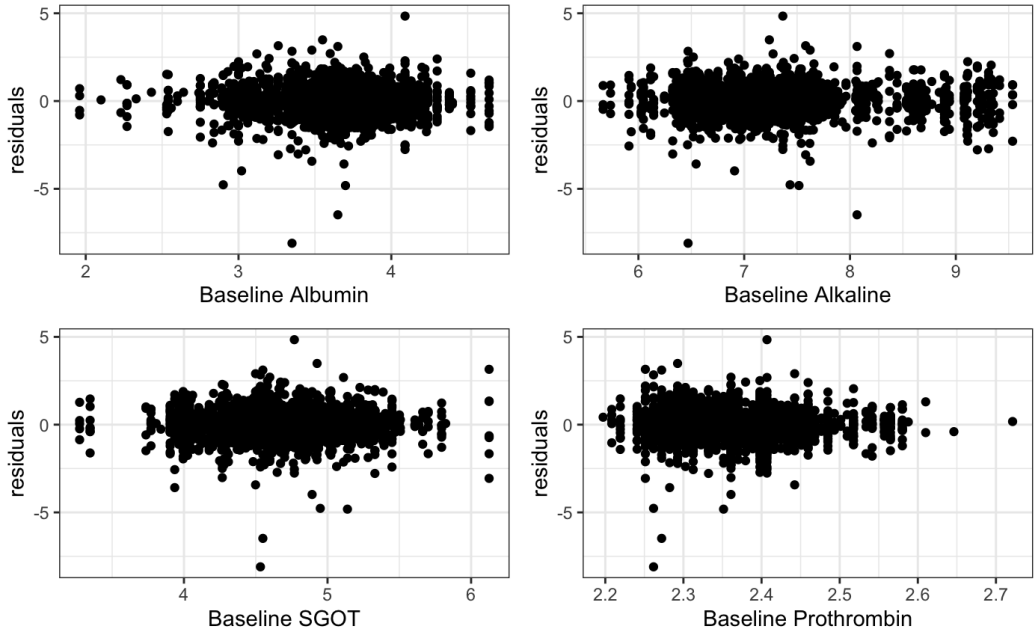

Figure A.18: Scatterplots of residuals from Linear Mixed-Effects Longitudinal Submodel of the Joint Model

Appendix B

Supplementary Analysis

This section includes the supplementary summary tables for the analysis of the PBC clinical data. In particular, they include the univariate and multivariate analyses of the Linear Mixed-Effects model as the longitudinal submodel of the Joint Model. In addition to that, there are summary tables of the multivariate analyses of the three statistical models: Cox PH, Time-Dependent Cox, and Joint Model using the different union sets of covariates that are statistically significant from the backward variable selection for each of these models. Last, it includes Bilirubin Analysis for each of the three models in which bilirubin is the only covariate.

Univariate Linear Mixed-Effects Model			
Characteristics	$\hat{\beta}$	**95% CI**	**p-value**
Drug: placebo	*ref*		0.4
D-penicillin	-0.08	(-0.26, -0.10)	
Age	0.00	(-0.01, 0.01)	0.9
Sex: male	*ref*		0.017
female	-0.35	(-0.64, -0.06)	
Albumin	-0.77	(-1.0, -0.57)	<0.001
Alkaline	0.37	(0.25, 0.49)	<0.001
Ascites: No	*ref*		<0.001
Yes	0.87	(0.53, 1.2)	
Edema: No edema	*ref*		<0.001
Edema no diuretics	0.13	(-0.12 , 0.39)	
Edema diuretics	1.0	(0.63, 1.4)	
Hepatomegaly: No	*ref*		<0.001
Yes	0.64	(0.47, 0.81)	
Histologic: 1	*ref*		<0.001
2	0.38	(-0.04, 0.79)	
3	0.60	(0.20, 1.0)	
4	1.1	(0.71, 1.5)	
Platelets	-0.45	(-0.67, -0.22)	<0.001
Prothrombin	3.5	(2.4, 4.5)	<0.001
SGOT	1.1	(0.9, 1.2)	<0.001
Spiders: No	*ref*		<0.001
Yes	0.63	(0.43, 0.82)	
Year	0.11	(0.09 , 0.12)	<0.001

Table B.1: Univariate Analysis for Linear Mixed-Effects Model Longitudinal Submodel of Joint Model

Multivariate Linear Mixed-Effects Model			
Characteristics	$\hat{\beta}$	**95% CI**	**p-value**
Albumin	-0.25	(-0.42, -0.07)	0.006
Alkaline	0.11	(0.02, 0.20)	0.018
Ascites: No	*ref*		0.027
Yes	0.31	(0.04, 0.59)	
Hepatomegaly: No	*ref*		<0.001
Yes	0.27	(0.13, 0.41)	
Prothrombin	1.3	(0.45, 2.2)	0.003
SGOT	0.82	(0.66, 1.0)	<0.001
Spiders: No	*ref*		<0.001
Yes	0.29	(0.13, 0.44)	
Sex: male	*ref*		0.003
female	-0.31	(-0.52, -0.11)	
Year	0.11	(0.09 , 0.12)	<0.001

Table B.2: Multivariate Analysis for Linear Mixed-Effects Model Longitudinal Submodel of Joint Model

Significant Covariates for Cox PH Model

Multivariate Cox PH Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.11**	**(1.06 - 1.15)**	**<0.001**
Albumin	0.52	(0.32 - 0.85)	0.008
Age	1.04	(1.03 - 1.06)	0.094
Edema: No edema	*ref*		0.019
Edema no diuretics	1.04	(0.65 - 1.67)	
Edema diuretics	2.37	(1.33 - 4.22)	
Histologic: 1	*ref*		0.014
2	4.49	(0.60 - 33.8)	
3	5.79	(0.79 - 42.5)	
4	8.04	(1.09 - 59.5)	
SGOT	1.00	(1.00 - 1.01)	0.012
Prothrombin	1.46	(1.20 - 1.78)	<0.001

Table B.3: Multivariate Cox PH Analysis with Significant Covariates from Variable Selection for Cox PH Model

Multivariate Time-Dependent Cox Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.20**	**(1.17 - 1.23)**	**<0.001**
Albumin	0.59	(0.36 - 0.97)	0.036
Age	1.06	(1.05 - 1.08)	<0.001
Edema: No edema	*ref*		0.003
Edema no diuretics	1.18	(0.72 - 1.93)	
Edema diuretics	3.09	(1.66 - 5.72)	
Histologic: 1	*ref*		<0.001
2	1.97	(0.26 - 14.9)	
3	4.48	(0.61 - 33.0)	
4	6.69	(0.90 - 49.4)	
SGOT	1.00	(0.99 - 1.00)	0.4
Prothrombin	1.11	(0.91 - 126)	0.3

Table B.4: Multivariate Time-Dependent Cox with Significant Covariates from Variable Selection for Cox PH Model

Survival Submodel of Joint Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.81**	**(1.60 - 2.04)**	**<0.0001**
Albumin	0.67	(0.41 - 1.08)	0.1011
Age	1.05	(1.03 - 1.07)	<0.0001
Edema: No edema	*ref*		0.003
Edema no diuretics	1.97	(1.18 - 3.30)	
Edema diuretics	3.11	(1.64 - 5.89)	
Histologic: 1	*ref*		0.3422
2	1.72	(0.26 - 11.26)	
3	2.02	(0.31 - 13.00)	
4	2.58	(0.40 - 16.70)	
SGOT	1.00	(0.99 - 1.00)	0.9569
Prothrombin	1.22	(0.91 - 1.51)	0.623

Table B.5: Survival Submodel of Joint Model with Significant Covariates from Variable Selection for Cox PH Model

Significant Covariates for Time-Dependent Cox

Multivariate Cox PH Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.15**	**(1.11 - 1.18)**	**<0.001**
Albumin	0.50	(0.31 - 0.81)	0.005
Age	1.04	(1.02 - 1.06)	<0.001
Edema: No edema	*ref*		0.002
Edema no diuretics	0.97	(0.60 - 1.55)	
Edema diuretics	2.92	(1.62 - 5.27)	
Histologic: 1	*ref*		<0.001
2	4.63	(0.62 - 34.5)	
3	6.39	(0.88 - 46.7)	
4	10.9	(1.49 - 80.1)	

Table B.6: Multivariate Cox PH Analysis with Significant Covariates from Variable Selection for Time-Dependent Cox Model

Multivariate Time-Dependent Cox Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.20**	**(1.17 - 1.22)**	**<0.001**
Albumin	0.62	(0.38 - 1.01)	0.005
Age	1.07	(1.05 - 1.09)	<0.001
Edema: No edema	*ref*		0.001
Edema no diuretics	1.26	(0.78 - 2.04)	
Edema diuretics	3.31	(1.80 - 6.09)	
Histologic: 1	*ref*		<0.001
2	1.85	(0.24 - 14.0)	
3	4.35	(0.59 - 32.0)	
4	7.09	(0.96 - 52.3)	

Table B.7: Multivariate Time-Dependent Cox with Significant Covariates from Variable Selection for Time-Dependent Cox Model

Survival Submodel of Joint Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.84**	**(1.65 - 2.06)**	**<0.0001**
Albumin	0.60	(0.37 - 0.97)	0.0364
Age	1.05	(1.03 - 1.07)	<0.0001
Edema: No edema	*ref*		<0.001
Edema no diuretics	2.16	(1.30 - 3.59)	
Edema diuretics	3.63	(1.96 - 6.70)	
Histologic: 1	*ref*		0.0878
2	0.92	(0.21 - 4.12)	
3	1.23	(0.28 - 5.29)	
4	1.70	(0.39 - 7.38)	

Table B.8: Survival Submodel of Joint Model with Significant Covariates from Variable Selection for Time-Dependent Cox Model

Significant Covariates for Joint Model

Multivariate Cox PH Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.14**	**(1.11 - 1.17)**	**<0.001**
Albumin	0.37	(0.23 - 0.58)	<0.001
Age	1.04	(1.03 - 1.06)	<0.001
Edema: No edema	*ref*		<0.001
Edema no diuretics	1.01	(0.63 - 1.62)	
Edema diuretics	3.32	(1.83 - 6.01)	

Table B.9: Multivariate Cox PH Analysis with Significant Covariates from Variable Selection for Joint Model

Multivariate Time-Dependent Cox Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.19**	**(1.16 - 1.21)**	**<0.001**
Albumin	0.57	(0.29 - 0.75)	0.002
Age	1.07	(1.05 - 1.09)	<0.001
Edema: No edema	*ref*		0.001
Edema no diuretics	1.32	(0.83 - 2.10)	
Edema diuretics	4.16	(2.25 - 7.70)	

Table B.10: Multivariate Time-Dependent Cox with Significant Covariates from Variable Selection for Joint Model

Survival Submodel of Joint Model			
Characteristics	**Hazard Ratio**	**95% CI**	**p-value**
Bilirubin	**1.84**	**(1.64 - 2.03)**	**<0.0001**
Albumin	0.52	(0.33 - 0.82)	0.054
Age	1.05	(1.04 - 1.07)	<0.0001
Edema: No edema	*ref*		<0.0001
Edema no diuretics	2.05	(1.25 - 3.36)	
Edema diuretics	3.84	(2.09 - 7.06)	

Table B.11: Survival Submodel of Joint Model with Significant Covariates from Variable Selection for Joint Model

Univariate Bilirubin Analysis

Model	Hazard Ratio	95% CI	p-value
Model 1: Cox PH	**1.11**	**(1.06 - 1.15)**	**<0.001**
Model 2: Time-Dependent Cox	1.20	(1.17 - 1.23)	<0.001
Model 3: Joint Model	1.81	(1.60 - 2.04)	<0.0001

Table B.12: Bilirubin Analysis with Significant Covariates for Cox PH Model

Model	Hazard Ratio	95% CI	p-value
Model 1: Cox PH	1.15	(1.11 - 1.18)	<0.001
Model 2: Time-Dependent Cox	**1.20**	**(1.17 - 1.22)**	**<0.001**
Model 3: Joint Model	1.84	(1.65 - 2.06)	<0.0001

Table B.13: Bilirubin Analysis with Significant Covariates for Extended Cox

Model	Hazard Ratio	95% CI	p-value
Model 1: Cox PH	1.14	(1.11 - 1.17)	<0.001
Model 2: Time-Dependent Cox	1.19	(1.16 - 1.21)	<0.0001
Model 3: Joint Model	**1.82**	**(1.64 - 2.03)**	**<0.0001**

Table B.14: Bilirubin Analysis with Significant Covariates for Joint Model

Appendix C

Mathematical Formulae

C.1 Kaplan-Meier Curve

In survival analysis, censoring is taken into account by an event indicator $\delta_i = I(T_i^* \leq C_i)$ such that T_i denotes the observed event time for Subject i and C_i denotes the censoring time. Event indicator δ_i takes value 1 if the true event time is observed and 0 otherwise. This censorship status can be used to construct a non-parametric estimator which makes no assumptions on the underlying distribution of the failure times (Rizopoulos, 2012). The *Kaplan-Meier* (KM) survival probability at failure time t is estimated using the Law of Total Probability:

$$\begin{aligned} \hat{S}(t) &= \hat{S}(t-1) \times Pr(T^* > t \mid T^* > t-1) \\ &= \prod_{i=0}^{t} Pr(T^* > t \mid T^* > t-1) \end{aligned} \tag{C.1}$$

where $Pr(T^* > t \mid T^* > t-1) = \frac{r_i - d_i}{r_i}$ such that r_i denotes the number of subjects at risk at unique time event t_i and d_i denotes number of event t_i.

C.2 Log-Rank Test

In survival analysis, the log-rank test is used to test the difference in survival between two or more independent groups. The formula is given by

$$\text{Log-rank statistic} = \sum_{i=1}^{G} \frac{(E_i - O_i)^2}{Var(E_i)} \tag{C.2}$$

where i denotes the group, and E_i and O_i denote the expected and observed number of events of interest respectively. The null hypothesis is that there is no difference among the survival of the groups while the alternative hypothesis states that there is a difference at any time t. In large samples, the log–rank statistic is approximately χ^2_{G-1} with $G-1$ degrees of freedom where G is the number of groups being compared (Kleinbaum and Klein, 2010).

C.3 Box-Cox Transformation

Box-Cox transformation is a useful family of transformation to approximately normalize the data. It involves logarithmic and power transformations, depending on the value of transformation parameter:

$$y^* = \begin{cases} \dfrac{y^\gamma - 1}{y}, & \gamma \neq 0 \\ \log(\mathrm{y}), & \gamma = 0 \end{cases} \tag{C.3}$$

where y is the response variable. For $\gamma = 0$, the natural logarithm is used and Box-Cox transformation is continuous in γ. Computationally, various choices of γ are considered and the optimal value is selected to provide the best approximation of a normal distribution of the response y.

Bibliography

Albert, P. S. (1999), 'Longitudinal data analysis (repeated measures) in clinical trials', *Statistics in medicine* **18**(13), 1707–1732.

Arisido, M. W., Antolini, L., Bernasconi, D. P., Valsecchi, M. G. and Rebora, P. (2019), 'Joint model robustness compared with the time-varying covariate cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint', *BMC medical research methodology* **19**(1), 1–13.

Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S. et al. (2002), *Analysis of longitudinal data*, Oxford University Press.

Fan, J., Li, G. and Li, R. (2005), 'An overview on variable selection for survival analysis', *Contemporary Multivariate Analysis And Design Of Experiments: In Celebration of Professor Kai-Tai Fang's 65th Birthday* pp. 315–336.

Harrell Jr, F. E., Lee, K. L. and Mark, D. B. (1996), 'Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors', *Statistics in medicine* **15**(4), 361–387.

Ibrahim, J. G., Chu, H. and Chen, L. M. (2010), 'Basic concepts and methods for joint models of longitudinal and survival data', *Journal of Clinical Oncology* **28**(16), 2796.

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

Kaplan, M. M. (1996), 'Primary biliary cirrhosis', *New England Journal of Medicine* **335**(21), 1570–1580.

Kleinbaum, D. G. and Klein, M. (2010), *Survival analysis*, Springer.

Lammers, W. J., Van Buuren, H. R., Hirschfield, G. M., Janssen, H. L., Invernizzi, P., Mason, A. L., Ponsioen, C. Y., Floreani, A., Corpechot, C., Mayo, M. J. et al. (2014), 'Levels of alkaline phosphatase and bilirubin are surrogate end points of outcomes of patients with primary biliary cirrhosis: an international follow-up study', *Gastroenterology* **147**(6), 1338–1349.

Lombard, M., Portmann, B., Neuberger, J., Williams, R., Tygstrup, N., Ranek, L., Ring-Larsen, H., Rodes, J., Navasa, M., Trepo, C. et al. (1993), 'Cyclosporin a treatment in primary biliary cirrhosis: results of a long-term placebo controlled trial', *Gastroenterology* **104**(2), 519–526.

Markus, B. H., Dickson, E. R., Grambsch, P. M., Fleming, T. R., Mazzaferro, V., Klintmalm, G. B. G., Wiesner, R. H., Van Thiel, D. H. and Starzl, T. E. (1989), 'Efficacy of liver transplantation in patients with primary biliary cirrhosis', *New England Journal of Medicine* **320**(26), 1709–1713.

Maruyama, K. (1987), 'The most important prognostic factors for gastric cancer patients: a study using univariate and multivariate analyses', *Scandinavian Journal of Gastroenterology* **22**(sup133), 63–68.

Mells, G. F., Floyd, J. A., Morley, K. I., Cordell, H. J., Franklin, C. S., Shin, S.-Y., Heneghan, M. A., Neuberger, J. M., Donaldson, P. T., Day, D. B. et al.

(2011), 'Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis', *Nature genetics* **43**(4), 329.

Murtaugh, P. A., Dickson, E. R., Van Dam, G. M., Malinchoc, M., Grambsch, P. M., Langworthy, A. L. and Gips, C. H. (1994), 'Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits', *Hepatology* **20**(1), 126–134.

Neuberger, J., Christensen, E., Portmann, B., Caballeria, J., Rodes, J., Ranek, L., Tygstrup, N. and Williams, R. (1985), 'Double blind controlled trial of d-penicillamine in patients with primary biliary cirrhosis.', *Gut* **26**(2), 114–119.

Nicolai, P., Redaelli de Zinis, L. O., Tomenzoli, D., Barezzani, M. G., Bertoni, F., Bignardi, M. and Antonelli, A. R. (1997), 'Prognostic determinants in supraglottic carcinoma: univariate and cox regression analysis', *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck* **19**(4), 323–334.

Paesmans, M., Sculier, J.-P., Libert, P., Bureau, G., Dabouis, G., Thiriaux, J., Michel, J., Van Cutsem, O., Sergysels, R. and Mommen, P. (1995), 'Prognostic factors for survival in advanced non-small-cell lung cancer: univariate and multivariate analyses including recursive partitioning and amalgamation algorithms in 1,052 patients. the european lung cancer working party.', *Journal of Clinical Oncology* **13**(5), 1221–1230.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and Team, R. C. (2007), 'Linear and nonlinear mixed effects models', *R package version* **3**(57), 1–89.

Rizopoulos, D. (2010), 'Jm: An r package for the joint modelling of longitudinal and time-to-event data', *Journal of Statistical Software (Online)* **35**(9), 1–33.

Rizopoulos, D. (2012), *Joint models for longitudinal and time-to-event data: With applications in R*, CRC press.

Rubin, D. B. (1976), 'Inference and missing data', *Biometrika* **63**(3), 581–592.

Schmitz-Moormann, P., Himmelmann, G., Baum, U. and Nilles, M. (1987), 'Morphological predictors of survival in colorectal carcinoma: univariate and multivariate analysis', *Journal of cancer research and clinical oncology* **113**(6), 586–592.

Schober, P. and Vetter, T. R. (2018), 'Survival analysis and interpretation of time-to-event data: the tortoise and the hare', *Anesthesia and analgesia* **127**(3), 792.

Schuppan, D. and Afdhal, N. H. (2008), 'Liver cirrhosis', *The Lancet* **371**(9615), 838–851.

Sjoberg, D., Hannum, M., Whiting, K. and Zabor, E. (2020), 'gtsummary: Presentation-ready data summary and analytic result tables', *Published online* .

Canadian Liver Foundation (2021), '*Cirrhosis of the Liver - Scarring of the Liver, Causes*'. Accessed Feb. 2021.
URL: *http://www.liver.ca/patients-caregivers/liver-diseases/cirrhosis*

Mayo Clinic (2019), '*Liver Function Tests*'. Accessed Feb. 2021.
URL: *http://www.mayoclinic.org/tests-procedures/liver-function-tests/about/pac-20394595*

Mayo Clinic (2021), '*Cirrhosis*'. Accessed Feb. 2021.

URL: *http://www.mayoclinic.org/diseases-conditions/cirrhosis/symptoms-causes/syc-20351487*

Therneau, T. M. and Grambsch, P. M. (2000), The cox model, *in* 'Modeling survival data: extending the Cox model', Springer, pp. 39–77.

Therneau, T. M. and Lumley, T. (2015), 'Package 'survival", *R Top Doc* **128**(10), 28–33.

Verbeke, G. (1997), Linear mixed models for longitudinal data, *in* 'Linear mixed models in practice', Springer, pp. 63–153.