Modelling hepatotoxicity in HIV/TB co-infected patients: Extensions of the Cox Proportional Hazards Model

A dissertation submitted to the Department of Statistical Sciences at the University of Cape Town, in fulfilment of the requirements for the degree of Master of Science in Statistics

by Vintia Philile Mlotshwa

Student number: MLTVIN003

Supervised by Professor Francesca Little

February, 2020

Declaration

I Vintia Phililе Mlotshwa declare that this dissertation is my own unaided work. It is being submitted for the Master of Science in Statistics degree to the University of Cape Town, South Africa. It has not been submitted before for any degree or examination to any other University

---

Student Signed by candidate Signature

On the 7th day of February 2020

Acknowledgements

I would like to acknowledge and sincerely thank Mr Jordache Ramjith for convincing me to pursue this degree and for assisting me with the outline of this project.

I would like to acknowledge and sincerely thank my supervisor, Professor Francesca Little. I am very grateful for her patience and encouragement throughout my research journey.

I would like to thank Dr Kogieleum Naidoo from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) for giving me permission to use the centre's data. I would also like to thank Dr Nonhlanhla Yende-Zuma from CAPRISA for providing insights on how to handle the data.

Lastly, I would like to thank the South African Centre for Epidemiological Modelling and Analysis (SACEMA), the South African Women in Science Awards (WISA) and the First National Bank (FNB) for funding my studies. A special thanks to Gavin Hitchcock from SACEMA and Celia van Niekerk from FNB.

Dedication

This dissertation is dedicated to my mother Thokoza Prudence Ntshele, I love you dearly mother, thank you for your love and support.

Abstract

Hepatotoxicity which is also known as liver damage is mainly caused by intake of medicine. It is common among patients who are co-administering Tuberculosis (TB) treatment and the antiretroviral therapy (ART) for the Human Immunodeficiency Viruses (HIV). If severe, hepatotoxicity sometimes necessitates cessation or interruption of treatment. Therefore, understanding, monitoring and managing hepatotoxicity in patients co-infected with TB and HIV is crucial for optimal treatment outcomes. Hepatotoxicity has been investigated in patients co-infected with TB and HIV, however, most studies have analyzed only the first occurrence of hepatotoxicity and discarded information relating to the resolution and recurrence of hepatotoxicity.

Data from the 'Starting Antiretroviral therapy at three Points in Tuberculosis' (SAPiT) trial is used in this project. This was a trial that was instrumental in finalizing treatment guidelines for patients co-infected with HIV and TB in South Africa.

The clinical objectives of this project are to estimate incidence rates and determine risk factors associated with hepatotoxicity. The statistical objectives are to fit a Cox regression model, the resolution model of hepatotoxicity, and the extended Cox models for recurring events, including the Andersen Gill (AG) model, the Shared frailty model, the Prentice, Williams and Peterson (PWP) total time (TT) model, the PWP gap time (GT) model, as well as a Cox based recurrent model, that models only the second occurrence of hepatotoxicity.

There were 593 patients assessed for hepatotoxicity in the study, 30% (179/593) developed the first occurrence of hepatotoxicity (grade >=1) and 2% (13/593) developed severe hepatotoxicity (grade >=3). Resolved cases (grade = 0) are 76% (136/179) and recurring cases (grade >=1) 24% (32/136). In the Cox multivariable analyses: time-varying treatment arm, older patients, alcohol consumption, low baseline total bilirubin and a positive baseline Hepatitis B surface antigen status, were associated with a higher risk of developing the first occurrence of hepatotoxicity. The extended Cox models (AG model, Shared frailty model, PWP TT model and PWP GT model) in combination identified that: time-varying treatment arm, older patients, alcohol consumption, baseline CD4 count that is greater than 50 cells per mm^3, low baseline total bilirubin, and a positive baseline Hepatitis B surface antigen status were associated with an increased risk of developing recurrent hepatotoxicity. In the resolution model multivariable analyses; non-consumers of alcohol and an abnormal liver function tests at baseline, were associated with an increased chance of

resolving the first occurrence of hepatotoxicity. In the multivariable analyses of the recurrent model: younger patients and the time-varying treatment arm were associated with the development of the second occurrence hepatotoxicity.

Since the Cox regression model utilized data up to the first occurrence of hepatotoxicity, in some instances, the time-varying treatment effect based on the Cox regression model was closer to unity and marginally significant. And the corresponding effect based on the recurrent event models (AG model, Shared frailty model, PWP TT model, PWP GT model and the recurrent model), that utilized data of the first and second occurrence of hepatotoxicity, generally produced a time-varying treatment effect slightly far from unity with a strong statistical significance. This trend was similar for other predictors of hepatotoxicity, like CD4 count and alcohol consumption.

In conclusion, hepatotoxicity is common in this study, however, it is often transient or mild and did not necessitate treatment interruption. However, close monitoring of patients especially in the first 5 months of TB-treatment is recommended. The PWP TT model seemed to be the best model for modelling recurring hepatotoxicity, since the identified risk factors that were associated with hepatotoxicity, changed from the first occurrence of hepatotoxicity to the second occurrence of hepatotoxicity.

Table of contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Background

South Africa carries the highest burden of the Human Immunodeficiency Virus (HIV) pandemic in the world (AVERT, 2019). In 2017, there were 7.2 million people in South Africa estimated to be living with HIV (AVERT, 2019). Making matters worse is the Tuberculosis (TB) pandemic, which is also a leading cause of death in South Africa. It is estimated that people living with HIV have a 60% chance of developing TB due to a weakened immune system caused by HIV (AVERT, 2019). TB is the most common cause of morbidity and mortality in HIV infected patients and people in the African region are most susceptible (World Health Organization, 2018).

The two pandemics are described as inextricably intertwined with a bidirectional relationship (Naidoo, et al., 2015). HIV increases the risk of primary and reactivation of TB infection and this risk increases with advanced HIV (Kwara, et al., 2005). It is pivotal therefore, to identify an effective way of treating these two pandemics.

Up to approximately year 2005, there was very little understanding of clinical management of concurrent HIV and TB. Clinical guidelines regarding the concomitant antiretroviral therapy (ART) and TB treatment were complicated with a high pill burden, overlapping drug toxicities, drug-drug interactions, paradoxical immune reconstitution reaction and hepatotoxicity. These were concerns that justified delaying ART therapy in patients treated for TB or the cessation of the TB treatment, and lives were lost as a result. At this point, prospective studies that identified the optimal timing for initiating ART in patients treated for TB patients were lacking (Kwara, et al., 2005).

In 2010 Abdool Karim, et al. published a paper regarding a prospective study that they conducted named ‘Starting Antiretroviral therapy at three Points in Tuberculosis’ (SAPiT) trial. Their objective was to identify an optimal timing for initiating ART in patients with TB. The trial designed three treamtnent arms, where patients either initiated ART during the intensive phase

of TB treatment or during the continuation phase of TB treatment, or after the completion of TB treatment. Mortality rates were reduced substantially in patients who initiated ART during TB treatment, compared to patients who initiated ART after completion of TB treatment. Other studies in line with the survival benefit of initiating ART during TB treatment, include the Cambodian Early versus Late Introduction of Antiretrovirals (CAMELIA) trial by Blanc, et al., (2011), a clinical trial conducted in Addis Ababa, Ethiopia by Degu, et al., (2012), and a trial that included study participants from different geographies to aid the generalizability of the outcome to inform policy by Havlir, et al., (2011). Subsequently, these studies and others solidified the World Health organization (WHO) guideline for treating the HIV/TB co-infection, which were before this point based on observational studies and expert opinion.

Although the SAPiT trial showed a survival benefit for patients that initiated ART earlier compared to those who initiate ART late, there were still concerns relating to the general wellbeing of HIV infected individuals. These were investigated as secondary outcomes of this trial. For example, the immune reconstitution inflammatory syndrome (IRIS) rates were reported to be significantly higher in the group that initiated ART early compared to the group that delayed ART initiation. However, this did not outweigh the survival benefit, and there were no deaths associated with IRIS, more detail on this is reported elsewhere (Naidoo, et al., 2012). Moreover, other reasons for delaying ART included; a high pill burden, overlapping drug toxicities and concerns about drug-drug complications arising from cotreatment of ART and TB treatment were reported elsewhere (Naidoo, et al., 2014).

Hepatotoxicity due to concomitant administration of ART and TB treatment has not been addressed well to date, given the proposed guidelines of integrating ART therapy earlier rather than late during TB treatment. Hepatotoxicity is commonly referred to as the liver damage that is caused by medicine, chemical, herbal or dietary supplement (U.S. Department of Health and Human Services, 2019). Hepatotoxicity is also commonly known as drug-induced hepatotoxicity, or drug-induced liver injury. Patients cotreated for TB and HIV, generally take four drugs for TB treatment and 3 drugs for ART therapy (now commonly administered as a single pill), these multiple drugs take a toll on the patients' liver, and this results in liver damage. Hepatotoxicity is a serious infection, that if not treated can be fatal (Tostmann, et al., 2008).

There is an extensive amount of literature reporting anti-tuberculosis drug induced hepatotoxicity (Kwara, et al., 2005; Shu, et al., 2013; Tostmann, et al., 2008). Additionally, studies done by Sulkowski, et al., (2000) and den Brinker, et al., (2000) have reported results

about ART induced hepatotoxicity. Several TB treatment drugs are more likely to induce hepatotoxicity than others (Tostmann, et al., 2008) and co-treatment of ART further aggravates hepatotoxicity (Kwara, Flanigan, & Carter, 2005).

There are a few studies in Sub-Saharan Africa that investigate hepatotoxicity in patients that are co-treating HIV and TB; one of which is a study by Hoffmann, et al., (2007). However, the generalizability of this study may be questionable, since the cohort population is predominantly male (94%). Male sex has been identified by Nagu, et al. (2012) as risk factor that increases the occurrence of hepatotoxicity. Therefore, Hoffmann's study may have reported high cases of hepatotoxicity due to the sampling bias of including a disproportionate high male cohort relative to the female cohort.

The intention of this project is to investigate hepatotoxicity in the SAPiT trial, to add to the knowledge base pertaining to hepatotoxicity in Sub-Saharan Africa. The SAPiT trial patients were co-treated for TB and HIV. Therefore investigating hepatotoxicity rates that occured prior and post ART initiation will be considered, where prior ART initiation results will be comparable to studies reporting anti-tuberculosis drug induced hepatotoxicity, and post ART initiation results would be comparable to studies reporting ART induced hepatotoxicity. Furthermore, the SAPiT trial is a randomized controlled clinical trial, thus sampling bias will not be a shortfall as it was for Hoffmann, et al., (2007).

Common risk factors associated with development of hepatotoxicity include, sex, alcohol consumption, and co-infection of Hepatitus B and C virus (Nagu, et al., 2012; den Brinker M. , et al., 2000; de Lima & de Melo, 2012; Yimer, et al., 2014; Tostmann, et al., 2008; Sulkowski, et al., 2000; Pukenyte, et al., 2007). In addittion, there are also risk factors associated with increased hepatotoxicity that are not common, these include, baseline abnormal liver function test and CD4+ count cells per mm3 identified by den Brinker, et al. (2000) and by Hoffmann, et al., (2007), respectively, these would be considered as potential confounders before investigating them as effect modifiers.

Furthermore, studies have merely stated the number of resolution and reccurence of hepatotoxicity cases (Hoffmann, et al., 2007; Nagu, et al., 2012), and do not necessarily analyse the risk factors associated with these events. The omission of this analyses may be partly due to lack of awareness of the advanced statistical methods available. This project will attempt to explore and expose the advanced methods which are an extension of the Cox proportional hazards model, such as the Andersen Gill model, the shared frailty model, the Prentice, Williams

and Peterson (PWP) total time (TT) model, and the PWP gap time (GT) model for recurring hepatotoxicity. Additionally, a simpler extension of Cox proportional hazards model will be considered to model the resolution of hepatotoxicity and the second occurrence of hepatotoxicity.

1.2 Objectives of this study

This research will have two main objectives.

Clinical objectives:

- Investigate the incidence rates of hepatotoxicity in each of the three treatment arms that are defined in the SAPiT trial.
- Identify factors that influence occurrence hepatotoxicity.
- Investigate factors associated with resolved and recurrent hepatotoxicity.

Statistical objectives;

- To assess/evaluate the application of the Cox PH model with non-proportional hazards
- To evaluate and compare models for recurrent events.
- To develop a resolution model for time to resolution of a hepatotoxicity event.

The dissertation proceeds as follows, in Chapter 2 the study design is outlined, and hepatotoxicity is defined. The incidence rates are presented. The survival analysis is introduced, and then the Kaplan-Meier method, the cumulative incidence function and the log-rank test are described and applied.

Chapter 3 introduces the Cox proportional hazards model. Descriptive statistics are presented, and then, an application of the univariate and multivariable Cox proportional hazards model determines the relevant predictor variables that describe the hazard of hepatotoxicity. The model validity is assessed, therafter the Cox model is adequately applied in a non-proportional hazard setting.

In Chapter 4 recurrent event models are described and compared to the Cox model with non-proportional hazards. A resolution model which analyses time to resolving the first hepatotoxicity event is introduced and applied.

Lastly, the Discussion in Chapter 5 refers to the performance of the different models summarizes and interprets the overall results from this project.

Chapter 2

Exploratory Data analysis

2.1 Introduction

This chapter will apply basic concepts and tools of epidemiology to describe the distribution and factors associated with the occurrence of hepatotoxicity.

Firstly, the design overview of the SAPiT trial is described and hepatotoxicity is defined. Upon defining hepatotoxicity, epidemiological tools such as incidence rates and incidence ratios are described and applied to quantify the occurrence of first hepatotoxicity, severe hepatotoxicity, resolved hepatotoxicity and recurring hepatotoxicity in the trial. The probability of surviving and the probability of developing first or recurring hepatotoxicity is evaluated through the Kaplan-Meier Method and the cumulative incidence function respectively. These aforementioned parameters are compared between the treatment arms and or between levels of potential confounders. Lastly, the baseline characteristics of patients in the SAPiT trial are examined and interpreted.

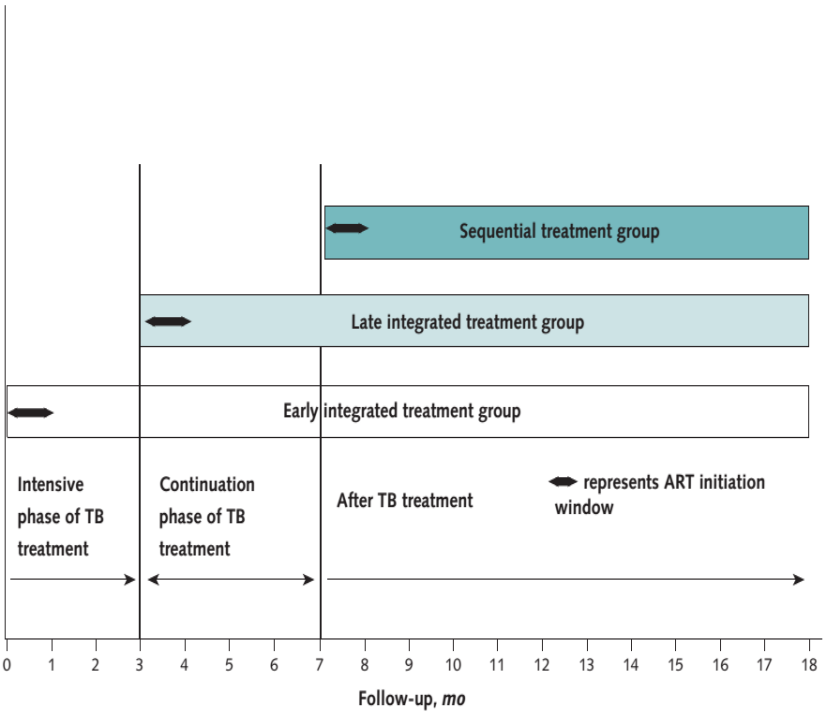
2.2 Design overview

The Starting Antiretroviral Therapy at three Points in Tuberculosis (SAPiT) trial was a randomized, open label clinical trial which was conducted from June 2005 to July 2010. The study was conducted at the Centre for the AIDS Programme of Research in South Africa (CAPRISA) eThekwini clinic for HIV and TB, which adjoins the Prince Cyril Zulu Communicable Disease Centre, an outpatient TB facility in Durban. There were 642 ambulatory patients recruited for this study. Recruited patients were 18 years and older and co-infected with HIV and pulmonary TB. A signed consent form was required from the recruited patients before partaking in the study.

Diagnosis of pulmonary TB was confirmed by a positive sputum smear for acid fast bacilli. The HIV infection was confirmed by two rapid screening tests for HIV. Only patients with CD4+ T cell count of less than 500 per cubic millimetre were included in the study. All 642 patients, initiated TB treatment right away.

The study was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal and the Medicines Control Council of the South African government (BREC Ref number: E107/05).

Patients were randomly assigned into three groups with a ratio of 1:1:1 (with the use of sealed envelopes) in permuted blocks of 6 or 9 with no stratification. Patients were assigned to initiate ART within four weeks of TB treatment initiation (early arm), within four weeks after completion of the intensive phase of TB treatment (late arm), or within completion of TB therapy (sequential arm). A graphical view of how the treatment arms were segmented is shown in Figure 2.1 below.



ART = antiretroviral therapy; SAPiT = Starting Antiretroviral Therapy at Three Points in Tuberculosis; TB = tuberculosis.

Figure 2.1: SAPiT trial study schema (Naidoo, et al., 2012)

Follow up visits for clinical monitoring were scheduled monthly for the first 24 months, and serum tests such as CD4 count, HIV RNA, and the liver serum chemestry tests, such as alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALK) and the gamma-glutamlytransferase (GGT) were performed at screening, at randomization and 6 months thereafter.

2.3 Definitions

2.3.1. Event definition

As mentioned previously, hepatotoxicity is liver damage that is caused by medicine, chemical, herbal or dietary supplement. Liver damage causes enzymes to be released into the bloodstream and the levels of these enzymes are used to monitor the liver function (U.S. Department of Health and Human Services, 2019). Subsequently, the liver function tests are then used to define the extent of hepatotoxicity (Kwara, et al., 2005). Serum liver chemistry tests of alanine aminotransferase (ALT) (Altman, et al., 1995), aspartate aminotransferase (AST) (Binquet, et al., 2009), alkaline phosphatase (ALK) and gamma-glutamlytransferase (GGT) are common chemistries used to classify hepatotoxicity.

It should be noted that hepatotoxicity is defined differently in different settings. In addition to liver function tests, clinical symptoms can also be used to identify hepatotoxicity. Symptoms such as nausea, vomiting, abdominal pain, loss of appetite, diarrhoea, feeling tired or weak, jaundice and hepatomegaly may occur due to the presence of hepatotoxicity (U.S. Department of Health and Human Services, 2019).

In this project, hepatotoxicity is classified according to the Division of AIDS Table Grading and Severity of adults and Paediatric Adverse effects (“DAIDS AE Grading Table”) Version 1.0 (U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Allergy and Infectious Diseases, Division of AIDS, 2004). Hepatotoxicity is classified into grades based on changes of ALT or AST levels relative to the upper limit of normal (ULN):

- grade 0: <1.25 × ULN
- grade 1: 1.25 - 2.5 × ULN
- grade 2: 2.6 - 5.0 × ULN
- grade 3: 5.1 - 10.0 × ULN
- grade 4: > 10 × ULN

The normal range is adjusted by gender for both ALT and AST. The normal range for ALT is 10 - 40 and 7 - 35 for males and females respectively, and the normal range for AST is 15 - 40 and 13 - 35 for males and females respectively. Furthermore, the ULN is replaced by the baseline level

measurement for subjects with elevated baseline liver enzymes (Becker, 2004; Sulkowski, et al., 2000).

As can be seen hepatotoxicity manifests as 5 grade types. However, in this project hepatotoxicity will be investigated as a binary event, where grade >=1 defines the occurrence of any hepatotoxicity, and grade 0 defines the absence of hepatotoxicity. Other ways of grouping the hepatotoxicity grades are considered, and investigated further if the quantity of the data permits. The distribution of hepatotoxicity will be examined interms of the following dimensions

Table 2.1: Dimensions of hepatotoxicity

Dimension	**Description**	**Rationale**
1) First occurrence of hepatotoxicity	As the name suggests, the first occurrence of any hepatotoxicity during the study follow-up.	To cover all grades (grade >=1) that indicate some elevating of liver enzymes above ULN.
2) Severe hepatotoxicity	Classified as grade >=3, otherwise not severe.	To examine the progression of hepatotoxicity in the study.
3) Resolved hepatotoxicity	Defined as the first drop from either grade 1, 2, 3 or 4 to a grade 0 hepatotoxicity.	To investigate protective factors of resolving hepatotoxicity.
4) Recurring hepatotoxicity	Considers repeated cases of hepatotoxicity.	The intention is to use all information relating to the occurrence of hepatotoxicity, to examine if the concluding results change.

It is important to note that severe hepatotoxicity is a composite of the first occurrence of hepatotoxicity. As can be seen, hepatotoxicity will not be investigated on merely the grade levels, but rather on grade groupings, that answer specific questions, for ease of interpretation

2.3.2. Time to event

Described as per the 4 dimensions of the event presented in Table 2.1.

First occurrence of hepatotoxicity

Time at risk of developing hepatotoxicity was calculated as time from randomization to the date when grade >=1 is detected. Time at risk for patients who did not experience hepatotoxicity was

calculated from randomization to: loss to follow up, withdrawal from the study, study termination and or, death which ever occurred first.

Severe hepatotoxicity

Time at risk of developing severe hepatotoxicity was calculated as time from randomization to the date when grade >=3 is detected. Time at risk for patients who did not experience severe hepatotoxicity was calculated from randomization to: loss to follow up, withdrawal from the study, study termination and or, death which ever occurred first.

Resolved hepatotoxicity

Time at risk of resolving hepatotoxicity was calculated as time from first hepatotoxicity (grade >=1) to the date when grade = 0 is detected. Time at risk for patients who did not resolve hepatotoxicity was calculated from first hepatotoxicity to: loss to follow up, withdrawal from the study, study termination and or, death which ever occurred first.

Recurring hepatotoxicity

Time at risk of developing first hepatotoxicity was calculated as time from randomization to the date when grade > = 1 is detected. Time at risk of developing second hepatotoxicity was calculated as time from when first hepatotoxicity was resolved to the date when grade >= 1 is detected again. Time at risk for patients who did not experience hepatotoxicity was calculated from randomization to: loss to follow up, withdrawal from the study, study termination and or, death which ever occurred first. For patients who resolved their first event of hepatotoxicity but did experience the second event of hepatotoxicity, time at risk for these patients was calculated from the resolution of hepatotoxicity of the first hepatotoxicity to: loss to follow of up, withdrawal from the study, study termination and or, death which ever occurred first.

2.3.3. Potential confounders

Confounders that make the patients biological make up vulnerable to hepatotoxicity, as well as increase the likelihood of hepatotoxicity, will consided as follows :

- A variable that classifies patients who had a baseline liver enzymes that was above ULN (i.e. where grade >=1) as abnormal liver function test (LFT) at baseline, otherwise normal LFT at baseline. Baseline LFTs so as to, identifying pre-existings abnormalities before any treatment is adminisistered, so that any elavation of liver enzymes that occur during the course of the study can be assumed to be attributable to the treatment taken.

- A variable indicating if the patient's HIV has progressed severely (baseline CD4 count < 50 cells/mm^3), or not (baseline CD4 count >= 50 cells/mm^3).
- A classification of when the patient developed hepatotoxicity, prior antiretrovial therapy (ART) initiation, or post - ART initiation.

2.4 Data management

The data used were received in a Microsoft excel format, the information was in multiple workbooks. Which was then imported, cleaned and manipulated in SAS Enterprise Guide version 7.1 (SAS Institute Inc., Cary, NC, USA).

2.4.1. Data inclusions and exclusions

All data received was used, except records of patients with missing baseline ALT and AST information. It is important to note that patients with a history of hepatotoxicity may have been included in data, as there was no data available to indicate if the patient had a history of hepatotoxicity or not.

2.5 Descriptive statistics

Statistical analyses were done using SAS Enterprise Guide version 7.1 (SAS Institute Inc., Cary, NC, USA).

2.5.1. Incidence rate

Incidence rates are a core concept of epidemiology. Incidence rates describe the frequency of the event of interest relative to the time each subject is at risk of developing the event. The formula for incidence rates takes the number of events that have develop during the study divided by the sum of the follow up times for all subjects that are at risk of experiencing the event of interest.

Incidence rate can simply be calculated as:

$$\text{Incidence rate } (IR_g) = \frac{R_g}{\sum_{i=1}^{N_g} M_{ig}}, \tag{2.1}$$

where R_g is the number of events that developed during the study period in group g, and M_{ig} is the length of time subject i $(i = 1, \ldots . N_g)$ is at risk for. Follow up time M_{ig} is often called person-time, depending on the time scale this expression could be called person-days, person-months

or person-years. In this project the incidence rate will be reported as person-years, for easier interpretation, incidence rates will be reported per 100 person-years this is, $IR_g \times 100$, will be interpreted. The Poisson approximation was used to determine the 95% confidence intervals (CIs) for incidence rates.

Generally, incidence rate estimates are based on counts of new events and the sum of follow-up times that begin at randomization and ends when the first event occurs or when the subject is lost of follow-up. However, incidence rates can be estimated for recurring events as well, the follow-up time for recurring events begins at the time a subject resolves hepatotoxicity to the time the subject develops the second occurrence of hepatotoxicity, and again time at risk for the third event of hepatotoxicity commences when the second hepatotoxicity has resolved until the time the subject develops third hepatotoxicity, and so on and so forth.

An appealing feature of incidence rate is that it can be estimated per 100 person years, however, you do not need to conduct a study that will last for 100 years to estimate this measurement. A shortfall of this estimate is that, it those not indicate how soon or how late during the study follow-up did the subject develop hepatotoxicity.

2.5.2. Incidence rate ratio

Difference in incidence rates is compared using estimated incidence rate ratios, a ratio is a measurement that indicates how much larger one quantity is compared to another. Incidence rate ratio when comparing group g to group k, is simply $IRR = IR_g/IR_k$. The chi-square test is used to test the difference in incidence rates between the two groups and the F-distribution is used to determine the 95% CIs for incidence rate ratios.

2.5.3. Introduction to survival analysis

The type of data that is considering in this project, tracks patients from baseline to the end of the study. This project is particulary concerned with the time from baseline of the study to the time when patient develops hepatotoxicity (if it is observable). This data is widely known as *time to event* data or the *survival time* data, thus the outcome of interest is the time to event.

2.5.4. Nature of data under survival analysis

The most distinguishing feature of survival time data is that it contains *censored* observations. An observation is said to be censored, if the time at which the event occurred was not observed

during the study period. In other words, the observation is incomplete, since we do not have the full information about the individuals' total survival time. If the observation time ends before the event of interest is observed, this type of censoring mechanism is called *right censoring*. In contrast, if the event of interest occurred before the individual commenced with the study follow up, then such censoring mechanism is called *left censoring*. Furthermore, an outcome of interest may be known to have occurred within some time interval, meaning that the time in which the event of interest occurred is not known exactly. This type of censoring mechanism is called *interval censoring*.

It is common practice to assume that the censoring time for a patient provides no further information about this person's likelihood of survival at a future time, had the individual continued in the study (Klein & Moeschberger, 1997). This assumption is called the non-informative censoring.

In this study, some patients were lost to follow up before developing hepatotoxicity, and some did not develop hepatotoxicity until the study ended. This implies that this study was subjected to the right censoring mechanism.

2.5.5. Main characteristics of survival analysis method

The below notation is introduced to aid the discussion in this section:

- T denotes a random time variable for the subjects in the study.
- Small letter t denotes some specific time of interest, during the study period.
- The Greek letter δ (delta) is an event indicator, $\delta = 1$ indicating event occurrence and $\delta = 0$ indicating indicating that no event was observed.

There are two functions that are frequently considered to characterize the outcome variable under survival analysis. Namely, the *survival function* which gives a probability of a subject surviving beyond time t ($S(t) = P(T > t)$). This function is a non-increasing function as t increases.

The survival function is estimable on a homogeneous population (ignoring any risk factors that may be present) just to see survival experience of the population as whole. Frequently the survival function is stratified by categories of a risk factor of interest, and the survival experience per risk level is comparable; in this project the risk factor will be the treatment arm.

Klein and Moeschberger (1997) mention that the survival function is a very important function that shows us the survival experience of individuals in the study; however, it is difficult to see the failure pattern from it. This motivates the consideration of this second function called the hazard function. The *hazard function* gives the probability that an individual will experience the event in the next instant (i.e. in next infinitesimal interval of Δt), given that the subject has survived up to time t ($P(t \leq T < t + \Delta t | T \geq t)$, per unit time. The 'per unit time' translates to the division of the conditional probability ($P(t \leq T < t + \Delta t | T \geq t)$ by the interval of Δt, to give a rate. The hazard function is denoted as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{2.2}$$

As Δt approaches zero, $\lambda(t)$ is equivalent to the limit, of the probability statement about the survival, divided by Δt (Kleinbaum & Klein, 2005). This function is always non-negative. The function $\lambda(t)$ is also known as the hazard rate, or the conditional failure rate, or instantaneous death rate, or intensity rate, or the force mortality and this list is not exhaustive.

The survival and the hazard function measure contradicting features of the event process, since the former is concerned with the survival and latter concerned with failure. There are general formulae that show how to mathematically manipulate either function to get the other. Given the hazard function, the survival function can be derived as

$$S(t) = \exp\left[-\int_0^t \lambda(u)du\right]. \tag{2.3}$$

The above formula is the exponentiation of the negative integral of the hazard function over the time interval $[0, t]$ to get the survival function. The integral of the hazard function $\int_0^t \lambda(u)du$ is known as the cumulative hazard, and it is widely denoted as $H(t)$.

The hazard function can be estimated for a homogeneous population; however, the hazard function estimated for the heterogeneous population (adjusting for the risk factors) is commonly of more interest.

I will now consider the Kaplan-Meier estimator, which is a method that is commonly used to estimate the survival function.

2.5.6. The Kaplan-Meier (KM) estimate of the survival function

The Kaplan-Meier (KM) method is a non-parametric estimator of the survivor function; non-parametric meaning that the distribution of the survival times, do not need to be specified. The KM method can estimate the survivor function per group, and plotting these group estimates of the survivor function on the same axes will facilitate the comparison of the survival distribution, for example, by treatment arm.

To obtain the Kaplan-Meier estimator, the time intervals based on distinct survival times observed in the study are constructed. Suppose that there are n subjects in a study, who have observed survival times of $t_1, t_2, \dots, t_n$ respectively. Of the n observed survival times, r events were experienced ($r \leq n$), and the rest of the observed times are due to right-censoring. If the observed survival times of the subjects that experienced the event are arranged in ascending order, the j^{th} observed survival time is denoted as $t_{(j)}$ for $j = 1, 2, \dots r$, and the ordered r observed survival times are $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$. Note that, there might be more than one subject who has experienced an event at the same time.

Let n_j be the number of subjects who have not experienced the event by $t_{(j)}$, including those who will experience the event at this time. And let d_j denote the number of subjects who experience the event at $t_{(j)}$. It follows that the conditional probability of a subject experiencing the event in a very small interval of $t_{(j)} - \Delta t$ to $t_{(j)}$ can be estimated by (d_j / n_j). Then the conditional survival probability of a subject surviving beyond time $t_{(j)}$ can be calculated as $(1 - d_j / n_j)$. Note that if $t_{(j)}$ corresponds to both a censored and an event survival time, then censoring is assumed to have occurred immediately after event in computing n_j. In a limit where Δt tends to zero, $(1 - d_j / n_j)$ becomes the estimate of surviving from $t_{(j)}$ to $t_{(j+1)}$ (Collett, 1994). The events that occur per interval are assumed to be independent of each other. Therefore, the Kaplan-Meier estimate of the survival function is the product of the conditional estimated survival probabilities which can be written as,

$$\hat{S}(t)_{KM} = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right). \tag{2.4}$$

The $\hat{S}(t)_{KM}$ estimate is always between 0 and 1, implying that the numerator in $\hat{S}(t)_{KM}$ is always less than the denominator. As patients develop hepatotoxicity and some are censored, both the

numerator and denominator decreases at every successive step. The $\hat{S}(t)_{KM}$ estimate drops in the times when the event occurs and remains constant between observed survival times, resulting in a non-increasing step function commonly called the *survival curve*.

2.5.7. Cumulative incidence function

A non-parametric probability of failure is estimated by the cumulative incidence function, which is not to be confused with the incidence proportion that is a point of estimate rather than a function of the probabily of failure, as it sometimes appears in public health literature (Boston University School of Public Health, 2016). It is intuitive to assume that the complement of the population surviving from the event gives the probability of failure (1 - $\hat{S}(t)_{KM}$). However, it is not that straightforward. Gooley et al (1999) in their comprehensive paper illustrate that (1 - $\hat{S}(t)_{KM}$) is not an unbiased estimate of the probability of failure, since it accounts for competing events as if there were censored events. A competing event is an event that precludes the occurrence of the event of interest from occurring.

When a patient is censored in (1 - $\hat{S}(t)_{KM}$) it is assumed that, that patient will have the same probability of experiencing the event of interest like patients who have not been censored, this affects the level of (1 - $\hat{S}(t)_{KM}$). However, in essence, if a patient experiences a competing event, the probability of failure of the event of interest becomes 0. It is essential therefore that the construct of the probability of failure be based on the hazard of experiencing the event of interest and the hazard of experiencing the competing event.

In the SAPiT trial the competing event is death, patients cannot experience hepatotoxicity if they are dead. It is difficult to ignore death in the study because there is quite a number of patients who die during the trial (Figure 2.2). The cumulative incidence function is proposed to estimate the probability of failure, the formulation is as below:

$$\text{Cumulative incidence Function (CIF)} = \sum_{j:t_j \le t} \frac{d_j}{n_j}\left(1 - \frac{d_j}{n_j}\right)\left(1 - \frac{e_j}{n_j}\right) = \sum_{j:t_j \le t} \frac{d_j}{n_j} \hat{S}(t)_{KM_1} \hat{S}(t)_{KM_2}. \quad (2.5)$$

The CIF formula is an estimate of the probability of failure, the formulae is a product of the hazard of experiencing the event of interest at time t_j denoted by $\frac{d_j}{n_j}$, on condition that the subject has

survived from the event of interest until this time, is estimated as $\hat{S}(t)_{KM_1} = \left(1 - \frac{d_j}{n_j}\right)$, and that this patient has not experienced the competing event $\hat{S}(t)_{KM_2} = \left(1 - \frac{e_j}{n_j}\right)$, this is, e_j is number of subjects who experience a competing event at $t_{(j)}$. The plot of CIF over the study period results in an increasing step function.

2.5.8. The log-rank test

The log-rank test is derived from the Mantel-Haenstel 2 x 2 table test of the difference between 2 groups (Mantel & Haenszel, 1959), Cox adopted this for survival analysis, the test is also referred to as the Cox-Mantel test.

The log-rank test is essentially a chi-square test that compares observed events to expected events under the null hypothesis of independence for each time interval (Kleinbaum & Klein, 2005). This test combines information of the extent of the difference between groups over each of the observed survival times into a single statistic.

Suppose that at time $t_{(j)}$ the below detail can be tabulated:

Group (i)	Number of events	Number subjects who are event free	Number at risk
1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
2	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Study totals at survival time $t_{(j)}$	d_j	$n_j - d_j$	n_j

Furthermore,

- $i = 1,\ 2$ is the group index
- $O_i = \sum_j^r d_{ij}$ is the sum of observed events in group i
- $E_i = \sum_j^r e_{ij} = \frac{n_{ij}}{n_j} \times d_j$, is the number of expected events in group i, where n_{ij} is the number subjects at risk in group i at survival time $t_{(j)}$, $n_j = n_{1j} + n_{2j}$ and $d_j = d_{1j} + d_{2j}$
- Var(O_i- E_i) = $\sum_j^r \left(\frac{n_{1j}n_{2j}(d_j)(n_j - d_j)}{n_j{}^2(n_j - 1)}\right)$ is the variance of the difference of observed events versus expected events in group i.

The log-rank test statistics comparing two groups is written as;

$$\text{log-rank test statistics} = \frac{(O_i - E_i)^2}{\text{Var}(O_i\text{-}\, E_i)} \sim \chi_{(1)} . \quad (2.6)$$

An alternative simpler formulation of the log-rank test statistc, which does not require the calculation of the variance terms, is written as:

$$\text{log-rank test statistics} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \sim \chi_{(1)} . \quad (2.7)$$

Equation (2.7), can be easily generalized to more than two groups by extending the summation to cover all groups. The resulting test statistic would have $k - 1$ degrees of freedom, where k is the number of groups.

A large log-rank test statistic corresponds to small *p*-value, indicating evidence against the null hypothesis that the survival distribution is the same across the groups. This test is a good test when the hazards are expected to be proportional and the sample has relatively few censored events. The power of the log-rank test depends on the number of observed failures rather than the sample sizes.

2.6 Results

2.6.1. Study flow

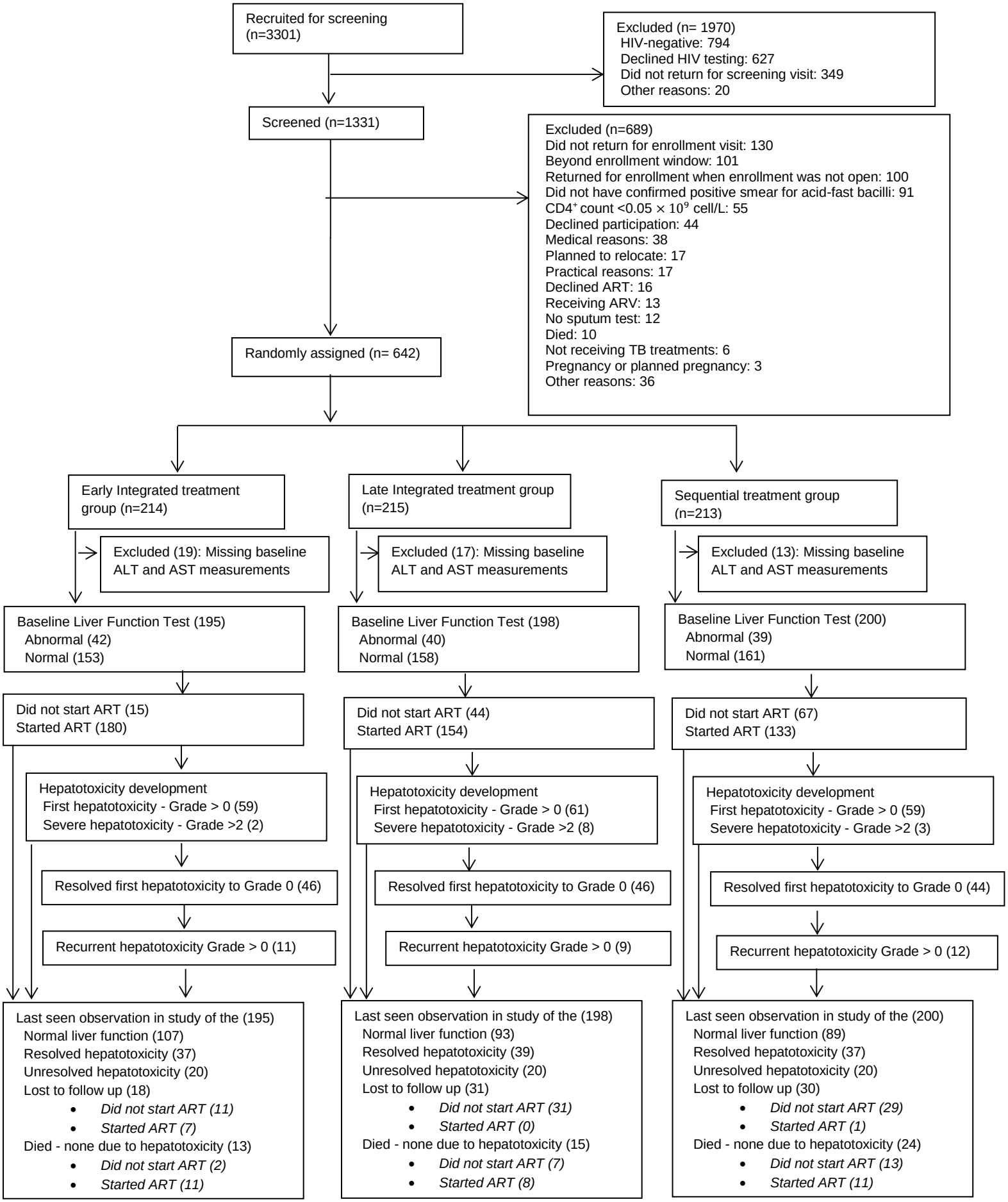
A study flow diagram of how the study evolved from study initiation until study termination is shown in Figure 2.2: Study flow. Of the 642 randomly assigned patients, 49 patients were missing both AST and ALT levels at baseline were subsequently excluded from the analysis, leaving 593 patients eligible for the study. Of 593 patients, 121 had abnormal liver function at baseline, of the 121, 42 patients were from the early arm, 40 were from the late arm and 39 were from the sequential arm. Of the 593, 126 of these patients did not initiate ART, of the 126, there were 15, 44 and 67 patients in the early arm, late arm and sequential arm, respectively. Of the 593, 179 patients developed first occurrence of hepatotoxicity in the study. Of the 179, 2 patients in the early arm, 8 patients in the late arm, and 3 patients in the sequential arm developed severe hepatotoxicity.

Of the 179, 136 patients resolved their first hepatotoxicity sometime in the study. In the early arm, there were 46 resolved first hepatotoxicity, 46 resolved resolved first hepatotoxicity in the late arm and 44 resolved resolved first hepatotoxicity in the sequential arm. Of the 136 resolved cases of

any hepatotoxicity, 11 patients in the early arm, 9 patients in the late arm and 12 patients in the sequential arm developed recurrent hepatotoxicity.

In Figure 2.2 it is shown that 18 patients were lost to follow in the early arm, 11 of these patients did not initiate ART, 31 patients were lost to follow in the late arm, all 31 patients did not initiate ART and 30 patients were lost to follow in the sequential arm, 29 of these patients did not initiate ART. Additionally, of the 593 patients, 52 died during the course of the study none of the deaths were due to hepatotoxicity. Thirteen patients died in the early arm, 2 of these patients did not initiate ART, 15 patients died in the late arm 7 of these patients did not initiate ART, and 24 patients died in the sequential arm, 13 of these patients did not initiate ART.

Figure 2.2: Study flow



2.6.2. Hepatotoxicity incidence

Overall incidence

First occurrence of any hepatotoxicity occurred in 179 (30.2%) patients during study follow-up. In Table 2.2, there were 29.0 incidence cases of hepatotoxicity per 100 person-years(py) that developed in the early arm over a median of 5.2 months, 32.6 incidence cases of hepatotoxicity per 100 py that developed in the late arm over a median of 4.7 months, and 32.6 incidence cases of hepatotoxicity per 100 py that developed in the sequential arm over a median of 5.1 months. Incidence rate ratios between the treatment arms were not statistically significant.

Severe hepatotoxicity

Severe hepatotoxicity (grade 3+) occurred in 13 (2.2%) patients during follow-up. In Table 2.2, there were 0.7 (95% CI 0.1 to 2.7) incidence cases of severe hepatotoxicity per 100 py that developed in the early arm over a median of 9.2 months, 3.2 (95% CI 1.4 to 6.3) incidence cases of severe hepatotoxicity per 100 py developed in the late arm over median of 4.6 months and 1.2 (95% CI 0.3 to 3.5) incidence cases of severe hepatotoxicity per 100 py developed in the sequential arm over a median 13.8 months. There were 77% less severe hepatotoxicity incidence cases in the early arm compared to the late arm, the IRR = 0.23, 95% CI (0,05; 1.1) is marginally significant with p-value = 0.0652.

Resolved hepatotoxicity

Of the patients that developed first hepatotoxicity 136 (77%) resolved their first hepatotoxicity during the study follow up. In Table 2.2, the early arm had 267.3 (95% CI 195.7 to 356.5) incidence cases of resolved hepatotoxicy per 100 py occurred within a median of 2.9 months, 243.1 (95% CI 178.0 to 324.2) incidence cases of resolved hepatotoxicity per 100 py occurred in the late arm within a median of 2.8 months, and 347.8 (95% CI 252.7 to 466.9) incidence cases of resolved hepatotoxicity per 100 py occurred in the sequential over a median of 2.1 months. There were 30% less resolved cases of hepatotoxicity in the late arm compared to the sequential arm, the IRR = 0.7, 95% CI (0,46; 1.1) is marginally significant with p-value = 0.0892 (Table 2.2).

Recurrent hepatotoxicity

Of the cases that resolved first occurrence of hepatotoxicity 32 (24%) developed a recurring hepatotoxicity during the study follow-up. In Table 2.2, there were 49.8 incidence cases of

recurring hepatotoxicity per 100 py that occurred in the early arm within a median of 2.9 months, 29.6 incidence cases of recurring hepatotoxicity per 100 py occurred in the late arm within a median of 2.8 months, and 42.8 incidence cases of recurring hepatotoxicity per 100 py occurred in the sequential arm within a median of 2.1 months. The incidence rate ratios between treatment arms for recurring hepatotoxicity indicated no difference in the incidence rates.

Table 2.2: Overall incidence rates and incidence rate ratios by treatment arm

Treatment Arm	Parameter	**First occurrence**	**Severe hepatotoxicity**	**Resolved hepatotoxicity**	**Recurring hepatotoxicity**
Early Arm	Events, n	59	2	46	11
	Person Years	203.4	269.6	17.2	22.1
	Incidence rate per 100 Person-Years (95% CI)	29 (22.1; 37.4)	0.7 (0.1; 2.7)	267.3 (195.7; 356.5)	49.8 (24.9; 89.2)
Late Arm	Events, n	61	8	46	9
	Person Years	187	251.0	18.93	30.43
	Incidence rate per 100 Person-Years (95% CI)	32.6 (25.0; 41.9)	3.2 (1.4; 6.3)	243.1 (178; 324.2)	29.6 (13.5; 56.2)
Sequential Arm	Events, n	59	3	44	12
	Person Years	184.46	247.5	12.7	28.1
	Incidence rate per 100 Person-Years (95% CI)	32 (24.3; 41.3)	1.2 (0.3; 3.5)	347.8 (252.7; 466.9)	42.8 (22.1; 74.7)
Early Arm vs Late Arm	Incidence Rate Ratio (95% CI); p-value	0.89 (0.62; 1.27); 0.5189	0.23 (0.05; 1.1); 0.0652	1.1 (0.73; 1.65); 0.649	1.68 (0.7; 4.07); 0.2459
Early Arm vs Sequential Arm		0.91 (0.63; 1.30); 0.5944	0.61 (0.1; 3.66); 0.5906	0.77 (0.51; 1.16); 0.2115	1.16 (0.51; 2.64); 0.7148
Late Arm vs Sequential Arm		1.02 (0.71; 1.46); 0.9139	2.63 (0.7; 9.91); 0.1534	0.7 (0.46; 1.06); 0.0892	0.69 (0.29; 1.64); 0.4027

2.6.3. Hepatotoxicity incidence by potential confounders

Hepatotoxicity incidence occurrence of abnormal LFT versus normal LFT at baseline

Of the patients with abnormal baseline LFT 27% (33/121) developed first hepatotoxicity. There were 14.1 cases of first occurrence of hepatotoxicity per 100 py observed in early arm (95% CI 5.7 to 29.0), 39.1 cases of first occurrence of hepatotoxicity per 100 py observed in the late arm (95% CI 21.4 to 65.5), and 38.2 cases of first occurrence of hepatotoxicity per 100 py observed

in the sequential arm (95% CI 19.8 to 66.8). Among patients with abnormal baseline LFT incidence cases of first hepatotoxicity is 64% less in the early arm compared to the late arm the IRR = 0.36, 95% CI (0,15; 0.89), significant with p-value = 0.0276. Similary, incidence cases of first occurrence of hepatotoxicity is 63% less in the early arm compared to the sequential arm, the IRR = 0.37, 95% CI (0,15; 0.94), significant with p-value = 0.0358 (Table 2.3).

Incidence rates among the group of patients who had a normal baseline LFT were similar across the treatment arms, comparative incidence rate ratios were close to unity.

Table 2.3: Incidence of hepatotoxicity for patients with abnormal baseline LFT versus normal baseline LFT

Treatment Arm	Parameter	Abnormal LFT	Normal LFT	**Grand Total**
Early Arm	Events, n	7	52	**59**
	Person Years	49.7	153.8	**203.5**
	Incidence rate per 100 Person-Years (95% CI)	14.1 (5.7; 29)	33.8 (25.3; 44.3)	**29 (22.1; 37.4)**
Late Arm	Events, n	14	47	**61**
	Person Years	35.8	151.2	**187.00**
	Incidence rate per 100 Person-Years (95% CI)	39.1 (21.4; 65.5)	31.1 (22.8; 41.3)	**32.6 (25; 41.9)**
Sequential Arm	Events, n	12	47	**59**
	Person Years	31.4	153.1	**184.49**
	Incidence rate per 100 Person-Years (95% CI)	38.2 (19.8; 66.8)	30.7 (22.6; 40.8)	**32.6 (25; 41.9)**
Early Arm vs Late Arm	Incidence Rate Ratio (95% CI); p-value	0.36 (0.15; 0.89); 0.0276	1.09 (0.73; 1.61); 0.6768	**0.89 (0.62; 1.27); 0.5189**
Early Arm vs Sequential Arm		0.37 (0.15; 0.94); 0.0358	1.1 (0.74; 1.63); 0.6314	**0.91 (0.63; 1.3); 0.5944**
Late Arm vs Sequential Arm		1.02 (0.47; 2.21); 0.9564	1.01 (0.68; 1.52); 0.951	**1.02 (0.71; 1.46); 0.9139**

Hepatotoxicity incidence prior and post ART initiation

There were notable differing incidence rates of hepatotoxicity, that occurred post ART initiation by treatment arm amongst patients with CD4+ cell count that are more than 50 cells/mm^3, 28.4 incidence cases of hepatotoxicity per 100 py developed in early arm (95% CI 20.9 to 37.8). This was 73% more than 16.4 cases per 100 py that developed in the late arm (95% CI 10.5to 24.4) –

p-value =0.0285. And 2.2 times more than 13.1 cases per 100 person-years that developed in the sequential arm (95% CI; 8.0 to 20.2) – p-value =0.0037 (Table 2.4).

Hepatotoxicity incidence stratified by CD4 count strata

In the subgroup of patients with CD4+ cell count that is less than 50 cells/mm^3, there were 26.6 incidence cases of hepatotoxicity per 100 person-years that developed in the early arm (95% CI 12.7 to 48.9) this is almost half of the 53.3 incidence cases of hepatotoxicity per 100 py that developed in the sequential arm (95% CI 28.4 to 91.2). The incidence ratio is 0.5 (95% CI; 0.22 to 1.14), and p-value = 0.0975. Further comparison of the hepatotoxicity incidence rate stratified by CD4 count strata and treatment arm were not significant (Table 2.4).

Table 2.4: Incidence of hepatotoxicity prior and post ART initiation, stratified by CD4 count strata

Treatment Arm	Parameter	CD4 count < 50 cell/L			CD4 count > 50 cell/L			Grand Total
		Prior ART	Post ART	*Subtotal*	Prior ART	Post ART	Subtotal	**Grand Total**
Early Arm	Events, n	2	8	*10*	2	47	*49*	**59**
	Person Years	0.00	37.6	*37.60*	0.3	165.5	*165.8*	**203.4**
	Incidence rate per 100 Person-Years (95% CI)	-	21.3 (9.2; 41.9)	*26.6 (12.7; 48.9)*	-	28.4 (20.9; 37.8)	*29.5 (21.9; 39.1)*	**29 (22.1; 37.4)**
Late Arm	Events, n	5	8	*13*	24	24	*48*	**61**
	Person Years	0.8	34.6	*35.4*	5.1	146.5	*151.6*	**187**
	Incidence rate per 100 Person-Years (95% CI)	-	23.1 (10; 45.5)	*36.7 (19.5; 62.8)*	-	16.4 (10.5; 24.4)	*31.7 (23.3; 42)*	**32.6 (25; 41.9)**
Sequential Arm	Events, n	7	6	*13*	26	20	*46*	**59**
	Person Years	2.1	22.3	*24.36*	7.1	153	*160.1*	**184.46**
	Incidence rate per 100 Person-Years (95% CI)	-	27 (9.9; 58.7)	*53.3 (28.4; 91.2)*	-	13.1 (8; 20.2)	*28.7 (21; 38.3)*	**32.6 (25; 41.9)**
Early Arm vs Late Arm	Incidence Rate Ratio (95% CI); p-value	-	0.92 (0.35; 2.45); 0.869	*0.72 (0.32; 1.65); 0.4421*	-	1.73 (1.06; 2.83); 0.0285	*0.93 (0.63; 1.39); 0.7332*	**0.89 (0.62; 1.27); 0.5189**
Early Arm vs Sequential Arm		-	0.79 (0.27; 2.27); 0.6614	*0.5 (0.22; 1.14); 0.0975*	-	2.17 (1.29; 3.66); 0.0037	*1.03 (0.69; 1.54); 0.8915*	**0.91 (0.63; 1.3); 0.5944**
Late Arm vs Sequential Arm		-	0.86 (0.3; 2.47); 0.7754	*0.69 (0.32; 1.48); 0.3406*	-	1.25 (0.69; 2.27); 0.4558	*1.1 (0.74; 1.65); 0.6374*	**1.02 (0.71; 1.46); 0.9139**

2.6.4. Application of Kaplan-Meier methodology

The hepatotoxicity free survival profiles of the three treatment groups were calculated using the Kaplan Meier method and compared using the log-rank test.

In Figure 2.3 the probability of remaining hepatotoxicity free for patients in the late arm dropped rapidly from beginning of study to the 6th month after study randomization, compared to the other treatment arms. After 6 months, the curves estimating the probability of remaining hepatotoxicity free for the treatment arms are overlapping for most of the study follow up period; this suggests that the probability of remaining hepatotoxicity free is similar across the three treatment arms at this period. The log-rank test statistics that compared the probability of remaining hepatotoxicity free in the three treatment arms is estimated to be 0.4247 with *p*-value =0.8076. The very large *p*-value indicates there is not enough evidence to suggest that survival experience in each of the treatment arms is different.

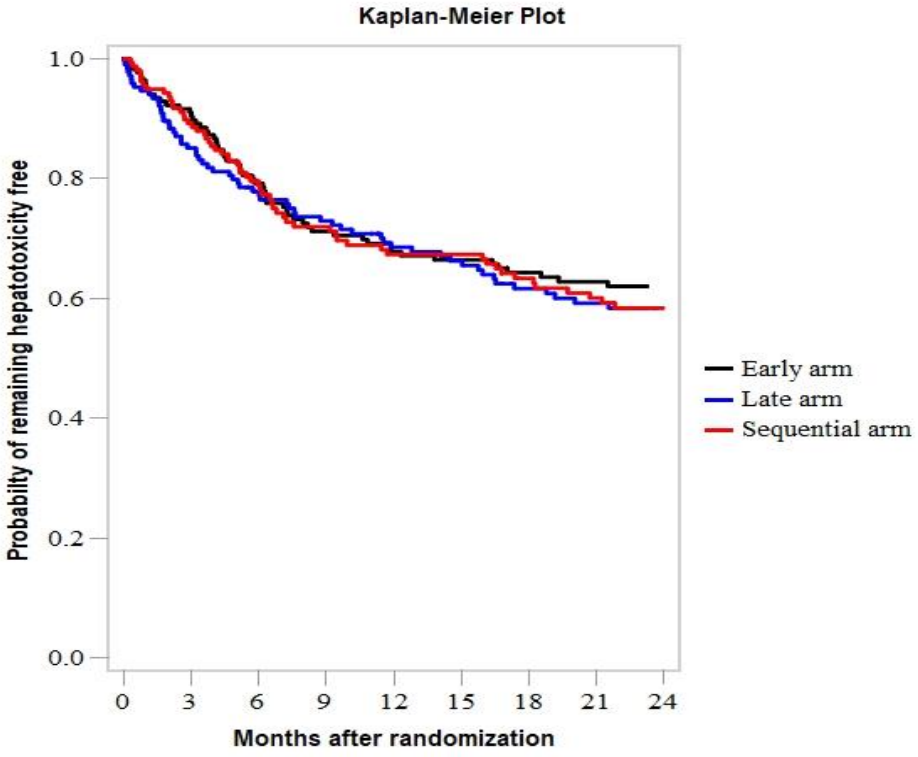


Figure 2.3: Kaplan-Meier estimates first occurrence of any hepatotoxity by treatment arm

2.6.5. Cumulative incidence function estimate

The probability of developing hepatotoxicity in the three treatment groups was calculated using the cumulative incidence function and compared using the log-rank test. In Figure 2.4 the probability of developing first occurrence of any hepatotoxicity is similar across the treatment arms, p-value =0.7395.

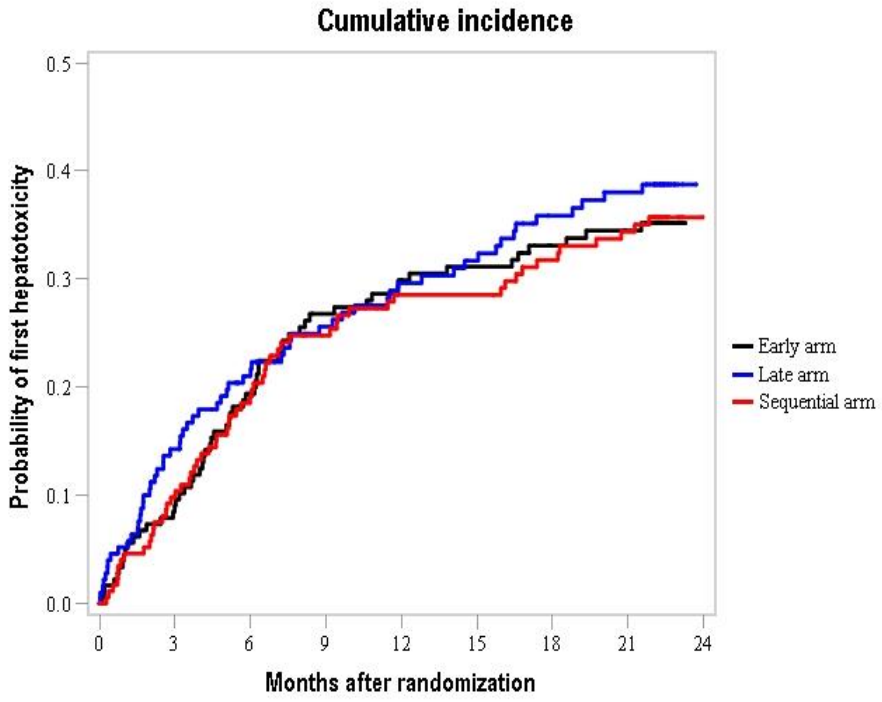


Figure 2.4: : Cumulative incidence function first occurrence of any hepatotoxity by treatment arm

2.7 Conclusion

There were 179 cases of first occurrence of hepatotoxicity in study, the cases were evenly distributed within each treatment arm and the corresponding person-time in each arm was similar. Severe hepatotoxicity occurred in 13 patients, 8 cases occurred in the late arm, a much higher number compared to 2 and 3 cases of severe hepatotoxicity that developed in the early and late arm respectively. There was quite a number of patients in study who resolved hepatotoxicity 76% (136/179). Suggesting that first hepatotoxicity was a transitory state for most patients. Furthermore, there were 24% (32/136) of patients that developed recurring hepatotoxicity after resolving their first hepatotoxicity. Nevertheless, the incidence rate ratio comparing the incidence

rates between the treatment arm for the occurrence of first hepatotoxicity, severe hepatotoxicity, resolved hepatotoxicity and recurring hepatotoxicity were no different to unity.

The KM estimator for the first occurrence of hepatotoxicity revealed that the probability of remaining hepatotoxicity free was similar across the treatment arms. There were also no signicant differences in the cumulative incidence functions between the treatment arms for the first occurrence of hepatotoxicity.

Incidence rates of first occurrence of hepatotoxicity were investigated between the treatment arms and the levels of potential confounders. And for the confounder that classifies patients as having abnormal LFT vs normal LFT at baseline; the analysis revealed that patients in the early arm with abnormal LFT had a statistically lower incidence rates of developing first hepatotoixicty compared to the other treatment arms.

Another confounder investigated was whether hepatotoxicity occured prior or post ART initiation, within each baseline CD4 count level. There were not enough persons at risk prior ART initiation, incidence rates were not estimated in this group and therefore not compared between the treatment arms as a result this confounder was not investigated further. Also this confounder may have been prone to bias as patients would have to have remained hepatotoxicity free in order to have experienced hepatotoxicity post-ART initiation. However, incidence rates in early arm were significantly higher in patients with CD4 greater than 50 cells/mm^3 and whose hepatotoxicity developed post ART initiation compared to the other treatment arms.

Chapter 3

The Cox Proportional Hazards Model

3.1 Introduction

The objective is to ascertain the association of the incidence of hepatotoxicity with treatment arm and other risk factors in the study. This can be done by modeling the hazard function that was introduced in Chapter 2.

The hazard function is typically used to determined event occurrence patterns and describe how the chance of experiencing the event changes over time. This function is modelled in one of two ways; using the multiplicative hazard model procedure or using the less popular approach, called the additive hazard model (Klein & Moeschberger, 1997). The former model, determines association between the risk factor and the occurrence of the disease by examining the relative hazards between risk factor levels, the latter model determines association between the risk factor and the occurrence of the disease by examining the absolute difference in hazards between risk factor levels (Madadizadeh, et al., 2017).

The relative hazard (hazard ratio) quantifies the strength of the association of the risk factor and the occurrence of the disease, whereas the absolute difference quantifies the public health impact of the risk factor, and focuses on the number of cases that could potentially be prevented by eliminating the risk factor (Boston University School of Public Health, 2018). As can be derived from descriptions above, the parameters estimated by multipilicative hazard model and the additive hazard model are equally important. However, in this dissertation only the multiplicative hazards model will be considered. As this model forms the foundation of the extended models that are considered in Chapter 4.

As discussed in Chapter 2, the occurrence of hepatotoxicity can be analysed in one of 4 ways, the focus of this chapter is to determine the risk factors that are associated with the first

occurrence of hepatotoxicity. The occurrence of severe hepatotoxicity will not be considered further in this project due to the sparse occurrence of this event in the study. Resolved hepatotoxicity and recurrent hepatotoxicity are considered in the next chapter.

3.2 Model form

The multiplicative hazard models (Therneau & Grambsch, 2000) are expressed as a product of a baseline hazard function and a non-negative function of covariates. The general formula for the multiplicative hazard models can thus be written as:

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t)c(\boldsymbol{\beta}'\boldsymbol{x}), \tag{3.1}$$

where $\lambda_0(t)$ is the baseline hazard function that can be specified using a parametric form, or left unspecified. Any function can be used for $c(.)$. In 1972, Sir David Cox proposed this function to be $c\left(\boldsymbol{\beta}^{'}\boldsymbol{x}\right) = \exp(\boldsymbol{\beta}^{'}\boldsymbol{x})$ to ensure positivity. The hazard formula of the Cox model for subject i from n subjects can be expressed as:

$$\lambda(t|\boldsymbol{x}_i) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}_i), \tag{3.2}$$

where the baseline hazard function $\lambda_0(t)$ is left unspecified. $\boldsymbol{x}_i$ is the subject specific vector of p covariates, and the $\boldsymbol{\beta}$ is the vector of the corresponding p unknown regression coefficients. Note that, the baseline hazard function is the hazard function for subject i when all the covariates are equal to 0 ($\boldsymbol{x}_i = \boldsymbol{0}$). Furthermore, the vector of covariates include time-invariant covariates and or time-varying covariates, in which case the vector will be denoted as $\boldsymbol{x}_i(\boldsymbol{t})$.

The Cox model is a type of a semi-parametric model because it has an unspecified $\lambda_0(t)$ and a specified $\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)$ element in its model form. A possible advantage of this model over a fully specified parametric model is that, specifying the probability distribution of the baseline hazard may give rise to the risk of choosing the wrong probability distribution (Kleinbaum & Klein, 2005).

3.3 Computing the hazard ratio

Results from the Cox model are often expressed as hazard ratios. Suppose that at time t, subject i has a covariate x_i and subject j has a covariate x_j. The hazard ratio formula between subject i and j is expressed as:

$$\widehat{HR} = \frac{\lambda(t|x_i)}{\lambda(t|x_j)} = \frac{\cancel{\lambda_0(t)}\exp(\beta x_i)}{\cancel{\lambda_0(t)}\exp(\beta x_j)} = \frac{\exp(\beta x_i)}{\exp(\beta x_j)} = \exp\left(\beta(x_i - x_j)\right) \quad (3.3)$$
$$= \hat{\theta},$$

where $\hat{\theta}$ denotes the constant difference in the effect of the covariate x on the relative hazards over time.

The 'constant difference in the effect' between subject i and j is another way of saying that the hazards between subject i and j are proportional over time, also called the *proportional hazards assumption.* As a result, this model is referred as a proportional hazards model.

The proportional hazards assumption is pivotal when considering the Cox model, because interpreting ratios from a Cox model that are breaching the PH assumption may lead to misleading model results. Therefore, it is crucial to check if the proportional hazards assumption holds per covariate considered in the Cox model. This will be demonstrated in subsequent sections.

3.4 Fitting the Cox model

To fit the Cox model presented in equation (3.2), the baseline hazard function $\lambda_0(t)$ and the regression coefficients vector of $\boldsymbol{\beta}$ need to be estimated. However, the interest of this study focuses on obtaining the hazard ratio shown in equation (3.3). An attractive feature of the Cox PH model is that hazard ratios can be estimated without estimating the unknown baseline hazard function (Collett, 1994).

Like in standard regression models, the vector of the β-coefficients is estimable by the method of maximum of likelihood. This method considers a function of the unknown β- coefficients that is a joint probability of obtaining the data that is observed, this function is called the *likelihood function*. The estimates of the β's are the values that are more likely in line with the observed data (Collett, 1994), these estimates are called the maximum likelihood estimates.

In the context of the Cox PH model, the likelihood function is defined by the unknown β-coefficients and the observed survival times that experienced the event of interest. The construction of the likelihood function is informed by the order of occurrence of the uncensored events only.

3.4.1. The Likelihood function

Following Collett (1994), suppose that there are n subjects in the study, r of these subjects experienced the event $(r \leq n)$, and the remaining observed times $(n - r)$ are due to right-censoring. Suppose further that the r subjects had distinct observed survival times, arranged in ascending order. The j^{th} observed survival time is denoted as $t_{(j)}$ for $j = 1, 2, \ldots r$, and the ordered r observed survival times are $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$.

In general, at time $t_{(j)}$ the likelihood function (L_j) is the hazard of experiencing the event for subject j (this is $\lambda_0(t)\exp(\boldsymbol{\beta}' \boldsymbol{x}_j)$), over the sum of the hazard of experiencing the event for each subject that is still under observation in the study (this is $\sum \lambda_0(t)\exp(\boldsymbol{\beta}' \boldsymbol{x}_i)$ if each subject i has not experienced the event of interest). Note that the subjects that are still under observation may include subjects that have not been censored by $t_{(j)}$, but will be censored some time after interval $t_{(j)}$. Since the observed survival times are assumed to be independent, the likelihood function of the study becomes the product of likelihoods at each of the observed survival times of the study $\prod L_j$. Evidently, only the observed survival time times due to an event are used in constructing the likelihood function, under such circumstances this function is called the *partial likelihood function* (Collett, 1994).

The general formula of the partial likelihood suggested by Cox is as follows:

$$\begin{aligned} l_p(\boldsymbol{\beta}, \boldsymbol{x}) &= L(\boldsymbol{\beta}) \\ &= \prod_{j=1}^{r} \left[\frac{\lambda_0(t)\exp(\boldsymbol{\beta}' \boldsymbol{x}_j)}{\sum_{i \epsilon R(t_j)} \lambda_0(t)\exp(\boldsymbol{\beta}' \boldsymbol{x}_i)} \right]^{\delta_j} \\ &= \prod_{j=1}^{r} \left[\frac{\exp(\boldsymbol{\beta}' \boldsymbol{x}_j)}{\sum_{i \epsilon R(t_j)} \exp(\boldsymbol{\beta}' \boldsymbol{x}_i)} \right]^{\delta_j}, \end{aligned} \tag{3.4}$$

where δ_j is a binary indicator of censoring that takes on a value of 1 if the event of interest occurs and 0 if the observation was censored, $\boldsymbol{\beta}$ is the vector of unknown regression coefficients, $\boldsymbol{x}_j$ is the vector of explanatory variables and $R(t_j)$ is the risk set of subjects who have not experienced

the event of interest by ordered observed time $t_{(j)}$. It can easily be seen that the baseline hazard function has cancelled out in the likelihood function formula.

3.4.2. Likelihood function in the presence of tied data

The notion of tied data occurs when subjects in the study have the same observed survival time, this is when both subject a and b contribute the same j^{th} observed survival time $t_{(j)}$ (i.e. $t_{(j)} = t_a = t_b$). For data in the SAPiT trial the observed survival time were computed as months between the date in which a subject was randomized into the study and date in which the subject developed hepatotoxicity or date when the subject was last seen, if the subject was censored. Since time to a hepatotoxicity event was measured in months, there is a high chance that some subjects will have observed survival times that are the same due to the rounding process.

The likelihood method discussed above requires distinct observed survival times in order to correctly calculate the probability (L_j) at each time interval $t_{(j)}$ for $j = 1, 2, \dots r$. Collett (1994) discussed a likelihood method proposed by Efron (1977), which will be applied in this study. Additional notation is required for this method to be discussed. Collett (1994), introduced the vector $\boldsymbol{s}_j$ to denote the vector of sums of each of the p covariates for those subjects who experience the event at the j^{th} time for $j = 1, 2, \dots r$. If there are d_j events at time $t_{(j)}$, the h^{th} element of $\boldsymbol{s}_j$ is $s_{hj} = \sum_{k=1}^{d_j} x_{hjk}$, where x_{hjk} is the value of the h^{th} covariate for $h = 1, 2, \dots p$, for the k^{th} of the d_j subjects, $k = 1, 2, \dots d_j$, who experience the j^{th} event, $j = 1, 2, \dots r$.

The approximation of the likelihood function proposed by Efron (1977) is written as:

$$l_{p(Efron)}(\boldsymbol{\beta}, \boldsymbol{x}) = \prod_{j=1}^{r} \frac{\exp(\boldsymbol{\beta}' \boldsymbol{s}_j)}{\prod_{k=1}^{d_j} \left[\sum_{i \epsilon R(t_j)} \exp(\boldsymbol{\beta}' \boldsymbol{x}_i) - (k-1) d_j^{-1} \sum_{i \epsilon D(t_j)} \exp(\boldsymbol{\beta}' \boldsymbol{x}_i) \right]} \tag{3.5}$$

in which $D(t_j)$ is the set of the subjects who experience the event at time $t_{(j)}$. When there are no ties present in the data, equation (3.5) is equivalent to equation (3.4). There are other methods discussed in literature that approximate the likelihood function for the Cox proportional hazards model in the presence of tied data. Authors such as Collett (1994) and Allison (1995) amongst others, have attested that Efron's method of approximating the likelihood function works quite

well, even when number of tied data points increase. The computational time when using this method is reduced compared to other methods.

3.5 Variable selection approach

The goal of this study is to determine the relationship between the treatment arm and the occurrence of hepatotoxicity, implying that treatment arm will be included in the final model automatically. Furthermore, studies conducted in the past that investigate hepatotoxicity in patients who are co-infected with HIV/TB have given indication of the risk factors that are consistently related to hepatotoxicity. Therefore, these risk factors will be included in the final model accordingly.

Additional risk factors will be included in the final model based on the 'backward elimation' procedure as proposed by (Jewell, 2003). His method follows 6 steps:

1. Fit univariate models for each potential risk factor and retain risk factors with level of significance of *p*-value < 0.2. In this study however, risk factors that will be retained will be those with level of significance of *p*-value < 0.1.
2. Secondly, fit a multivariable model with all risk factors that are retained in Step 1.
3. Thirdly, remove risk factors from this model one by one if they are no longer significant at *p*-value < 0.1. Compare models using the likelihood ratio statistics to ensure that deletion of a variable does not cause a significantly poorer fit.
4. In the fourth step, consider each of the risk factors that were discarded in step 1 on their own to determine whether any one of these risk factors should be added to the model.
5. In the fifth step, consider any relevant interaction terms between pairs of included risk factors.
6. Lastly, assess the final model for goodness of fit.

3.5.1. The Wald test statistic

The Wald test statistic is essentially a ratio of the estimated coefficient to its estimated standard error. This statistic will be introduced as illustrated by Hosmer Jr, et al., (2008). Suppose that $\widehat{\boldsymbol{\beta}}$ is a $p \times 1$ vector regression of coefficients that are maximum likelihood estimates from equation (3.5). And let $\boldsymbol{I}(\boldsymbol{\beta})$ be the $p \times p$ observed information matrix evaluated at $\boldsymbol{\beta}$, this is equivalent to the second derivative of the log partial likelihood ($\boldsymbol{I}(\boldsymbol{\beta}) = \partial^2 L_p(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^{\boldsymbol{2}}$). The estimate of the variance is obtainable by inverting the observed information matrix that is evaluated at $\widehat{\boldsymbol{\beta}}$, this is

$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{I}^{-1}(\widehat{\boldsymbol{\beta}})$. The estimator of the standard error is the positive square root of the variance estimator, denoted as $\widehat{\mathrm{SE}}(\widehat{\beta})$**.** Note that this is notation for a single covariate, for multiple covariates, a variance-covariance matrix is estimated by $\widehat{\mathrm{Var}}\,(\widehat{\boldsymbol{\beta}})$.

The Wald test statistic to test the null hypothesis that $\widehat{\boldsymbol{\beta}}$ has no effect (H_0: $\boldsymbol{\beta} = \mathbf{0}$) can be written as:

$$X_W^2 = \widehat{\boldsymbol{\beta}}\,'\boldsymbol{I}(\widehat{\boldsymbol{\beta}}\,)\widehat{\boldsymbol{\beta}}\,.$$

This test statistic is asymptotically equivalent to a chi-square random variable with p degrees of freedom. It is assumed that asymptotically the p estimated coefficients of $\widehat{\boldsymbol{\beta}}$ are normally distributed with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\boldsymbol{I}^{-1}(\widehat{\boldsymbol{\beta}}\,)$.

The $100(1-\alpha)$ confidence interval for a single coefficient $\widehat{\beta}$ can be defined as:

$$\widehat{\beta} \pm z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{\beta})\mathbf{.}$$

This confidence interval for coefficient $\widehat{\beta}$ is also called the Wald statistic based confidence interval. Therefore, the endpoints of this interval follow the same assumptions that are specified for the Wald test statistic discussed above. These confidence limits can be exponentiated to obtain confidence intervals on the hazard ratio scale.

3.6 Model adequacy assessment

The multivariable Cox model will include treatment arm, age, gender, CD4 count, baseline LFT results and the covariates with strong association identified through Jewell's backwards elimination procedure described above. Model adequacy will be based on quantities called residuals. Residuals are estimated for each individual in the study, and their expected behavior is known when the model is correctly fitted (Collett, 1994).

The residual techniques require estimates of the baseline hazard function. This makes intuitive sense, because to assess if the estimated model fits the data well, you need to have estimates that are based on the full model.

3.6.1. Estimate of the baseline hazard function

The estimation of the baseline hazard function illustrated here is as proposed by Kalbfleisch & Prentice, (1973). Suppose that there are r subjects who experienced the event $(r \leq n)$, and the remaining observed times $(n - r)$ reflect right-censoring. Suppose further that the r subjects had observed survival times (not necessarily all distinct), arranged in ascending order. The j^{th} observed survival time is denoted as $t_{(j)}$ for $j = 1, 2, \ldots r$, and the ordered r observed survival times are $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$.

The estimate of the baseline hazard function at time $t_{(j)}$ is written as:

$$\hat{\lambda}_0(t_j) = 1 - \hat{\alpha}_j, \tag{3.6}$$

in which $\hat{\alpha}_j$ is solution of equation:

$$\sum_{i \epsilon D(t_j)} \frac{\exp(\widehat{\boldsymbol{\beta}}' \boldsymbol{x}_i)}{1 - \hat{\alpha}_j^{\exp(\widehat{\boldsymbol{\beta}}' \boldsymbol{x}_i)}} = \sum_{i \epsilon R(t_j)} \exp(\widehat{\boldsymbol{\beta}}' \boldsymbol{x}_i), \tag{3.7}$$

where $D(t_j)$ is the set of all d_j subjects who experience the event at j^{th} ordered observed time $t_{(j)}$, and $R(t_j)$ is the risk set of subjects who have not experienced the event of interest, at ordered observed time $t_{(j)}$. $\widehat{\boldsymbol{\beta}}$ is a vector of the maximum likelihood estimates that will be obtained from equation (3.5). Collett (1994) pointed out that equation (3.7) requires to be solved by an iterative scheme, as it cannot not be solved explicitly.

Once the baseline hazard has been estimated, then the estimated hazard function for subject i is given by:

$$\hat{\lambda}_i(t|\boldsymbol{x}_i) = \hat{\lambda}_0(t)\exp(\widehat{\boldsymbol{\beta}}' \boldsymbol{x}_i). \tag{3.8}$$

And the cumulative hazard function for subject i is simply the integral of the hazard function denoted as

$$\begin{aligned} \hat{\Lambda}_i(t|\boldsymbol{x}_i) &= \int_0^t \hat{\lambda}_i(u|\boldsymbol{x}_i)du \\ &= \hat{\Lambda}_0(t)\exp(\widehat{\boldsymbol{\beta}}' \boldsymbol{x}_i)\,, \end{aligned} \tag{3.9}$$

where $\hat{\Lambda}_0(\mathrm{t})$ is the cumulative baseline hazard function.

3.6.2. Estimate of the adjusted survival function

Once the baseline hazard function has been estimated, the estimated adjusted survival function of the study can be obtained. Assuming that time t is continuous then the estimate of the adjusted survival function for subject i given by

$$\hat{S}_i(t|\boldsymbol{x}_i) = \left[\hat{S}_o(t)\right]^{\exp(\widehat{\boldsymbol{\beta}}\prime \boldsymbol{x}_i)} \tag{3.10}$$

in which $\widehat{\boldsymbol{S}}_{\boldsymbol{o}}(\boldsymbol{t})$ is the estimated baseline survival function,

$$\hat{S}_o(t) = \prod_{j=1}^{r-1} \hat{\alpha}_j$$

for $t_{(j)} < \ t < t_{(j+1)}$, $j = 1, 2, \ldots r-1$ where $\hat{\alpha}_j$ is the solution of equation (3.7).

3.6.3. Cox-Snell residuals

Cox & Snell (1968) proposed the Cox-Snell residuals that allows an assessment of the overall fit of the model. The Cox-Snell residuals (r_{C_i}) for the i^{th} subject where $i = 1,2, \ldots n$ is given by:

$$r_{C_i} = \hat{\Lambda}_0(t_i)\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{x}_i), \tag{3.11}$$

where $\hat{\Lambda}_0(t)$ is the estimated cumulative baseline hazard function at observed time t_i. $\boldsymbol{x_i}$ is the vector of explanatory variables for subject i and $\widehat{\boldsymbol{\beta}}$ are the corresponding estimated maximum likelihood estimates from equation (3.5). Note that the Cox-Snell residual of r_{C_i}, is the estimated point of equation (3.9) when time is t_i.

If the model is a good fit of the data, this is, if the assumed Cox model holds and $\widehat{\boldsymbol{\beta}}$, $\hat{\Lambda}_0$ are close to the true values $\boldsymbol{\beta}$, Λ_0, then the Cox-Snell residuals r_{C_i} should indicate that they come from an exponential distribution with a mean and variance of 1. This implies that if $\hat{\Lambda}_i(r_{C_i})$ is the cumulative hazard function evaluated at each r_{C_i} for $i = 1,2, \ldots n$, the plot of $\hat{\Lambda}_i(r_{C_i}) = \widehat{\Lambda}_0(r_{C_i})\exp(\widehat{\boldsymbol{\beta}}'\boldsymbol{x}_i)$ against the actual Cox-Snell residual r_{C_i} will have a slope that is equivalent to 1.

3.6.4. Martingale residuals

Wilson (2013) discussed that the partial likelihood method estimates regression coefficients of the respective covariates on the basis that the covariates operate linearly on the hazard function. If the linear relationship is not satisfied than the interpretation of the hazard would be incorrect. The assessment of the linearity is often called the assessment of the functional form. The martingale residual will be used to assess the functional form of the covariate, as discussed extensively by Therneau & Grambsch (2000) amongst other authors.

The martingale residual is basically the difference between the observed number of events and the expected number of events predicted by the fitted model, given the observed survival time and the observed covariate. The martingale residual for the i^{th} subject is written as:

$$r_{M_i} = \delta_j - r_{C_i}, \tag{3.12}$$

where δ_j takes on a value of 1 if the event of interest occurred and 0 if subject i was censored. r_{C_i} is the Cox-Snell residual of subject i as illustrated in equation (3.11), this element represents the number of events expected from subject i based on the fitted model.

The martingale residuals are estimated without the covariate which the functional form is been assessed for, and the resulting martingale residual is plotted for this covariate in question for each subject. If the covariate is fitted correctly in the model, the fit of a loess regression (Cleveland, 1979) line will be a horizontal line at 0. The values of martingale residuals range $-\infty$ to 1, implying that the resulting plot will be asymmetrical.

3.6.5. Deviance residuals

Deviance residuals are conceptually martingale residuals that are transformed to produce values that are symmetric about zero when the fitted model is appropriate (Collett, 1994). They are defined by

$$r_{D_i} = \operatorname{sgn}\left(r_{M_i}\right)[-2\,\{r_{M_i} + \delta_j \log(\delta_j - r_{M_i})\}]^{1/2}\,, \tag{3.13}$$

where r_{M_i} is the martingale residual, δ_j the event indicator of 1 if the event has occurred or 0 and otherwise. Deviance residuals are helpful for identifying outlying observations.

3.7 Proportional hazard assumption test

Two methods discussed by Kleinbaum & Klein (2005) will be used to assess whether the proportional hazards assumption holds for the covariates included in the multivariable Cox model.

3.7.1. Applying the time dependent covariates method

The time-dependent covariates method specifies a model that is an extension of the Cox Model in equation (3.2). This model fits the main effects of the covariates and the interaction effects between the covariate and some function of time ($g(t)$). The form of this model for subject i can be written as:

$$\lambda(t|\boldsymbol{x}_i) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}_i + \boldsymbol{\varphi}'\boldsymbol{x}_i(t)), \quad (3.14)$$

where $\lambda_0(t)$ is the baseline hazard function, $\boldsymbol{x}_i$ is the vector of p covariates, and the $\boldsymbol{\beta}$ is the vector of the corresponding p unknown regression coefficients, $\boldsymbol{x}_i(t) = [\boldsymbol{g}_i(\boldsymbol{t})]'\boldsymbol{x}_i$ is a vector of interaction terms of the $\boldsymbol{x}_i$ vector of p covariates and $\boldsymbol{g_i}(\boldsymbol{t})$ vector of p functions of time, and lastly $\boldsymbol{\varphi}$ is the vector of the corresponding p unknown interaction term coefficients.

This model in equation (3.14) is fitted by a likelihood function specified in equation (3.5), with a modification that, the hazard that a subject contributes into the construction of the likelihood function per time interval will vary with time. Once the model is fitted, the effects of the interaction terms $\boldsymbol{\varphi}$ are tested for significance in the model, using the Wald test statistic with p degrees of freedom. If there is no evidence that the interaction term has an effect on the hazard of event occurrence this would imply that the proportional hazards assumptions is satisfied.

3.7.2. Graphical approach assessment

The graphical approach is based on a double log transformation of the survival curves. This method transforms the estimated survival curves by taking the natural log twice. Suppose the estimated survival curve for subject i given that $\boldsymbol{x}_i$ is vector of covariates is given by:

$$\hat{S}_i(t|\boldsymbol{x}_i) = \left[\hat{S}_o(t)\right]^{\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_i)}. \quad \text{Step 1}$$

Replacing the vector multiplication of $\widehat{\boldsymbol{\beta}}'\boldsymbol{x}_i$ by the summation form $\sum_{l=1}^{p}\widehat{\beta}_l\, x_l$ and taking a natural log on both sides of the equation;

Step 2

$$\ln[\hat{S}_i(t|\boldsymbol{x}_i)] = \exp\left(\sum_{l=1}^{p}\hat{\beta}_l\, x_l\right) \times \ln\left[\hat{S}_o(t)\right].$$

Note that the survival function $\hat{S}_i(t|\boldsymbol{x}_i)$ and the baseline survival function are quantities that range between 0 to unity over time. It follows that the natural log of a number between 0 to unity is negative. This implies that in order to take the second natural log we need to negate both sides of the equation in Step 2 and then take the second log as below:

$$\ln[-\ln[\hat{S}_i(t|\boldsymbol{x}_i)]] = \sum_{l=1}^{p}\hat{\beta}_l\, x_l + \ln\left[-\ln\left[\hat{S}_o(t)\right]\right]. \qquad \text{Step 3}$$

Lastly, both sides of the equation are negated; note that this is usually an optional step, is

$$-\ln[-\ln[\hat{S}_i(t|\boldsymbol{x}_i)]] = -\sum_{l=1}^{p}\hat{\beta}_l\, x_l - \ln\left[-\ln\left[\hat{S}_o(t)\right]\right]. \qquad (3.15)$$

It follows that if equation (3.15) is specified for subject i and j, the difference of these two specifications is independent of time, since element of $\ln\left[-\ln\left[\hat{S}_o(t)\right]\right]$ cancels out. This means that the plot of equation (3.15) over time, for subjects from different categories of the covariate variable should result in curves which are parallel if the proportional hazards assumption is satisfied. This is straight forward for covariates that are categorical in nature. Continuous variables need to be categorized to assess proportional hazards using the graphical approach.

3.8 Treatment of predictors that do not satisfy PH assumption

3.8.1. Cox model with non-proportional hazards

One way to adjust the Cox model for non-proportional hazards is to allow the covariates to interact with time. The general formula for the Cox model with non-proportional hazards for subject i is written as:

$$\lambda(t|\boldsymbol{x}_i) = \lambda_0(t)\exp\left(\sum_{j=1}^{p-q}\beta_j\, x_{ij} + \sum_{l=1}^{q}\varphi_l\, x_{il}(t)\right), \qquad (3.16)$$

where $\lambda_0(t)$ is the baseline hazard function, x_{ij} is the j^{th} covariate that is fixed for the subject i and β_j is the corresponding unknown regression coefficient, lastly $x_{il}(t)$ is the l^{th} covariate that

is time-dependent for the subject i, and φ_l is the corresponding unknown regression coefficient. $x_{il}(t)$ is a covariate which is function of time that can be defined as $x_{il}(t) = x_{il} \times g_l(t)$ where x_{il} is the fixed l^{th} covariate for subject i and $g_l(t)$ is the function of time. Note that equation (3.16) can be fitted by the partial likelihood already specified in equation (3.5); however, it will require more computational time.

The choice of $g_l(t)$ is left to the discretion of the modeler. In general, the choice of $g_l(t)$ affects how the hazard function is interpreted. If $g_l(t)$ is chosen to be t and the resulting estimate of $\hat{\varphi}_l$ is a negative value, this means that the effect of covariate x_{il} increases the hazard of experiencing the event as time increases. And if the estimate of $\hat{\varphi}_l$ is a positive value, then this suggests that the effect of covariate x_{il} decreases the hazard of experiencing the event as time increases.

There are many choices for the form of $g_l(t)$. For example, choosing t as a form of $g_l(t)$ is typically justified if the difference in hazard between the risk levels is monotonically increasing or decreasing over time. Likewise, $\log(t)$ is considered as a form of $g_l(t)$ for the same reason as mentioned above, the natural $\log$ is applied when the monotonic pattern is not readily seen. Kleinbaum and Klein (2005) discussed the heavyside function as a choice for the function time $g_l(t)$.

The heavyside function

The heavyside function is a function that segments the study period into intervals, in order to estimate a different coefficient for covariate x_{il} in each defined interval. The estimated coefficient is constant within the interval, however varies from interval to interval. The general form of heavyside function is expressed as:

$$g_{ls}(t) = \begin{cases} 1 & \text{if } t \leq t_l \\ 0 & \text{if } t > t_l \end{cases} \qquad (3.17)$$

where t_l represents the maximum time for which the hazard ratio will be estimated, otherwise the hazard ratio will be equivalent to zero. The heavyside function defined in equation (3.17), is ideal when the modeler suspects that the hazard ratio in the interval $[0, t_l]$ will constant and then equivalent to zero beyond t_l. Note that the hazard ratio is only obtained for a single interval, since only a single heavyside function has been defined here. If the modeler is interested in more intervals than corresponding number of heavy side functions will need to be defined.

3.8.2. The stratified Cox procedure

Another way to control for variables which do not satisfy the PH assumption is by stratification of the Cox model. Variables that do not satisfy the PH assumption are used to stratify the Cox model, and variables that satisfy the PH assumption are included in the Cox model. The general stratified Cox model discussed here will be as illustrated by Kleinbaum & Klein (2005).

Suppose that there are p explanatory variables that are considered in the Cox model and q of the p variables do not satisfy the PH assumption. For the i^{th} subject, variables that satisfy the PH assumption will be denoted as x_{ij} where $j = 1, 2, \cdots p - q$, and variables that do not satisfy the PH assumption will be denoted as z_{il} where $l = 1, 2, \cdots, q$.

To perform a stratified cox model, the categories of $z_.'s$ are considered, if the variable is continuous it is categorized. The combination of the categories is formed, and these combinations form the stratas. A combination of the categories form a new variable denoted as a z^*. The stratification variable z^* has k^* categories, where k^* is the total number of possible combinations (or strata) formed after categorizing each of the $z_.'s$. The general stratified Cox model is written as

$$\lambda_g(t|\boldsymbol{x}_i) = \lambda_{0g}(t)\exp\left(\sum_{j=1}^{p-q} \beta_j\, x_{ij}\right) \tag{3.18}$$

where $g = 1, 2, \cdots, k^*$ indicates a stratum of z^* and $\lambda_{0g}(t)$ is the baseline hazard function for the gth stratum, x_{ij} is the j^{th} covariate that is fixed for the subject i and β_j is the corresponding unknown regression coefficient. As can be seen the variable z^*, is not explicitly included in the in the model but the $x_.'$s which are assumed to satisfy the PH assumption are included.

The regression coeffiecients $\beta_.$'s are estimable by the maximum of likelihood method. The likelihood function L is the product of partial likelihood functions that are determined for each stratum, $L = \prod_{g=1}^{k^*} L_g$, each L_g is determined by equation (3.5). The $\beta_.$'s estimates are then the values that maximize the likelihood function.

Note that the baseline hazard function is allowed to vary per stratum. However, the regression coefficients are the same for each statum, this feature is what is called a 'no-interaction assumption'. The no-interaction assumption is tested by fitting an alternative model that assumes that, there is interaction in the model, this model is written as :

$$\lambda_g(t|\boldsymbol{x}_i) = \lambda_{0g}(t)\exp\left(\sum_{j=1}^{p-q} \beta_{jg}\, x_{ij}\right) \tag{3.19}$$

where $g = 1, 2, \cdots, k^*$ indicates a stratum of z^*, and $\lambda_{0g}(t)$ is the baseline hazard function for the gth stratum, x_{ij} is the j^{th} covariate that is fixed for the subject i and β_{jg} is the corresponding unknown regression coefficient for the gth stratum. The log-likehood test is then performed between the no-intereaction models equation (3.18) and equation (3.19), with $p - q(k^* - 1)$ degrees of freedon, if the outcome of the test is statitistically significant, then it will mean that the interaction model is the appropriate model to interpret, otherwise the no-interaction will be more appropriate.

3.9 Application

3.9.1. Background

The primary aim of the SAPiT trial was to identify the optimal timing for initiating ART therapy in patients who are HIV/TB co-infected, and are on TB treatment. All patients in the trial initiated TB treatment at or before randomization. Patients were then randomized to either initiate ART within four weeks of the TB therapy (early arm), or within four weeks after completion of the intensive phase of tuberculosis treatment (late arm), or within completion of TB therapy (sequential arm).

Abdool Karim, et al., (2010) found that the mortality rates in trial were at least doubled in patients who were in the sequential arm, compared to patients in the early and the late arm combined. They also reported that the HIV RNA was suppressed in higher levels for patients in the early and late arm compared to patients in the sequential arm; this is after 12 months of study randomization. This study indicated that there was a survival benefit for those patients who were in the early and late arm compared to patients in the sequential arm.

Of interest in this analysis is the relationship of liver injury (hepatotoxicity) within the three treatment arms. Patients in the early and late arm co-administered more drugs at the same time, compared to the number of drugs co-administered by patients in the sequential arm, leading to a clinical hypothesis that patients in the early and late arm will be at a higher risk of hepatotoxicity compared to patients in the sequential arm; because of the number drugs taken by each arm. Of secondary interest in this project is to model the impact of the covariates on the occurrence hepatotoxicity. Thus, the summary statistics of these covariates by treatment arm is reviewed next. Moreover, the assessment of the covariate characteristic that describes a subject with baseline abnormal liver function to those with normal liver function is performed to test if this variable is confounding or not. If necessary, a description of the covariate and the reasons for including these covariates in the analysis is discussed.

3.9.2. Covariates distribution at baseline

Continuous variables were summarised using means and standard errors, or medians and interquartile ranges. Categorical variables were summarized using frequencies and proportions. Wilcoxon rank sum test or unpaired t-test were used for the analysis of continuous data. Fisher's exact test was used for the analyses of categorical variables. Observed p-values were reported for group comparisons.

3.9.3. Baseline characteristics

The demographic data variables that were considered in the SAPiT trial were age and gender. The overall median and mode of the age in the study is 33 and 31 years of age respectively, this indicates that the age distribution is positively skewed under this study. The median age of a study subject was 34 years old in the early arm, 33 years old in the late arm and 33 years old in the sequential arm. The proportion of males was 45% in the early arm, 52% in the late arm and 52% in the sequential arm (Table 3.1).

There were not that many subjects that were underweight across the treatment arms. Only 12% of subjects were underweight in the early arm, 13% in the late arm and 14% in sequential arm (Table 3.1). There were 37% of the subjects in the early arm, 32% of the subjects in the late arm, and 31% of the subjects in sequential arm, reporting a history of TB. Furthermore, patients with abnormal liver function at baseline were less likely to have a history of TB than patients with a normal liver function at baseline (22% versus 36%, p-value = 0.0024).

Patients, who reported suffering from extra pulmonary TB were not that common in the study, with only 5% of the study subjects in the early arm, 4% of the study subjects in the late arm, and 4% of the study subjects in the sequential arm. In subjects with extra pulmonary TB, none had TB of the liver. The World Health Organization (WHO) clinical stage 4 (severely symptomatic stage, designation includes all of the AIDS-defining illnesses) is not frequent in the study, with a proportion 7% of subjects in the early arm, 5% of subjects in the late arm and 6% of subjects in the sequential arm presenting with WHO stage 4.

Alcohol use has been investigated in literature primarily because alcohol use may cause liver disease. There were 12%, 11% and 14% patients in the early, late and sequential arm respectively who reported to be occasional consumers of alcohol. In patients with abnormal liver function 13% of the patients reported to be occasional consumers of alcohol, this is comparable to 12% of patients, who have a normal liver function and reported to be an occasional consumer of alcohol.

$CD4^+$ count $<0.05 \times 10^9$ cell/L is commonly a threshold that indicates severe HIV progression, 17% of subjects in the early arm, 16% of subjects in the late arm and 19% of subjects in the sequential arm were at a severe HIV progression stage. Furthermore, patients with abnormal liver function at baseline were more likely to have CD4+ count $< 0.05 \times 10^9$ compared to patients with a normal liver function at baseline (35% versus 13%, p-value < 0.001).

The $CD8^+$ count median was 0.7×10^9 cell/L across the treatment arms, which is within the range $[0.2 – 1.0] \times 10^9$ cell/L that is considered normal. However, patients with abnormal liver function at baseline had a median of $0.5\times 10^{\mathbf{9}}$ CD8+ count, this is significantly lower than the median of $0.7\times 10^{\mathbf{9}}$ CD8+ count observed in patients with normal liver function at baseline (p-value = 0.005). The medians for $\log_{10}$ HIV RNA were the same across all treatment arms.

Alkaline phosphatase (ALK), total bilirubin (BIL) and lactate dehydrogenase (LDH) are liver function tests that are performed along with ALT and AST liver function tests. It is of interest to also explore the relationship of these serum chemistries to hepatotoxicity, given that they come from the same source as that of ALT and AST. The median levels for ALK, BIL and LDH were similar across the treatment arms. While, patients with abnormal liver function at baseline had higher ALK, BIL and LDH medians compared to patients with normal liver function at baseline.

The proportion of patients who had Hepatitis B surface antigen (HBsAg) positive status compared to patients with a negative HBsAg status at baseline, was the same across the treatment arms. Additionally, there was no significant difference between patients with abnormal liver function compared to patients with normal baseline function with respect to HBsAg status (p-value = 0.17). However, 17% of observations were missing a HBsAg status at baseline and this result may not be precise, since the missing number of observations are above the threshold of 5 % (Marshall, et al., 2010).

Table 3.1: Baseline characteristics of participants in the SAPiT trial						
Variable	**Early arm (n=212)**	**Late arm Group (n=211)**	**Sequential arm (n=211)**	**Patients with liver enzyme abnormalities (n=121)**	**Patients with normal liver function tests (n=472)**	***p*- Value [a]**
Median age (IQR)	34 (28 - 39)	33 (28 - 40)	33 (28 - 38)	34 (28 - 39)	32 (28 - 39)	0.6096
Male, n (%)	97 (45.3)	112 (52.1)	110 (51.6)	63 (52.1)	230 (48.7)	0.5417
BMI<18.5 kg/m^2, n (%) [b]	25 (11.7)	28 (13.0)	29 (13.6)	12 (10.0)	63 (13.6)	0.3592
History of tuberculosis n (%)	80 (37.4)	68 (31.6)	66 (31.0)	26 (21.5)	169 (35.8)	0.0024
Extra pulmonary Tuberculosis, n (%) [c]	10 (4.7)	9 (4.2)	9 (4.3)	8 (6.6)	18 (3.8)	0.212
WHO stage 4, n (%)	14 (6.5)	11 (5.1)	13 (6.1)	11 (9.1)	24 (5.1)	0.127
Alcohol occasionally consumed, n (%) [d]	24 (11.7)	23 (11.0)	28 (13.7)	15 (12.8)	53 (11.6)	0.8769
Alcohol frequently consumed, n (%) [d]	6 (2.9)	9 (4.3)	9 (4.4)	4 (3.4)	20 (4.4)	
Patients with CD4$^+$ count <0.05 $\times$ 10^9 cell/L n (%)	37 (17.3)	35 (16.3)	41 (19.25)	42 (34.7)	61 (12.9)	<.0001
Median CD8$^+$ count (IQR), $\times$ 10^9 cell/L	0.697 (0.417 - 1.030)	0.660 (0.455 to 1.084)	0.663 (0.476 to 0.957)	0.521 (0.336 – 0.877)	0.690 (0.482 – 1.021)	0.0048
Median log$_{10}$ HIV RNA (IQR), copies/Ml	5.0 (0.9)	5.0 (0.9)	5.1 (0.7)	5.1(0.9)	5.0 (0.9)	0.277
Median AST levels (IQR)	29.0 (22.0 to 41.0)	29.0 (22.0 to 42.0)	29.0 (22.0 to 42.0)			
Median ALT levels (IQR)	20.0 (13 - 33)	19.0 (13 - 29)	17.0 (12 - 28)			
Median Alkaline Phosphate (IQR)	84.0 (68.0 to 109.0)	86.5 (69.0 to 111.0)	83.0 (67.0 to 122.0)	102.0 (74.0 to 149.0)	82.0 (68.0 to 105.0)	<.0001
Median Total bilirubin (IQR)	7.0 (5.0 to 10.0)	7.0 (5.0 to 10.0)	7.0 (5.0 to 11.0)	8.0 (5.0 to 14.0)	7.0 (5.0 to 10.0)	0.0019
Median Lactate Dehydrogenase (IQR)	254.0 (215.0 to 306.0)	243.0 (212.0 to 295.0)	253.0 (217.0 to 297.0)	306.0 (264.0 to 376.0)	240.0 (210.0 to 281.0)	<.0001
HBsAg Positive n (%) [e]	15 (8.8)	15 (8.1)	13 (7.5)	12 (11.3)	30 (7.6)	0.2366

SD = Standard Deviation BMI = Body Mass Index; IQR =Interquartile range; WHO = World Health Organization

[a] p-value for the comparison of patients with liver enzyme abnormalities to those without.

[b] Two patients in the late integrated arm and six patients in the sequential arm had missing baseline BMI data, which were not included in the percentage calculation.

[c] Among these patients, none had tuberculosis of the liver. One patient in the late integrated arm, 2 patients in the sequential arm had missing extra pulmonary tuberculosis data, which were not included in the percentage calculation.

[d] Eight patients in the early integrated arm, 4 patients in the late integrated arm and 6 patients in the sequential arm had missing alcohol records, which were not included in the percentage calculation.

[e] The Hepatitis B Surface Antigen status was missing for 37, 22 and 33 patients in early integrated arm, late integrated arm and sequential arm respectively.

3.9.4. Fitting the Cox Model

The candidate covariates to be included in the Cox model are age, sex, body mass index (BMI) strata, history of TB, extra pulmonary TB, World Health Organization (WHO) stage, alcohol consumed, CD4 count category, CD8 count, baseline log viral load, alkaline phosphate (ALK), total bilirubin (BILI), lactate dehydrogenase (LDH), hepatitis B surface antigen (HBsAg) and baseline LFTs.

It is sensible to assess the association of baseline measurement with the occurrence of hepatotoxicity, as it is at baseline that the appropriate treatment regimen for a patient is determined and administered. Therefore, the analysis in this chapter will examine the baseline characteristics of a patient that are associated with the occurrence of hepatotoxicity, so as to inform clinicians on what precautionary measures to take when treating the patients with a high risk of developing hepaototoxicity.

It is important to note the SAPiT is a randomised control trial, and therefore the abovementioned variables will be included in the model, with aim to quantify the association of these variables in relation to the occurrence of hepatotoxicity, and these variables are not included merely to adjust for confounding.

The model build process will follow the backward elimination procedure as proposed by Jewell (2003).

Univariate assessment

A univariate Cox PH model is fitted per baseline exposure variable considered in the study. Table 3.2 presents the parameter estimate ($\hat{\beta}_.$), standard error and a p-value based on the univariate assessment.

Table 3.2: Cox PH model Univariate assessment

	Univariate assessment			
Baseline exposure variables	**Parameter estimate**	**Standard error**	**Hazard Ratio**	***p* – value**
Early arm vs. Sequential arm (Reference)	-0.082	0.184	0.921	0.6573
Late arm vs. Sequential arm (Reference)	0.035	0.183	1.036	0.8488
Age (years)	0.019	0.008	1.019	**0.0215**
Sex = Male vs. Sex = Female (Reference)	0.306	0.150	1.358	**0.0414**
BMI category < 18.5 kg/m^2 vs. ≥18.5 kg/m^2 (Reference)	-0.130	0.243	0.878	0.5929
History of TB Yes vs. No (Reference)	0.026	0.160	1.026	0.8720
Extra pulmonary TB Yes vs. No (Reference)	-0.057	0.362	0.945	0.8748
WHO stage 4 vs. 3 (Reference)	-0.044	0.325	0.957	0.8921
Alcohol = Consumer vs. Non-consumer (Reference)	0.393	0.194	1.481	**0.0424**
CD4Cat $\geq 0.05 \times 10^9$ vs. CD4Cat < 0.05×10^9 (Reference)	0.205	0.186	1.228	0.2707
CD8 count $\times 10^9$ cell/L	0.086	0.092	1.090	0.7032
Log viral load	0.086	0.092	1.090	0.3512
ALK	0.001	0.001	1.001	0.2753
BILI	-0.037	0.017	0.964	**0.0299**
LDH	-0.001	0.001	0.999	0.3079
HBsAg = Positive status vs. HBsAg = Negative status (Reference)	0.697	0.241	2.008	**0.0038**
Baseline LFT= Abnormal vs. Normal (Reference)	-0.117	0.193	0.890	0.5425

Age, sex, alcohol, BILI and HBsAg status were all significant at 10% level significance in the univariable assessment. These variables will then be considered in the next step of the Cox model build.

First multivariable model

A multivariable Cox PH model is fitted based on age, sex, alcohol, bilirubin and HBsAg status, model results are shown in the first four columns of Table 3.3. It is apparent that the sex variable is not significant in the multivariable model. Consequently, this variable will be omitted in the multivariable Cox PH model going forward.

Table 3.3: Multivariable estimation of the preliminary effects model

	Multivariable assessment							
Baseline exposure variable	***Preliminary main effects***				***Preliminary main effects excluding Sex***			
	Parameter estimate	Standard error	Hazard Ratio	p-value	Parameter estimate	Standard error	Hazard Ratio	p-value
Age (years)	0.0252	0.0097	1.026	**0.0094**	0.0261	0.0096	1.026	**0.0063**
Sex = Male vs. Sex = Female (Reference)	0.1085	0.1870	1.115	0.5617				
Alcohol = Consumer vs. Non-consumer (Reference)	0.4448	0.2389	1.560	**0.0626**	0.4944	0.2195	1.640	**0.0243**
BILI	-0.0563	0.0199	0.945	**0.0046**	-0.0553	0.0199	0.946	**0.0054**
HBsAg = Positive status vs. HBsAg = Negative status (Reference)	0.7167	0.2545	2.048	**0.0049**	0.7400	0.2470	2.096	**0.0027**
-2 Log L	1579.086				1579.419			

Elimanating variable that are not significant in the first multivariable model

Sex is excluded in the multivariable model (6^{th}, 7^{th} , 8^{th} and 9^{th} column of Table 3.3), the log likelihood ratio test statistic for excluding sex in the model is 1579.419 - 1579.086 = 0.333 with 1 degrees of freedom, this is not statistically significant. Meaning that the inclusion of sex in the current multivariable model does not improve the model fit.

Re-considering risk factors that were eliminated in the univariable assessment

The multivariable Cox PH model now consists of age, alcohol, BILI and HBsAg status variable. We will now assess the significance of the variables that were discarded at the univariable assessment stage. The significance of these variables will be examined by adding each variable one by one, into the current multivariable Cox PH model (that contains age, alcohol, BILI and HBsAg status). The p-values for this process are shown in Table 3.4. It is important to note that the p-value for the variables that are already in the model (age, alcohol consumed, BILI and HBsAg status) are not presented here, only the p-values for variables of interest are presented Table 3.4.

Table 3.4: Re-considering risk factors that were eliminated in the univariable assessment

Variable added to the model	**p-value***
BMI < 18.5 kg/m^2 vs. ≥18.5 kg/m^2 (Reference)	0.7427
History of TB Yes vs. No (Reference)	0.6260
Extra pulmonary TB Yes vs. No (Reference)	0.7396
WHO stage 4 vs. 3 (Reference)	0.7860
CD4Cat $\geq 0.05 \times 10^9$ vs. CD4Cat $< 0.05 \times 10^9$ (Reference)	0.1708
CD8 count cell/L	0.9874
Baseline Log viral load	0.4809
Baseline ALK	0.7024
Baseline LDH	0.8112
Baseline LFT= Abnormal vs. Normal (Reference)	0.6124
***p-value from a multivariable model containing age, alcohol, BILI and HBsAg status	

None of these variables had improved significance and thus none of these variables will be added to the model. Therefore, the model to be considered forward is unchanged and contains age, alcohol, BILI and HBsAg variable. Possible interactions will now be examined based on this model.

Examining possible interactions effects in the model

All possible interaction or mixed terms that can be formed between age, alcohol consumed, BILI and HBsAg status are considered in the model. Model results are presented in Table 3.5 .

Table 3.5: First review of the multivariable assessment of main and interaction effects

Baseline exposure variables	**Parameter estimate**	**Standard Error**	**Hazard Ratio**	***p* – value**
Age (years)	0.020	0.018	1.020	0.2807
Alcohol = Consumer vs. Non-consumer (Reference)	-0.090	1.163	0.914	0.9385
BILI	-0.059	0.066	0.943	0.3697
HBsAg = Positive status vs. HBsAg = Negative status (Reference)	0.507	1.404	1.660	0.7179
Age * Alcohol = Consumer vs. Non-consumer (Reference)	0.013	0.029	1.013	0.6635
Age * BILI	0.000	0.002	1.000	0.8308
Age * (HBsAg = Positive status vs. HBsAg = Negative status (Reference))	0.039	0.034	1.040	0.2566
Alcohol = Consumer vs. Non-consumer (Reference)* BILI	0.002	0.043	1.002	0.9706
Alcohol = Consumer vs. Non-consumer (Reference)* HBsAg = Positive status vs. HBsAg = Negative status (Reference)	1.825	0.901	6.203	0.0427
BILI * HBsAg = Positive status vs. HBsAg = Negative status (Reference)	-0.171	0.081	0.843	0.0346
-2 Log L	1571.886			

Interaction of alcohol with the HBsAg status and the interaction of BILI with the HBsAg status were significant at 10% level of significance. Subsequently, these terms were added to the model. Model results are presented in Table 3.6.

Table 3.6: Second review of the multivariable assessment of main and interaction effects

Baseline exposure variables	Parameter estimate	Standard Error	Hazard Ratio	*p* – value
Age (years)	0.027	0.009	1.027	0.0042
Alcohol = Consumer vs. Non-consumer (Reference)	0.388	0.229	1.474	0.0906
BILI	-0.046	0.020	0.955	0.0237
HBsAg = Positive status vs. HBsAg = Negative status (Reference)	1.805	0.630	6.080	0.0042
Alcohol = Consumer vs. Non-consumer (Reference)* HBsAg = Positive status vs. HBsAg = Negative status (Reference)	1.719	0.925	5.579	0.0633
BILI * HBsAg = Positive status vs. HBsAg = Negative status (Reference)	-0.159	0.082	0.853	0.0538
-2 Log L	1573.093			

The interaction terms included in the model are all significant Table 3.6. However, interaction term between alcohol and HBsAg status yielded a high standard error, indicating an overfit of the model. Therefore, this interaction term between alcohol and HBsAg will be excluded in the modelling process.

Table 3.7: Third review of the multivariable assessment of main and interaction effects

Baseline exposure variables	Parameter estimate	Standard Error	Hazard Ratio	*p* – value
Age (years)	0.026	0.010	1.026	0.006
Alcohol = Consumer vs. Non-consumer (Reference)	0.513	0.220	1.670	0.0198
BILI	-0.047	0.020	0.954	0.0213
HBsAg = Positive status vs. HBsAg = Negative status (Reference)	1.430	0.604	4.179	0.0180
BILI * HBsAg = Positive status vs. HBsAg = Negative status (Reference)	-0.081	0.071	0.922	0.2536
-2 Log L	1577.822			

The remaining mixed term between BILI and HBsAg status is now statistically not significant and will be excluded in the modeling process (Table 3.7).

Final multivariable Cox PH model

At this point all variables were included in model based on statistical significance and no variable has been included in the model based on clinical justification or widespread practice. Naturally,

treatment arm will be included in the final model since the aim of this project is to explain the association of treatment arm with the hazard of hepatotoxicity. Additionally, sex, CD4 count category and the baseline LFT variable will also be added in the model since these variables have commonly been predictive in literature. The final multivariable Cox model PH is as below

$$\lambda(t, \boldsymbol{X}(t)) = \lambda_0(t)\exp\big(\beta_{Arm_1} \times Arm_1 + \beta_{Arm_2} \times Arm_2 + \beta_{Sex} \times Sex + \beta_{Age} \times Age + \beta_{Alc} \times Alc + \beta_{CD4cat} \times CD4cat + \beta_{Bilirubin} \times Bilirubin + \beta_{HBsAg_{Pos}} \times HBsAg_{Pos} + \beta_{Baseline\ LFT_{Abn}} \times Baseline\ LFT_{Abn}\big). \quad (3.20)$$

The model results of the final multivariable Cox model are presented in Table 3.8

Table 3.8: Final multivariable Cox PH model

Baseline exposure variable	Parameter estimate	Standard error	Hazard Ratio	p-value
Early arm vs. Sequential arm (Reference)	-0.047	0.212	0.954	0.8262
Late arm vs. Sequential arm (Reference)	0.167	0.208	1.182	0.4214
Age (years)	0.024	0.010	1.024	0.0146
Sex = Male vs. Sex = Female (Reference)	0.066	0.193	1.068	0.7341
Alcohol = Consumer vs. Non-consumer (Reference)	0.522	0.250	1.685	0.0372
CD4Cat $\geq 0.05 \times 10^9$ vs. CD4Cat $< 0.05 \times 10^9$ (Reference)	0.336	0.235	1.399	0.1526
BILI	-0.058	0.021	0.944	0.0051
HBsAg = Positive status vs. HBsAg = Negative status (Reference)	0.747	0.259	2.111	0.0040
Baseline LFT= Abnormal vs. Normal (Reference)	-0.212	0.229	0.809	0.3535
-2 Log L	1575.451			

The next step is to assess the adequacy of the final multivariable Cox PH model.

3.10 Model adequacy assessments

Model adequacy assessment is crucial, because it helps us understand how well the estimated model fits the data, before proceeding to model interpretation.

3.10.1. Assessment of the Cox Snell residuals

The Cox Snell residuals for the final multivariable Cox PH model (Table 3.8) are displayed in Figure 3.1. The residuals (blue line) seem to map over the red line reasonably well for earlier events, however, towards the end of the study these lines vary. This suggests that the model fits the observed data quite well at the beginning of the study and fits the observed data poorly

towards the end of the study. This however is common is studies that have right censored observed survival times.

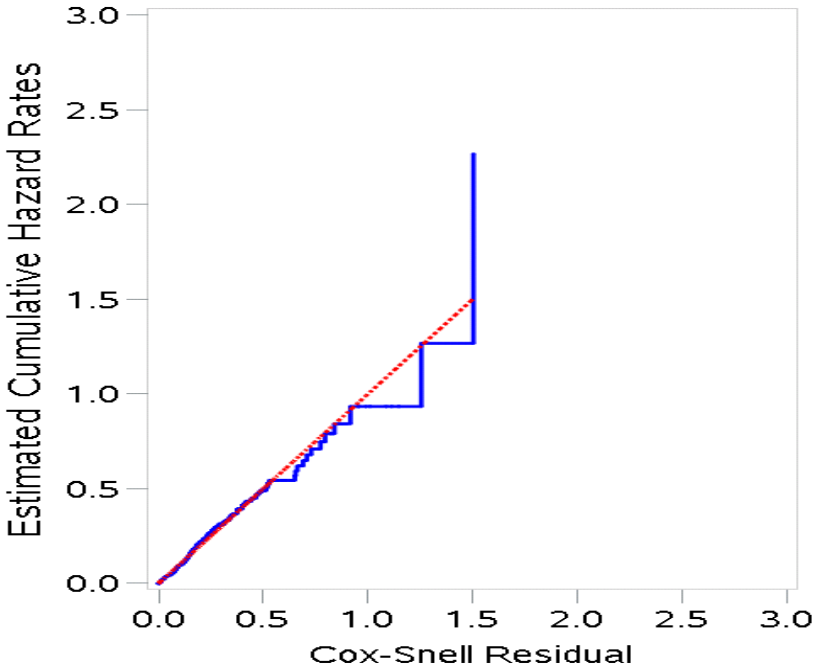


Figure 3.1: Cox Snell residuals

3.10.2. Assessment of the martingale and deviance residuals

The martingale and deviance residual plots for each covariate fitted in the Cox multivariable model are presented in Figure 3.2. Graphical analysis of the martingale and deviance residuals for categorical variables is based on a box and whisker plot per categorical level and a smooth line is fitted on the residuals for continuous variables.

The distribution of the residuals is the same in the early and late arm, and slightly different in the sequential arm (Figure 3.2 (a & b)). Marked in red are the outlier(s), and these are apparent in the sequential arm. Judging by the width size of the box and whisker plot for each treatment arm, there were more observations from the late arm used in the estimation of model parameters compared to the other arms. Recall that the number of observation for each arm was 195, 198, and 200 for the early arm, late arm and sequential arm respectively in Figure 2.2. However, there were observations that had a missing HBsAg status, 37 in the early arm, 22 in late arm and 33 in the sequential arm. Consequently, these observations are discarded in the process of estimating

the modeling parameters. Essentially, a subject is deleted in the modeling sample if they have a missing record for any of the variables included in the model.

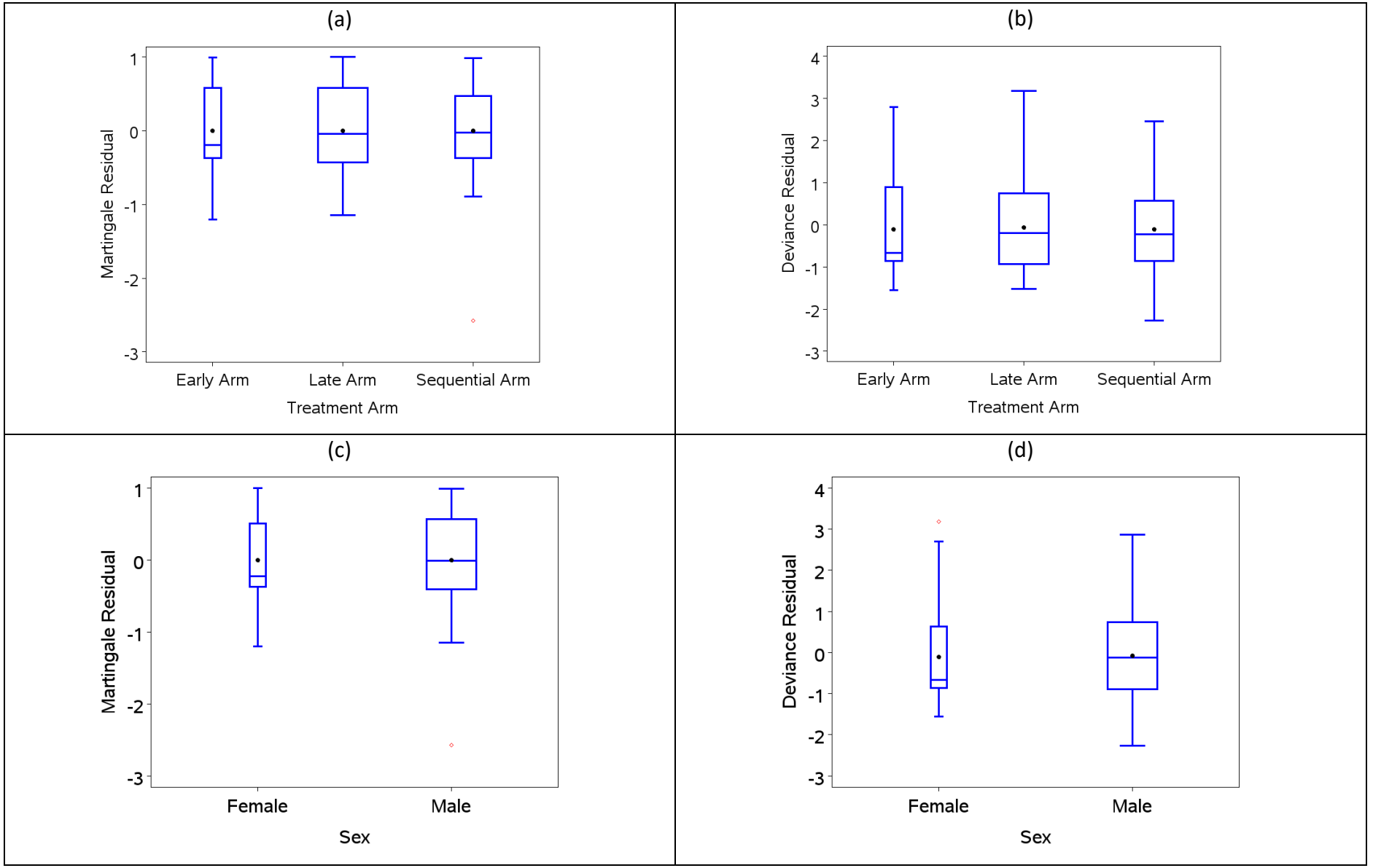
The martingale and deviance residual distribution of the sex variable is same between females and males (Figure 3.2 (c & d)). Additionally, there were more male observations used in the model sample compared to females when comparing distribution of this variable as reported in Table 3.1, for reasons mentioned above.

The martingale residual distribution of age variable reveals that age is linearly related to the hazard of developing hepatotoxicity for ages less than 40 years old, from 40 to 50 years old the model underestimates occurrence of the hepatotoxicity event and beyond 50 years old the model overestimates the occurrence of the hepatotoxicity event (Figure 3.2 (e)). The deviance residual for age displays the same pattern as discussed above, however, the deviation of the smooth line is slightly more pronounced beyond 40 years old (Figure 3.2 (f)).

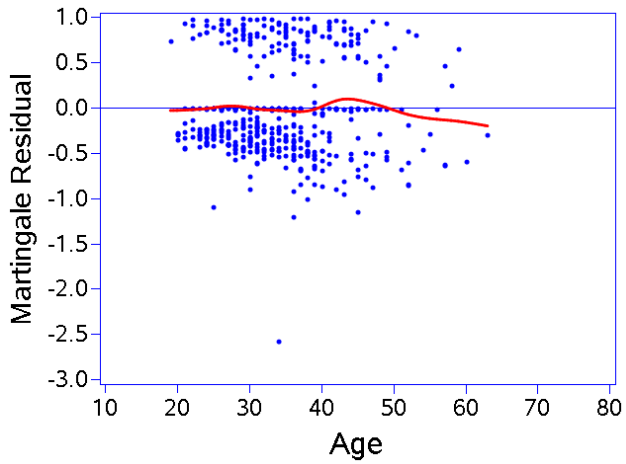
The distribution of the martingale and deviance residuals is the same across the risk levels of the alcohol consumed and the CD4 count category (Figure 3.2 (g, h, i, & j)). Note that, unlike the other categorical variables, observations in each risk level of alcohol consumed and CD4 count category are not distorted by those observations that were discarded in the model estimation as a result of missing HBsAG status.

The smooth line superimposed on the martingale and deviance residuals for BILI is roughly a straight line (Figure 3.2 (k & l)). This implies that BILI is linearly related to log hazard of hepatotoxicity. The number of observations used in the estimation of parameters per risk level of HBsAg, is different depending on the width size of the box whisker plot of HBsAg status (Figure 3.2 (m & n)). However, the distribution of the martingale and deviance residuals is the same across the risk levels based on the mean. Lastly, it can be seen that the model poorly fits the categories of the baseline LFT variable, and there is a large number of outliers, especially for the abnormal LFT category (Figure 3.2 (o & p)).

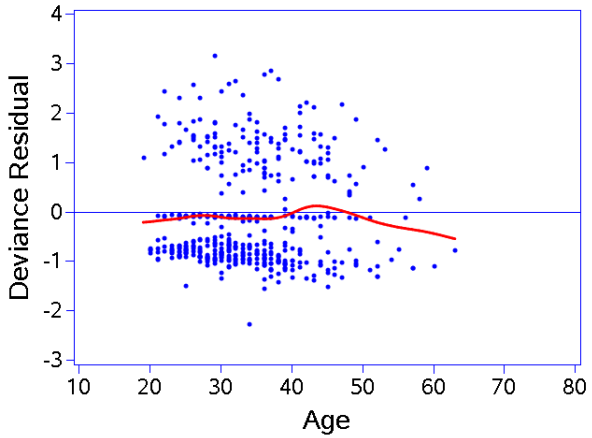
Figure 3.2: Martingale and deviance residual assessment



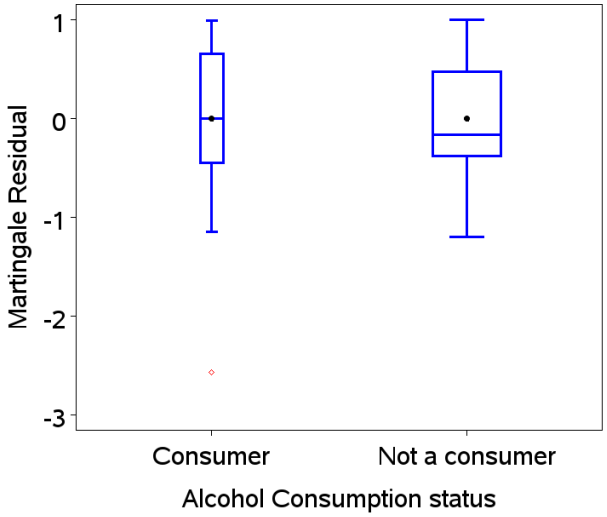
(e)


Martingale Residual
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
10
20
30
40
50
60
70
80
Age

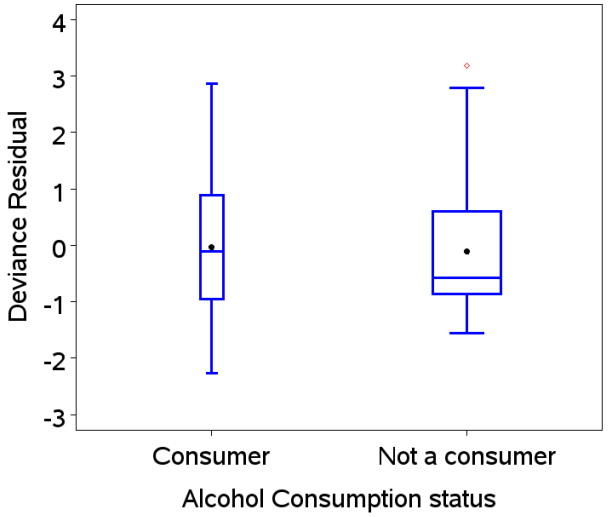
(f)


Deviance Residual
4
3
2
1
0
-1
-2
-3
10
20
30
40
50
60
70
80
Age

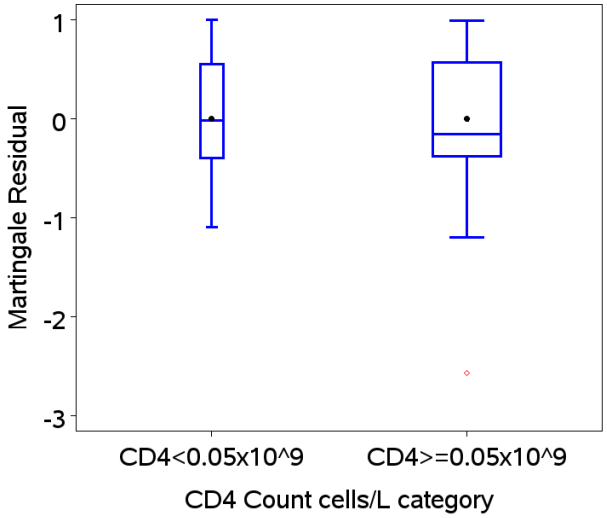
(g)


Martingale Residual
1
0
-1
-2
-3
Consumer
Not a consumer
Alcohol Consumption status

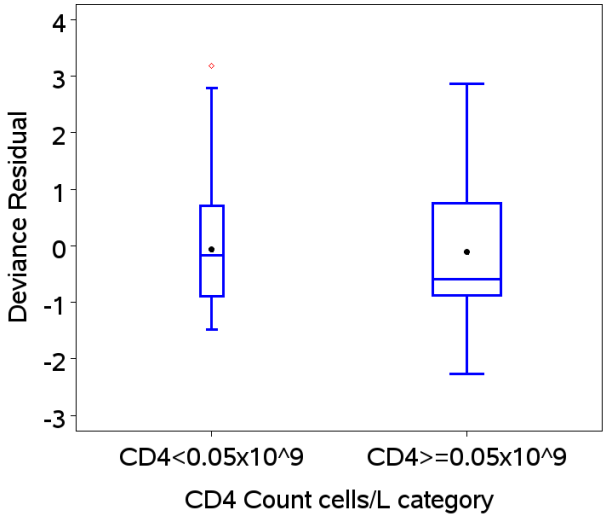
(h)


Deviance Residual
4
3
2
1
0
-1
-2
-3
Consumer
Not a consumer
Alcohol Consumption status

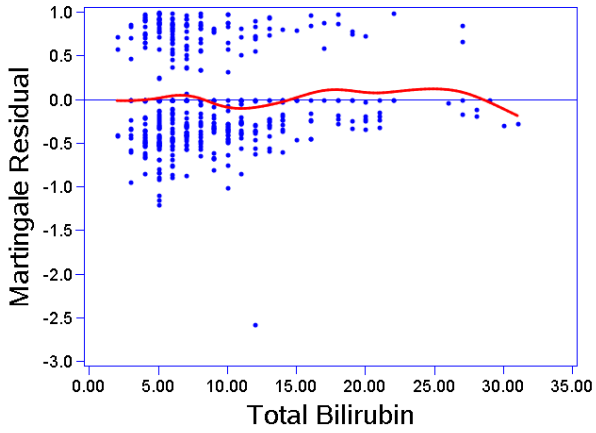
(i)


Martingale Residual
1
0
-1
-2
-3
CD4<0.05x10^9
CD4>=0.05x10^9
CD4 Count cells/L category

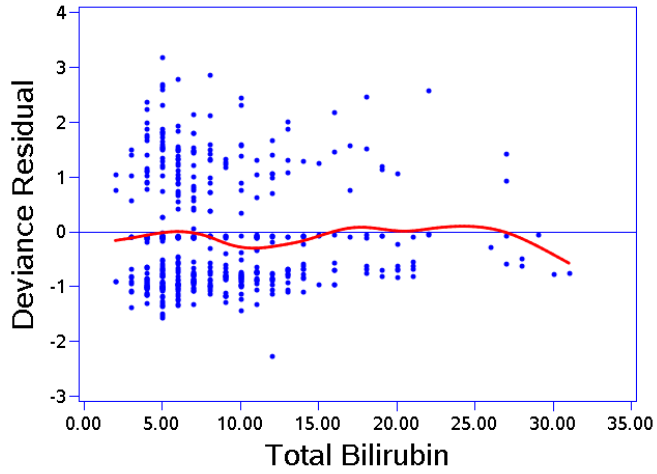
(j)

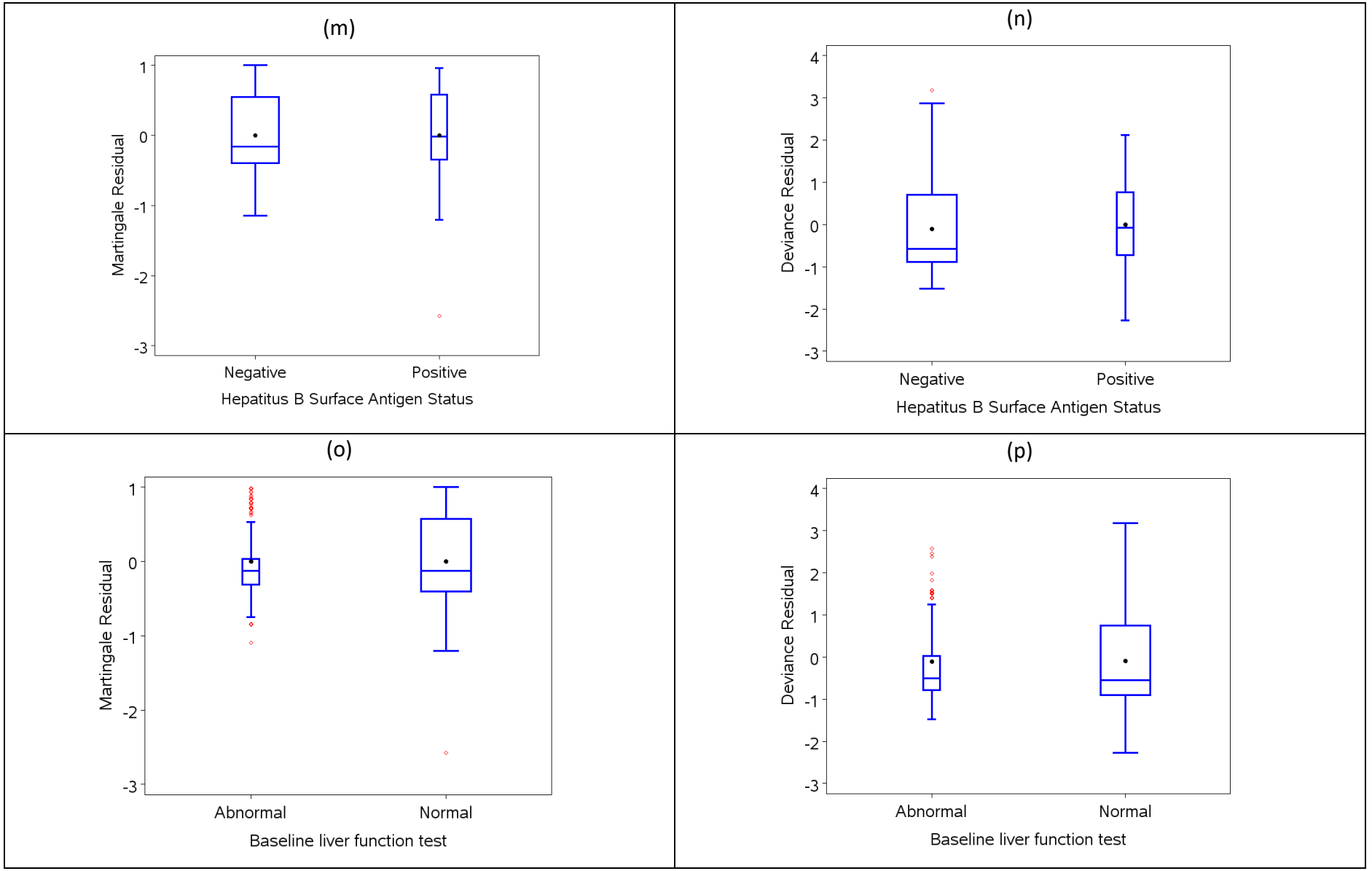

Deviance Residual
4
3
2
1
0
-1
-2
-3
CD4<0.05x10^9
CD4>=0.05x10^9
CD4 Count cells/L category

(k)


Martingale Residual
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
35.00
Total Bilirubin

(l)


Deviance Residual
4
3
2
1
0
-1
-2
-3
0.00
5.00
10.00
15.00
20.00
25.00
30.00
35.00
Total Bilirubin


(m)
Martingale Residual
1
0
-1
-2
-3
Negative
Positive
Hepatitus B Surface Antigen Status
(n)
Deviance Residual
4
3
2
1
0
-1
-2
-3
Negative
Positive
Hepatitus B Surface Antigen Status
(o)
Martingale Residual
1
0
-1
-2
-3
Abnormal
Normal
Baseline liver function test
(p)
Deviance Residual
4
3
2
1
0
-1
-2
-3
Abnormal
Normal
Baseline liver function test

3.11 Proportional hazard assumption test

In this section, the two methods to assess the validity of the proportional hazard assumption are used; namely time dependent covariates and the log-log survival curves method.

3.11.1. Applying the time dependent covariates method

Here the PH assumption is investigated by including interaction terms between the covariates and a function of time ($log\ (t)$), to assess whether this function of time is making a significant contribution. The Cox model that includes time dependent covariate is as below:

$$\begin{aligned}\lambda(\, t, \boldsymbol{X}(t)\,) = \lambda_0(t)\exp\big(&\beta_{Arm_1} \times Arm_1 + \beta_{Arm_2} \times Arm_2 + \beta_{Sex} \times Sex \\ &+ \beta_{Age} \times Age + \beta_{Alc} \times Alc + \beta_{CD4cat} \times CD4cat \\ &+ \beta_{HBsAg_{Pos}} \times HBsAg_{Pos} + \beta_{Bilirubin} \times Bilirubin \\ &+ \beta_{Baseline\ LFT_{Abn}} \times Baseline\ LFT_{Abn} + \varphi_{Arm_1} \times Arm_1 \times log\ (t) \\ &+ \varphi_{Arm_2} \times Arm_2 \times log\ (t) + \varphi_{Sex} \times Sex \times log\ (t) \\ &+ \varphi_{Age} \times Age \times log\ (t) + \varphi_{Alc} \times Alc \times log\ (t) \\ &+ \varphi_{CD4cat} \times CD4cat \times log\ (t) + \varphi_{Bilirubin} \times Bilirubin \times log\ (t) \\ &+ \varphi_{HBsAg_{Pos}} \times HBsAg_{Pos} \times log\ (t) \\ &+ \varphi_{Baseline\ LFT_{Abn}} \times Baseline\ LFT_{Abn} \times log\ (t)\big).\end{aligned} \tag{3.21}$$

In Table 3.9, the results for the estimated extended Cox model (3.21) are presented. The time dependent binary indicator based on Late arm ($\times\ log\ (t)$) vs. Sequential arm ($\times\ log\ (t)$) (Reference), and time dependent binary indicator based Alcohol = Consumer ($\times\ log\ (t)$) vs. Alcohol = Non-consumer ($\times\ log\ (t)$) (Reference), are significant with *p*-values of 0.0139 and 0.0341, respectively. This indicates that the proportional hazard assumption might be violated by the treatment arm and the alcohol variable.

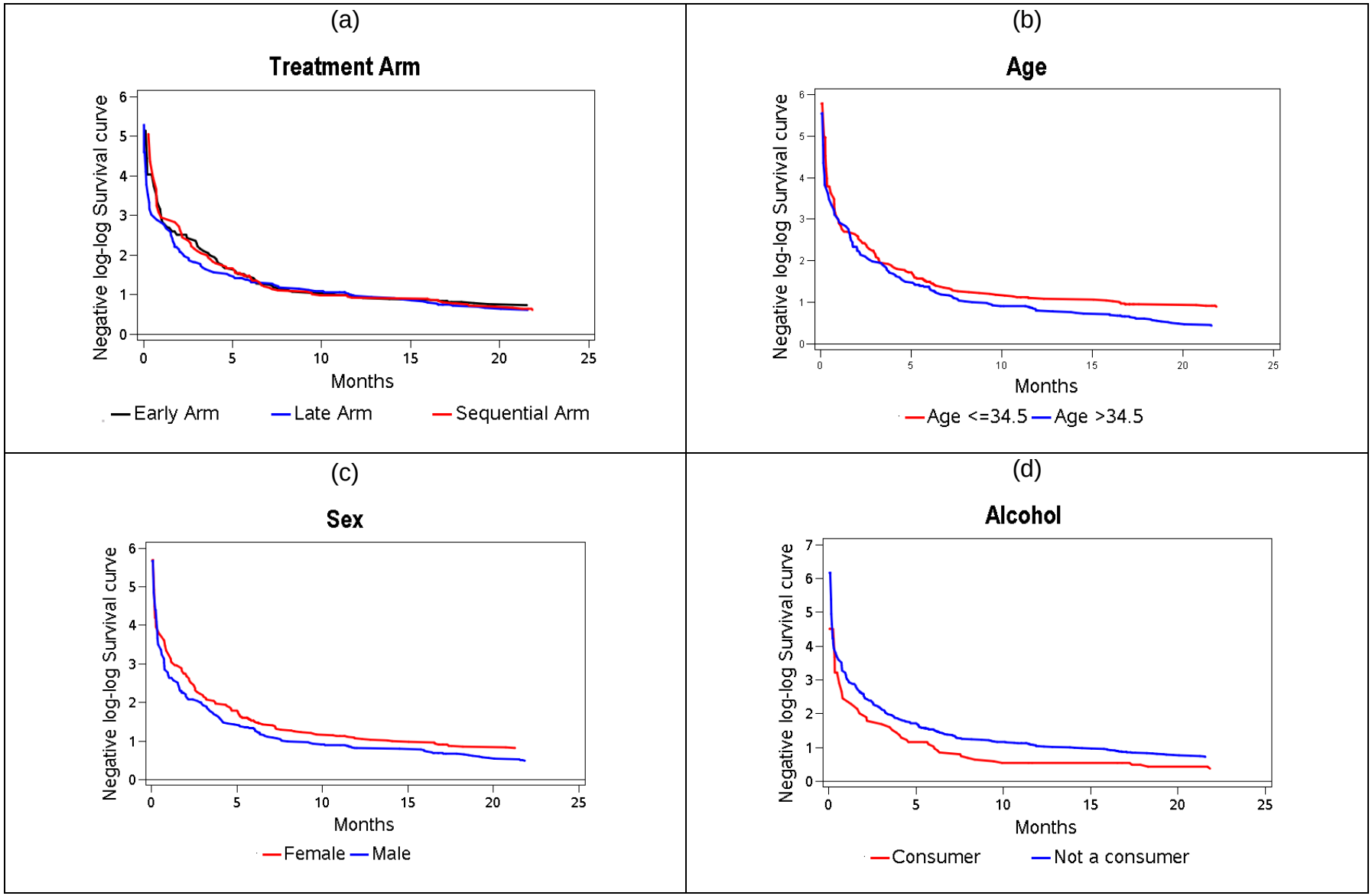
Table 3.9: Estimated Time dependent Cox model

Baseline Exposure	Parameter estimate	Standard Error	Wald Chi-square statistic	**Hazard Ratio**	*p* – value
Early arm vs. Sequential arm (Reference)	0.218	0.342	0.408	1.244	0.5232
Late arm vs. Sequential arm (Reference)	0.750	0.312	5.793	2.117	0.0161
Age (years)	0.007	0.015	0.228	1.007	0.6331
Sex = Male vs. Sex = Female (Reference)	-0.148	0.306	0.233	0.862	0.6294
Alcohol = Consumer vs. Non-consumer (Reference)	0.969	0.316	9.440	2.635	0.0021
CD4Cat $\geq 0.05 \times 10^9$ vs. CD4Cat $< 0.05 \times 10^9$ (Reference)	0.542	0.394	1.892	1.719	0.1690
Bilirubin	-0.111	0.044	6.362	0.895	0.0117
HBsAgPos vs. HBsAgNeg (Reference)	0.770	0.376	4.204	2.160	0.0403
Baseline LFT= Abnormal vs. Normal (Reference)	-0.460	0.423	1.183	0.631	0.2767
Early arm ($\times\ log\ (t)$) vs. Sequential arm ($\times\ log\ (t)$) (Reference)	-0.166	0.171	0.945	0.847	0.3309
Late arm ($\times\ log\ (t)$) vs. Sequential arm ($\times\ log\ (t)$) (Reference)	-0.402	0.164	6.047	0.669	**0.0139**
Age ($\times\ log\ (t)$)	0.149	0.155	0.927	1.161	0.3356
Sex = Male ($\times\ log\ (t)$) vs. Sex = Female ($\times\ log\ (t)$) (Reference)	0.011	0.008	2.018	1.011	0.1554
Alcohol = Consumer ($\times\ log\ (t)$) vs. Non-consumer (Reference) ($\times\ log\ (t)$) (Reference)	-0.322	0.152	4.490	0.725	**0.0341**
CD4Cat $\geq 0.05 \times 10^9$ ($\times\ log\ (t)$) vs. CD4Cat $< 0.05 \times 10^9$ ($\times\ log\ (t)$) (Reference)	-0.137	0.214	0.411	0.872	0.5217
Bilirubin ($\times\ log\ (t)$) ϑ	0.032	0.020	2.467	1.033	0.1163
HBsAgPos ($\times\ log\ (t)$) vs. HBsAgNeg ($\times\ log\ (t)$) (Reference)	-0.044	0.186	0.056	0.957	0.8135
Baseline LFT= Abnormal ($\times\ log\ (t)$) vs. Normal ($\times\ log\ (t)$) (Reference)	0.164	0.213	0.594	1.178	0.4410

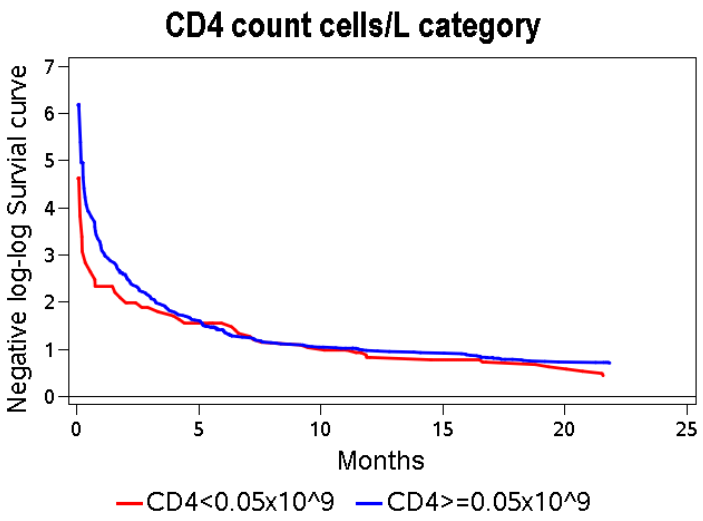
3.11.2. Graphical approach assessment of the PH assumption

Recall that this method plots $-ln(-\ln \hat{S}(t))$ for each category of a covariates against observed survival time t, if the curves intersect and or the curves show a lack of parallelism, then that implies that the covariate under investigation violates that proportional hazard assumption. The plots of the negative log-log survival curves per covariate fitted in the Cox multivariable are presented in Figure 3.3.

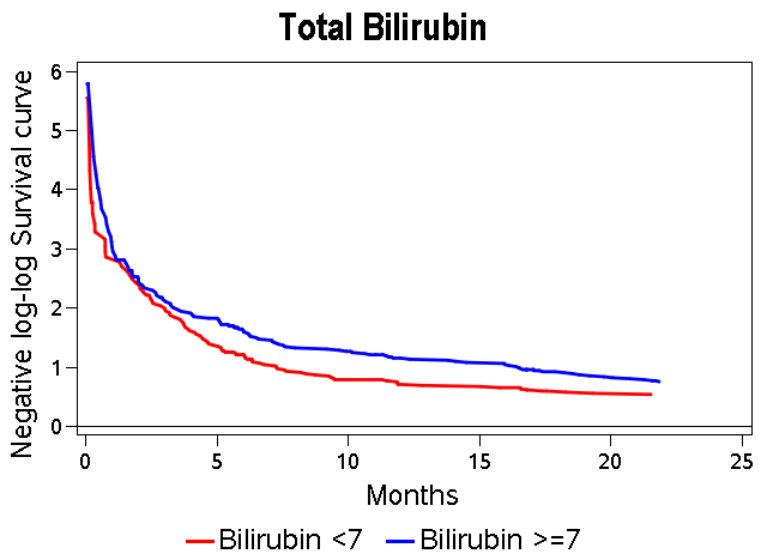
Figure 3.3: Negative log-log survival curves



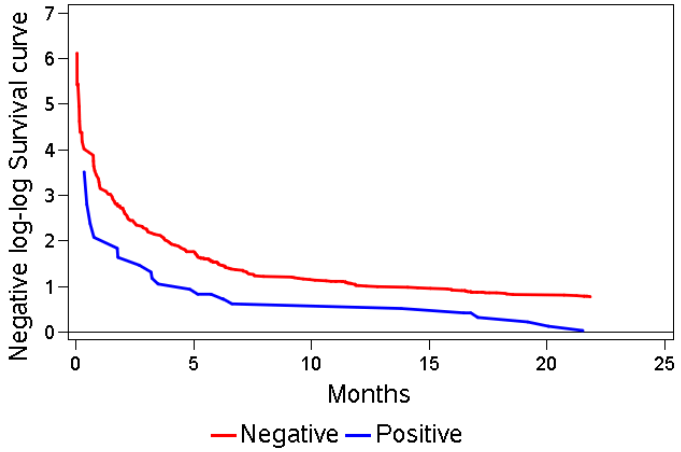
(e)


CD4 count cells/L category
Negative log-log Survial curve
0
1
2
3
4
5
6
7
0
5
10
15
20
25
Months
CD4<0.05x10^9
CD4>=0.05x10^9

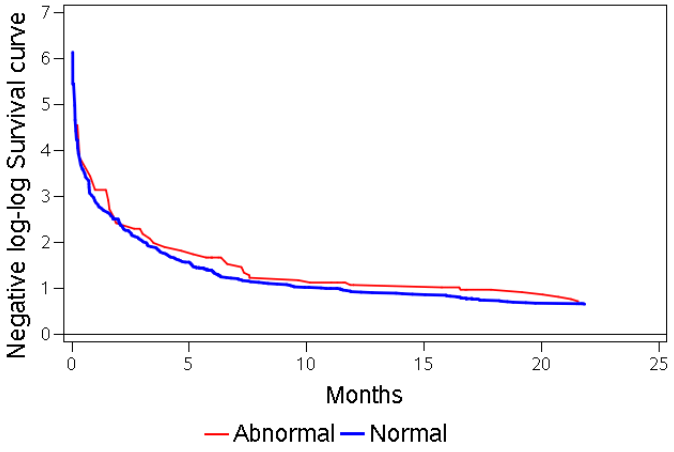
(f)


Total Bilirubin
Negative log-log Survival curve
0
1
2
3
4
5
6
0
5
10
15
20
25
Months
Bilirubin <7
Bilirubin >=7

(g)


Hepatitus B surface Antigen status
Negative log-log Survival curve
0
1
2
3
4
5
6
7
0
5
10
15
20
25
Months
Negative
Positive

(h)


Baseline Liver Function Test
Negative log-log Survival curve
0
1
2
3
4
5
6
7
0
5
10
15
20
25
Months
Abnormal
Normal

The negative log-log survival curves for each risk level of age, sex, alcohol and HBsAG status show reasonable level of parallelism (Figure 3.3 (b, c, d & g). In contrast, the negative log-log survival curves for each risk level of treatment arm, CD4 count category, BILI (risk levels based on a median of 7) and baseline LFT all show little parallelism (Figure 3.3 (a, e, f & h). The proportionality hazard assessment based on negative log-log survival curves for treatment arm, is agreeing with the results obtained in time dependent covariate approach. This suggests that the treatment arm is violating the proportional hazard assumption. Therefore, the treatment arm variable should be modelled as time dependent variable in the model going forward.

There is conflicting evidence from the two approaches for alcohol, CD4 count category, BILI and Baseline LFT, therefore, there is not enough evidence to suggest that these covariates are violating the proportional hazards assumption.

The graphical approach of assessing the proportional hazard assumption is quite subjective, because there is no specification of how parallel the curves should be to justify a violation of the proportional hazard assumption.

3.12 The Cox model with non-proportional hazards

3.12.1. Choosing the appropriate function of time

Including the treatment arm variable as time dependent variable, requires the choice of a function of time. It can be seen from the negative log-log survival curves, that before 5 months, the late arm has the poorest survival rates compared to the other treatment arms. Beyond 5 months the survival rate in the late arm improves and then eventually matches the survival level of the other treatment arm.

3.12.2. Defining the heavyside function

Recall that the heavyside functions is a function that defines intervals so that the constant hazards can be estimated within each interval. A two sided heavyside functions will allow for estimation of different hazard in each interval defined by the heavyside function for the treatment arm variable. The heavyside function in equation (3.22) that will be used is as follows:

$$g_1(t) = \begin{cases} 1 & \text{if } t \leq 5 \\ 0 & \text{if } t > 5 \end{cases} \tag{3.22}$$

$$g_2(t) = \begin{cases} 1 & \text{if } t > 5 \\ 0 & \text{if } t \leq 5 \end{cases} \tag{3.23}$$

where the 5th month is the chosen cut-off point to define two intervals in the study. Conceptually when the observed survival time is 5 months, or less ($t \leq 5$), the heavyside function takes $g_1(t) = 1$ and $g_2(t) = 0$. And when the observed survival time is greater than 5 months ($t > 5$), the heavyside function takes $g_1(t) = 0$ and $g_2(t) = 1$. A different hazard will be estimable for the treatment arm in each interval $t \leq 5$ and when $t > 5$.

The choice of 5 months will facilitate the comparison of the hazard by treatment arm during the intensive phase of TB treatment through the function $g_1(t)$. Moreover, the second function $g_2(t)$ will facilitate comparison of the hazards during the continuation phase of the TB treatment and after completion of the TB treatment.

3.12.3. Fitting the Cox model with non-proportional hazards

A Cox model with non-proportional hazards, of the treatment arm variable, can be built by including interactions with the specified function of time. Therefore, the indicator variable Arm_1 in equation expands to $Arm_1 \times g_1(t)$ or $Arm_1 \times g_2(t)$. Moreover, the indicator variable Arm_2 expands to $Arm_2 \times g_1(t)$ or $Arm_2 \times g_2(t)$ in the Cox regression model that was specified in equation (3.20). The Cox regression model with a time dependent effects takes the form:

$$\begin{aligned} \lambda(\, t, \boldsymbol{X}(t)\,) = \lambda_0(t) \exp\big(&\varphi_{Arm_{11}} \times Arm_1 \times g_1(t) + \varphi_{Arm_{12}} \times Arm_1 \times g_2(t) \\ &+ \varphi_{Arm_{21}} \times Arm_2 \times g_1(t) + \varphi_{Arm_{22}} \times Arm_2 \times g_2(t) \\ &+ \beta_{Sex} \times Sex + \beta_{Age} \times Age + \beta_{Alc} \times Alc + \beta_{CD4cat} \times CD4cat \\ &+ \beta_{CD8} \times CD8 + \beta_{HBsAg_{Mis}} \times HBsAg_{Mis} + \beta_{Bilirubin} \times Bilirubin \\ &+ \beta_{HBsAg_{Pos}} \times HBsAg_{Pos} + \beta_{Baseline\ LFT_{Abn}} \times Baseline\ LFT_{Abn}\big). \end{aligned} \tag{3.24}$$

The estimated Cox PH and Cox regression model multivariate model are presented in Table 3.10.

Table 3.10: Multivariable Cox PH and Cox model with non-proportional hazards

Baseline exposure Variables	Cox PH model		Cox regression	
	Hazard Ratio (95% CI)	**p – value**	**Hazard Ratio (95% CI)**	**p – value**
Early arm vs. Sequential arm (Reference) (0 ≤ months ≤ 5)	0.954 (0.629; 1.447)	0.8262	0.973 (0.515; 1.839)	0.9335
Early arm vs. Sequential arm (Reference) (months > 5)			0.933 (0.534; 1.629)	0.8074
Late arm vs. Sequential arm (Reference) (0 ≤ months ≤ 5)	1.182 (0.787; 1.775)	0.4214	1.648 (0.923; 2.940)	0.0912
Late arm vs. Sequential arm (Reference) (months > 5)			0.861 (0.494; 1.502)	0.5986
Age (years)	1.025 (1.005; 1.045)	0.0146	1.025 (1.005; 1.045)	0.0139
Sex = Male vs. Sex = Female (Reference)	1.068 (0.732; 1.558)	0.7341	1.061 (0.724; 1.554)	0.7618
Alcohol = Consumer vs. Non-consumer (Reference)	1.685 (1.032; 2.753)	0.0372	1.702 (1.039; 2.789)	0.0346
CD4Cat $\geq 0.05 \times 10^9$ vs. CD4Cat $< 0.05 \times 10^9$ (Reference)	1.399 (0.883; 2.218)	0.1526	1.417 (0.892; 2.251)	0.1397
Bilirubin	0.944 (0.907; 0.983)	0.0051	0.944 (0.907; 0.983)	0.0050
HBsAgPos vs. HBsAgNeg (Reference)	2.11 (1.269; 3.508)	0.0040	2.097 (1.262; 3.483)	0.0042
Baseline LFT= Abnormal vs. Normal (Reference)	0.809 (0.517; 1.266)	0.3535	0.807 (0.515; 1.265)	0.3499

3.12.4. Model fit statistics

Table 3.11 shows the model fit statistics, the Akaike Information Criterion (AIC) statistic is reduced in the Cox model with non-proportional hazards compared to the Cox PH model. This indicates that the Cox model with non-proportional hazards fits the data better when the treatment arm is included in the model as having time dependent effects on the hazard of developing hepatotoxicity.

Table 3.11: Model Fit Statistics		
Criterion	**Cox PH model**	**Cox model with non-proportional hazards**
AIC	1575.451	1572.245

3.12.5. Interpreting the model Cox model with non-proportional hazards

Within the first 5 months, patients in the early arm have 2.7% decreased hazard of developing hepatotoxicity, then patients in the sequential arm, the hazards ratio is (HR=0.973, 95%CI (0.515, 1.839)), and this ratio is very close to unity and thus not statistically significant (p-value = 0.9335). However, within the same period, the adjusted hazard of developing hepatotoxicity is 1.6 times more for patients in late arm compared to patients in the sequential arm (HR=1.648, 95% CI (0.923, 2.940)) and a p-value of 0.0912.

After 5 months, the hazard of developing hepatotoxicity is lower in both the early and late arm compared to the sequential arm. During this period patients in the early arm have a 6.7% decreased hazard of developing hepatotoxicity (HR=0.933, 95% CI (0.534, 1.629)) compared to patients in the sequential arm, however this not statistically significant. Also in the late arm, the hazard of developing hepatotoxicity is reduced by 13.9% compared to the hazard of developing hepatotoxicity in patients from the sequential arm (HR=0.861, 95% CI (0.494, 1.502)), this is not statistically significant (*p*-value = 0.5986).

A year increase in the subjects age at baseline increases the hazard of developing hepatotoxicity by 2.5% (HR=1.025, 95% CI (1.005, 1.045)), statistically significant (*p*-value = 0.0139). The assessment of sex and CD4 count did not give compelling evidence that suggests that these variables affect the hazard of developing hepatotoxicity.

Subjects who consumed alcohol are 1.7 times more likely to develop hepatotoxicity compared to subjects who never consume alcohol (HR=1.702, 95% CI (1.039, 2.789)) this is statistically significant (*p*-value 0.0346).

An unit increase in the total bilirubin decreased the hazard of developing hepatotoxicity 5.6%, the 95% confidence interval for the hazard ratio does not include unity (HR = 0.944, 95% CI (0.907, 0.983)), and the confidence interval is narrow, this implies that the hazard ratio is statistically significant and precise. The hazard of developing hepatotoxicity for patients with a positive HBsAg was 2-fold the hazard of developing hepatotoxicity for patients with a negative HBsAg (HR=2.097, 95% CI (1.262, 3.483)) this is statistically significant (*p*-value = 0.0042).

The assessment of baseline LFT did not give statistical evidence that suggests that this variable affects the hazard of developing hepatotoxicity.

3.13 Conclusion

Upon fitting the Cox model based on the backward elimation procedure, the final model consisted of treatment arm, age, sex, alcohol consumed, CD4 count, BILI, HBsAg status and baseline LFT. Model adequacy evaluations were within expectactions, except for the baseline LFT variable, which had too many outliers under martingale and deviance residuals assesments. However, no modification was made to the model post model adequacy assessment.

The outcome from assessing the validity of PH assumption indicated that the treatment arm variable violates the PH assumption. Some authors do not stress the importance of PH assumption, as they describe the resulting estimate as being the 'average' effect on the hazard (Allison, et al., 2010). However, in this project, since the main objective is to describe the relationship of treatment arm and the occurrence of hepatotoxicity, the intention was to estimate the most granular effect of treatment arm that the model could estimate.

The treatment arm variable was then modified in the Cox model for non-proportional hazards. There is a choice of fitting a stratified Cox model to address the violated PH assumption, however, the shortfall of fitting a stratified Cox model is that effect of the treatment arm variable cannot be quantifed. Therefore, the effects of treatment arm were modeled as time varying, through the operation of a heavyside function that divides the observed time into two intervals, at a 5-month cut-off.

The choice of the heavyside function was motivated by the pattern observed in the Kaplan-Meier (Figure 2.3) and the Cumulative incidence function (Figure 2.4). This highlight's the importance of plotting the Kaplan-Meier and cumulative incidence function, as it has enhanced the resulting Cox model that is fitted. The Cox model with the time-varying effect of treatment arm resulted in

an AIC =1568.759, this is lower than the AIC of 1572.245 that was observed in the Cox model that modeled the effect of the treatment arm as fixed over time (Table 3.11).

After considering the model modifications, the late arm appeared as the group with the highest risk of hepatotoxicity. In the first 5 months, patients in the late arm were 2 times more likely to develop hepatotoxicity, this statistic was marginally significant with p-value =0.0912. The 95% CI of this estimate is quite wide (0.923, 2.940), thus this statistic might be slightly imprecise. The risk of hepatotoxicity was similar between the treatment arms after 5 months of study randomization. Despite this, it is important to note that, the hazard ratios for treatment arm estimated for the period beyond 5 months, will not be comparable to hazard ratios estimated under a randomized cohort. Since after 5 months, high risk patients would have already experienced hepatotoxicity.

Furthermore, age, alcohol, total bilirubin and HBsAg status remained strong predictors of the occurrence of first hepatotoxicity, under the univariate and multivariable analysis. The outcome of these variables is consistent with what has been observed in literature (Hoffmann, et al., 2007)). Sex, CD4 count category and baseline LFT variable were not statistically predictive in the model, contrary to what has been observed in the literature (Shu, et al., 2013).

Through the Cox regression model risk factors associated with the first occurrence of hepatotoxicity have been identified. However, the Cox regression model as introduced in this chapter has failed to simultaneously identify factors associated with resolving hepatotoxicity, and factors associated with recurring hepatotoxicity. Thus, the following chapter will attempt to address this shortfall.

Chapter 4

Alternative models for recurrent hepatotoxicity events

4.1 Introduction

The Cox model introduced in Chapter 3, considers timing of the first hepatotoxicity event and determines the effect of the treatment arm and other associated risk factors on this timing. However, it is apparent in the data, that some patients experienced a recurrent episode of hepatotoxicity (Figure 2.2). The application of the Cox model thus far used information up to the first hepatotoxicity event, and discarded information that occurred at subsequent events.

Recall that the purpose of the SAPiT trial was to effectively integrate TB treatment and ART, in patients co-infected with TB and HIV. The primary objective of integrating TB treatment and ART in the SAPiT trial, was to determine a treatment arm that reduces the rates of mortality (Abdool Karim, et al., 2010). The secondary objective of this trial was to evaluate the patient's quality of life, more generally, by examining per treatment arm, how prone the patient is to opportunistic infections – like hepatotoxicity. If a patient experiences more than a single hepatotoxicity event, it is sensible therefore, to evaluate not only risk factors associated with the first event, but also risk factors associated with subsequent events to better understand the patient's quality of life over time.

The Cox PH Model has been generalized to analyze recurring events – by modifying four features of the Cox PH model (Ullah, et al., 2012):

- The definition of a risk interval along the study period and as consequence whether a model is marginal or conditional.
- The definition of a risk set at a certain point in time.
- Whether the model will have an event specific or a common baseline hazard.
- Incorporation of within subject correlation.

In light of the above-mentioned features of recurrent events models, five models will be discussed in this current study, namely the Andersen Gill (AG) model, the shared frailty model,

the Prentice-Williams-Petersen total time (PWP-TT) model, the Prentice-Williams-Petersen gap time (PWP-GP) model, and the transition and reverse transition models (recurrent and resolution models) as specified by Islam (1994). I will give a brief introduction of each model and the dynamics of each model in explaining the recurrent events.

4.2 Islam's multi-state models

4.2.1. Description of the model

The experience of a patient in a survival study may be modelled as a process with two states and one possible transition from an "alive" state to a "dead" state, this model is also called a two-state mortality model (Meira-Machado, et al., 2009). Multi-state survival models evolved from the two-state mortality model, where the intensity (hazard) function $\lambda(t)$ is modeled by a Cox PH model introduced by Cox in 1972. Then shortly after Cox, Prentice et al. (1978), extended the theory of the Cox PH model to the analysis of failure times with competing causes of failure. They proved that the cause-specific hazard functions are estimable in the competing risk framework under a multi-state process.

Kay (1982) extended the competing causes of failure model as formulated by Prentice et al. (1978), to a more general multi-state stochastic process allowing several transient states between entry and death. That is, if there is more than one transient state, subjects are open to transition from one state to another during a study follow up. Kay's model was hierarchical in nature and the partial likelihood is identical to the partial likelihood presented by Cox in 1972, except for the definitions of the risk sets (Islam, 1994).

Islam (1994) developed a model framework, by extending Kay's model for reverse and repeated transitions. Islam's extension is suitable for the study of multi-state hepatotoxicity transitions as they occur in the SAPiT trial. Given that there is quite a significant rate of reverse and repeated transitions in the trial, Islam's framework makes it convenient to explain the dynamics of the hepatotoxic event in the data in more dimensions.

4.2.2.Model assumptions

Figure 4.1 below is a graphical illustration of the hepatotoxicity states and transitions that are possible in the data, where $H_1 \rightarrow H_2$ represents a transition (transition from a non-hepatotoxicity state to a hepatotoxicity state); $H_2 \rightarrow H_1$ represents a reverse transition (transition from a hepatotoxicity state to a resolved hepatotoxicity state); $H_1 \rightarrow H_2$ represents a repeated transition, on condition that $H_1 \rightarrow H_2$ has occurred at least once (transition from a resolved hepatotoxicity state to a repeated hepatotoxicity state); and $H_1 \rightarrow C$ or $H_2 \rightarrow C$

indicate a transition from a hepatotoxicity state or resolved hepatotoxicity state to a censored state, which is a state reached if a patient dies or is lost to follow up during the study period.

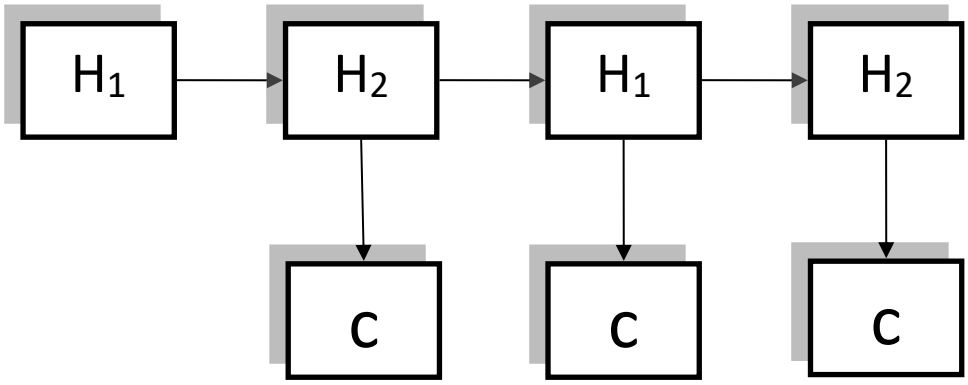


Figure 4.1:Transition dynamics of hepatotoxicity episodes

Therefore, three models will be considered as defined below:

- Model 1 – refers to the incidence of $H_1 \rightarrow H_2$ this is identical to the Cox PH model discussed in Chapter 4.
- Model 2 - refers to the incidence of $H_2 \rightarrow H_1$, which will be called the resolution model hereafter.
- Model 3 - refers to the repeat of $H_1 \rightarrow H_2$ on condition that $H_1 \rightarrow H_2$ has occurred, and this will be called the recurrent model hereafter.

4.2.3. Risk interval and risk set

In Table 4.1 the format of the data that will used to build the models in this section is presented.

The following fields are included in Table 4.1;

- participant id is a unique identifier for each subject,
- tstart is the time at which the observation of the subject starts,
- tstop is the time at which the subject developed hepatotoxicity.
- Event is the event indicator 1 if the subject developed hepatotoxicity and 0 otherwise.
- Resolved is 1 if a patient has resolved hepatotoxicity and 0 otherwise.
- And the last column is called ‘ model number’ indicates the model number in which the specific observation will be used in.

It is important to note that I will discard information where the patient has resolved their second hepatotoxicity event as there were too few of these observations. This is illustrated by a record

from participant_id = 121002 in Table 4.1 where the model number for this record is indicated by N/A, indicating that this record will not considered in any of the models.

Table 4.1: Islam's model framework data structure

Participant ID	Tstart	Tstop	Time	Event	Resolved	Arm	Model number
121002	01/Jan/2007	31/Mar/2007	3	1	.	Late Arm	1
121002	**31/Mar/2007**	**10/Jun/2007**	**2**	**0**	**1**	**Late Arm**	**2**
121002	10/Jun/2007	30/Sep/2007	4	1	0	Late Arm	3
121002	30/Sep/2007	18/Nov/2007	2	0	1	Late Arm	N/A
121002	18/Nov/2007	01/Jan/2008	1	0	0	Late Arm	3
121139	01/Jan/2007	30/Sep/2007	9	1	.	Late Arm	1
121139	**30/Sep/2007**	**01/Nov/2007**	**1**	**0**	**1**	**Late Arm**	**2**
121139	01/Nov/2007	01/Mar/2008	4	0	0	Late Arm	3
121170	01/Jan/2007	31/Jan/2007	1	1	.	Sequential Arm	1
121170	**31/Jan/2007**	**17/Jun/2007**	**4**	**0**	**1**	**Sequential Arm**	**2**
121170	17/Jun/2007	01/Apr/2008	9	0	0	Sequential Arm	3
121178	01/Jan/2007	31/Jan/2008	13	1	.	Late Arm	1
121178	**17/Jun/2007**	**07/Jun/2008**	**12**	**0**	**1**	**Late Arm**	**2**
121178	07/Jun/2008	30/Sep/2008	4	0	0	Late Arm	3
121179	01/Jan/2007	01/Jan/2009	24	1	.	Early Arm	1
121179	**01/Jan/2009**	**12/Feb/2009**	**1**	**0**	**1**	**Early Arm**	**2**
121179	12/Feb/2009	01/Apr/2009	2	0	0	Early Arm	3

4.2.4. Model form

Notation used here will be as employed by (Islam, 1994). Let f $(f = 1, 2..)$ be the state of origin and g $(g = 1, 2...)$ the state of destination. The general hazard function model is written as:

$$\lambda_{fgi} = \lambda_{0fg}(t) \exp\left(\boldsymbol{\beta}'_{fg}\boldsymbol{x}_{fgi}\right). \tag{4.1}$$

Here

- λ_{fgi} is the hazard function for moving from state f to state g,
- $\lambda_{0fg}(t)$ denotes the baseline hazard function specific for sequential move from state f to state g,
- $\boldsymbol{x}_{fgi}$ represents covariates for i^{th} subject in respect of moving from state f to state g,
- $\boldsymbol{\beta}_{fg}$ represents the corresponding regression coefficients.

4.2.5. Likelihood function

It follows that the partial likelihood for estimating $\boldsymbol{\beta}_{fg}$ is as follows;

$$L(\beta_{fg}) = \prod_{i=1}^{n} \left\{ \frac{\exp\left(\boldsymbol{\beta}_{fg}^{'} \boldsymbol{x}_{fgi}(t_{fgi})\right)}{\sum_{s \in R(t_{fgs}, f)} \exp\left(\boldsymbol{\beta}_{fg}^{'} \boldsymbol{x}_{s}(t_{s})\right)} \right\}, \quad (4.2)$$

where $R(t_{fgs}, f)$ denotes the risk set of patients who are in state f, thus are at risk of making a transition to state g.

4.3 Andersen Gill model

4.3.1. Description of the model

Andersen & Gill (1982) proposed the Andersen-Gill (AG) counting process model, which models timing of recurrent events. This model, extends the Cox PH model by using more information about the subject. This is, subjects not only contribute the survival time of the first event into the model, but also survival times related to subsequent events.

This model simplifies the approach of examining the overall treatment effect, when events are likely to occur recurrently. There is also minimal effort of transitioning from a commonly applied Cox PH model to the AG model. Moreover, this model is a building block to more sophisticated recurring events model - to be discussed in proceeding sections. It thus, plays a role in explaining the intermediate step that is found between applying the Cox PH to applying a more sophisticated recurrent events model.

4.3.2. Model assumptions

The AG model assumes that the multiple event times from the same subject are independent. Suppose that a subject develops their first event, this model assumes that, this does not change the subject's likelihood of experiencing successive events. Hence, a common baseline hazard function for each possible event is assumed and a global parameter is estimated for the factor of interest (Amorim & Cai, 2014).

If there is correlation between events from the same subject, and it is ignored or unknown, then the confidence intervals for the estimated rates could be artificially narrow, and the null hypothesis is rejected more often than it should (Amorim & Cai, 2014). This issue could be averted in one of two ways, first by assuming that the correlation between events is measured by a covariate, thus the dependence is captured by specifying a time-varying covariate, such as the number of previous events, or some function of previous events (Sagara, et al., 2014). Secondly, Lin & Wei (1989) proposed a sandwich robust standard error: this is a variance-correction technique which is employed together with the Cox extended model, to avoid

inflation of Type I error, due to multiple observation per subject which do not require specification of a correlation matrix (Sagara, et al., 2014).

4.3.3. Risk interval and risk set

An example in Table 4.2 to illustrate the data structure, by specifically defining the risk interval and risk set construct of the AG model.

Table 4.2: AG model data structure

Participant ID	Tstart	Tstop	Event	Arm
121002	**01/Jan/2007**	**31/Mar/2007**	**1**	**Late Arm**
121002	10/Jun/2007	30/Sep/2007	1	Late Arm
121002	18/Nov/2007	01/Jan/2008	0	Late Arm
121139	**01/Jan/2007**	**30/Sep/2007**	**1**	**Late Arm**
121139	01/Nov/2007	01/Mar/2008	0	Late Arm
121170	**01/Jan/2007**	**31/Jan/2007**	**1**	**Sequential Arm**
121170	17/Jun/2007	01/Apr/2008	0	Sequential Arm
121178	**01/Jan/2007**	**31/Jan/2008**	**1**	**Late Arm**
121178	07/Jun/2008	30/Sep/2008	0	Late Arm
121179	**23/Nov/2008**	**01/Jan/2009**	**1**	**Early Arm**
121179	12/Feb/2009	01/Apr/2009	0	Early Arm

The risk set of a Cox PH model comprises of the observations that are in bold, whereas the risk set of the AG model comprises of all the observations that have been recorded in the longitudinal data for each subject – i.e. every record in Table 4.2.

The survival time (risk interval) variable is discontinuous in nature (Gisolf, et al., 2000). This means that the time in which the subject sustains hepatotoxicity is excluded, since the subject is not at risk of hepatotoxicity at that point. For example, in Table 4.2 subject 121139 developed the first hepatotoxicity event on the 30th of September 2007, and this hepatotoxicity episode was not resolved until 1st of November 2007. The time between 30th of September 2007 and 1st of November 2007 is not included in the analysis, since the subject already has hepatotoxicity in this time and therefore not at a risk of developing hepatotoxicity.

The notation that is used to describe the models from now onwards is as adopted by Sagara, et al., (2014).

4.3.4. Model form

The hazard function for the AG model for subject i can be specified as below:

$$\lambda_{ik} = Y_{ik}\lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ik}). \quad (4.3)$$

Here:

- λ_{ik} represents the hazard of the k^{th} event for the i^{th} subject.
- $\lambda_0(t)$ represents the common baseline hazard function over time.
- $\boldsymbol{x}_{ik}$ is the vector of explanatory variables for subject i relevant to the k^{th} event.
- $\boldsymbol{\beta}$ represents a vector of global regression coefficients.
- Y_{ik} represents a predictable process taking value [0,1] indicating when a subject is under observation.

The model in equation (4.3) is identical to a standard Cox PH model in equation (3.2). The difference lies in the definition of Y_{ik}. In the Cox PH model, once the individual experiences an event, Y_{ik} goes from one to zero. Whereas for the AG model Y_{ik} remains one as the events occur (Therneau & Grambsch, 2000).

4.3.5. Likelihood function

To estimate the regression coefficients let $t_1, \dots, t_n$ be n ordered possible survival times, corresponding to vectors of explanatory variables $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$. The partial likelihood to estimate the regression coefficients is as below:

$$L(\beta) = \prod_{i=1}\prod_{k=1}\left(\frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{\sum_{i=1}\sum_{k=1} Y_{ik}\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}\right)^{\delta_{ik}}, \quad (4.4)$$

where Y_{ik} takes on 1 if the subject is at risk of event k and 0 otherwise (Therneau & Grambsch, 2000) and δ_{ik} represents the censoring variable for the k^{th} event and subject i (Amorim & Cai, 2015).

4.4 Shared frailty model

4.4.1. Model background

Although the AG model claims that survival times are independent, it does happen however, that subjects present survival times that are not independent. This typically occurs when the subjects in the study share some common risk. For instance, suppose the study is conducted in multiple clinics, and so subjects that attend the same clinic may have different survival times, compared to subjects that attend a different clinic. This may be due to different clinical practice standards that are followed by treating physicians per clinic (Collet, 2003).

Let's suppose a trial is conducted across 20 different clinics. To adjust for this clinic variable in the model, a design variable containing 19 factors is included in the model, clearly, analyzing this model might a bit cumbersome. Consequently, a random effect component that is realized by a probability distribution function that has a zero mean, and a variance equal to θ is included in a survival model (Ullah, et al., 2012). This then reduces the number of estimated coefficients related to the clinic variable from 19 to a single parameter θ. More generally, when there are numerous clinical sites that need to be considered in the survival analysis, including a random effect in a Cox PH model is a more efficient way of modeling timing of the event than considering modeling for each level of the clinical site in Cox PH model.

Under the current study of recurring events, some subjects present more than a single survival time in the study, and thus the survival times within a subject may not be independent. It follows then from the clinic analogy, that a subject-specific random effect can be included in the model, to explain the dependent survival times that arise in within the same subject. More generally, a random effect is included in the AG model – to reflect a suitable model for recurring events.

4.4.2. Description of the model

The random effect is a continuous variable that describes excess risk for distinct categories such as individuals, clinic sites or innate groups like families – the term 'excess risk' is known as frailty in the survival analysis context (Collet, 2003).

The interpretation of this model is that individuals have different frailties, and that those that are more frail experience the event earlier than others (Therneau & Grambsch, 2000). Accordingly, the frail term captures unobserved heterogeneity/covariates that is not captured by the observed covariates (Therneau & Grambsch, 2000). If individuals in a group share the same frailty the model is then called a shared frailty model as proposed by Yashin, et al.,

(1995). In the context of recurrent event models, given that an individual contributes more than a single observation in the study, a frailty term/random effect is then modeled for each individual to capture within subject heterogeneity.

4.4.3. Model assumptions

The shared frailty model assumes that all observations included in the model are independent, and the dependence structure is explained by the random effect. It also assumes a common baseline hazard function for all k events.

It is important to realize that the AG and shared frailty model share the same assumptions except that the dependence structure in the shared frailty model is captured by the random effect, this is in addition to using a robust variance estimator that accounts for possible correlation in the AG model. Therefore, the difference in the estimated coefficients between the AG and shared frailty model depend largely on how significant the random effect is in the model (Sagara, et al., 2014).

The random covariate can have a Gamma, a Gaussian or any other distribution. However, the Gamma distribution is widely preferred over other distributions because of its tractability (Sagara, et al., 2014).

4.4.4. Risk interval and risk set

The data structure used for the shared frailty model is the same as specified for the AG model in Table 4.2, since the only additional variable that will considered in the model is the participant ID and this is already included in Table 4.2.

4.4.5. Model form

The model notation employed here is as adopted by Therneau & Grambsch, (2000). Suppose that there are n total number of subject observations in the study and q subjects considered in the model, then the proportional hazard function for the subjects' i^{th} observation ($i = 1, \dots, n$) in the j^{th} subjects ($j = 1, \dots, q$) is written as:

$$\begin{aligned} \lambda_{ijk} &= Y_{ik}\lambda_0(t)U_j \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ik}) \\ &= Y_{ik}\lambda_0(t)\exp\,(\boldsymbol{\beta}'\boldsymbol{x}_{ik} + \boldsymbol{u}_j{}'\boldsymbol{z}_j). \end{aligned} \tag{4.5}$$

Here:

- λ_{ijk} represents the hazard of event k for the subjects' i^{th} observation from the j^{th} subject

- $\lambda_0(t)$ the baseline hazard function which is the same for all events in the study.
- $\boldsymbol{x}_{ik}$ represents the vector of explanatory variables for the subjects' i^{th} observation and event k
- $\boldsymbol{\beta}$ is a vector of regression coefficients**.**
- $\boldsymbol{u}_j$ is a vector of unknown random effects.
- $\boldsymbol{z}_j$ is the design vector that is 1 if the observation belongs to j^{th} subject, and 0 otherwise.
- Y_{ik} represents a predictable process taking value [0,1] indicating when a subject is under observation.

The density function for the random effect variable is written as:

$$f(u) = \frac{v^{1/\theta - 1}\exp(-\frac{u}{\theta})}{\theta^{1/\theta}\Gamma(1/\theta)},$$

where Γ is the Gamma function, and θ is the tuning parameter.

4.4.6. Likelihood function

The partial likelihood function that estimates the regression coefficient $\boldsymbol{\beta}$ is as follows

$$L(\beta) = \prod_{i=1}\prod_{k=1}\prod_{j=1}\left(\frac{U_j\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{\sum_{i=1}\sum_{k=1}Y_{ik}U_j\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}\right)^{\delta_{ik}}, \quad (4.6)$$

where Y_{ik} takes on 1 if the subject is at risk of event k and 0 otherwise (Therneau & Grambsch, 2000) and δ_{ik} represents the censoring variable for the k^{th} event and subject i.

4.5 Prentice, Williams and Peterson (PWP) model

4.5.1. Description of the model

When choosing a model for the timing to recurrent event, the analyst should consider the underlying dynamics of the recurring event within a subject. For instance, if the interest of the study is to model timing of recurrent infections, it is possible that after experiencing the first infection the risk of the next infection may increase per subject. This could happen if each infection permanently compromised the ability of the subjects' immune system to respond to subsequent infections (Therneau & Grambsch, 2000). Therefore, it of interest to evaluate whether the recurring events weakens the subjects' immune system, as result, the subject is likely to experience more events in the foreseeable future. Or rather the subject builds

immunity in response to the recurring events, as a result, will experience fewer events in the foreseeable future.

Prentice, et al., (1981) proposed a conditional model which is widely known as the Prentice, Williams and Peterson (PWP) model, this conditional model analyses ordered recurring events by a way of stratification, based on the prior number of events during the study follow-up (Sagara, et al., 2014). In general, a subject cannot be at risk for the k^{th} event, if they haven't terminated the $(k-1)^{th}$ event. This model analyses event specific effects and overall effects per covariate (Sagara, et al., 2014).

4.5.2. Model assumptions

This model preserves the sequence of event occurrence implying that the event dependence is accounted for in the model (Olawumi, et al., 2014). This model considerers the time scale of events in one of two ways:

- Total time (TT) – which is the time since study entry, also known as calendar time. The same time scale for all events is referenced to a fixed point in time, but does not allow an overlap of the risk periods for a given subject (Ullah, et al., 2012).
- Gap time (GT) – which is the time since previous event and is not relative to the actual time scale (Ullah, et al., 2012). When using gap or waiting time scale the time index is reset to zero after each recurrence of event with an assumption of a renewal process. Gaps between events are often useful for infrequent events this is when a renewal occurs after an event, or when the interest lies in the prediction of the next event (Amorim & Cai, 2014).

Hence, there are two types of PWP models based on the above-mentioned time scales: first of which is the PWP TT model, which evaluates the effect of the covariate for time since entry into the study. And secondly there is the PWP GT model, which evaluates the effect of covariate on the k^{th} event since time from the previous event (Amorim & Cai, 2014). The use of a time-dependent strata means that the underlying baseline hazard function varies from event-to-event, unlike the AG model that assumes that all events are identical (Therneau & Grambsch, 2000).

Moreover, unlike the AG model the effects of a covariate may vary from event to event in the stratified model. Therefore, the PWP model may be preferred over the AG model when the effects are different in subsequent events, which is likely to be the case for diseases such as viral infections because of immunity, or weakened immune system due to severe disease progression (Amorim & Cai, 2014).

4.5.3. Risk interval and risk set

The data structure of the PWP models is the same as that of the AG and shared frailty model Table 4.2. However, there are two added fields that show the event number – which will be used as the stratification variable in the models and a field called 'time' which is the value of the gap time between successive events (time suggests time is duration to event not gap time between events or start of the next interval), this field will be used in the PWP GT model only.

Table 4.3: PWP model data structure

Participant ID	Tstart	Tstop	Time	Event	Arm	Event number
121002	**01/Jan/2007**	**31/Mar/2007**	**3**	**1**	**Late Arm**	**1**
121002	10/Jun/2007	30/Sep/2007	4	1	Late Arm	2
121002	18/Nov/2007	01/Jan/2008	1	0	Late Arm	3
121139	**01/Jan/2007**	**30/Sep/2007**	**9**	**1**	**Late Arm**	**1**
121139	01/Nov/2007	01/Mar/2008	4	0	Late Arm	2
121170	**01/Jan/2007**	**31/Jan/2007**	**1**	**1**	**Sequential Arm**	**1**
121170	17/Jun/2007	01/Apr/2008	9	0	Sequential Arm	2
121178	**01/Jan/2007**	**31/Jan/2008**	**13**	**1**	**Late Arm**	**1**
121178	07/Jun/2008	30/Sep/2008	4	0	Late Arm	2
121179	**01/Jan/2007**	**01/Jan/2009**	**24**	**1**	**Early Arm**	**1**
121179	12/Feb/2009	01/Apr/2009	2	0	Early Arm	2

4.5.4. Model form

The PWP TT hazard function for i^{th} subject ($i = 1, \dots, n$) for event k ($k = 1, \dots, l_i$) is written as:

$$\lambda_{ik} = Y_{ik}\lambda_{0k}(t)\exp(\boldsymbol{\beta}_k'\boldsymbol{x}_{ik}). \tag{4.7}$$

The PWP GT hazard function for i^{th} subject ($i = 1, \dots, n$) for event k ($k = 1, \dots, l_i$) is written as:

$$\lambda_{ik} = Y_{ik}\lambda_{0k}(t_k - t_{k-1})\exp(\boldsymbol{\beta}_k'\boldsymbol{x}_{ik}). \tag{4.8}$$

Here:

- λ_{ik} represents the hazard of the k^{th} event for the i^{th} subject.
- $\lambda_{0k}(t)$ represents the event specific baseline hazard function of event k
- $\boldsymbol{x}_{ik}$ is the vector of explanatory variables for subject i.
- $\boldsymbol{\beta}_k$ represents a vector of regression coefficients specific for event k.

- $t_k - t_{k-1}$ represents survival times for the PWP GT model, this is the difference between the time when the k^{th} event occurred (t_k), and the time when the event before the k^{th} event occurred (t_{k-1}).
- Y_{ik} represents a predictable process taking value [0,1] indicating when a subject is under observation

4.5.5. Likelihood function

The regression coefficients are estimable with a partial likelihood function that is specified by (Labarga, et al., 2007). Let $t_{ki} < \cdots < t_{kd_k}$ denotes the ordered distinct survival times in stratum k, where d_k is the number of distinct events in stratum k. Suppose subject i experiences an event in stratum k at time t_{ki} and let $\boldsymbol{x}_{ki}(t_{ki})$ denote this subject's covariate vector at t_{ki}. Let $N(t) = \{n(u): u \leq t\}$, where $n(u)$, is the number of events of a subject in study prior to time u.

The partial likelihood to estimate $\boldsymbol{\beta}_k$ in the PWP TT is:

$$L(\beta) = \prod_{k\geq 1}\prod_{i=1}^{d_k} \frac{\exp(\beta_k' x_{ki}(t_{ki}))}{\sum_{m\in R(t_{ki},k)} \exp(\beta_k' x_m(t_{ki}))}. \quad (4.9)$$

Here $R(t,k)$ denotes the set of subjects in stratum k just prior to time t. The partial likelihood is valid for the PWP TT model provided that the stratification is restricted to be $k = n(t) + 1$ so that a subject can contribute at most a single survival time in a specific stratum (Labarga et al., 2007).

Furthermore, let $u_{ki} < \cdots < u_{kd_k}$, denote the distinct gap times from immediately preceding failure on the same subject, for the d_k events occurring in stratum k. Suppose subject i gives rise to the failure with gap time u_{ki} at time t_{ki}, and let $\boldsymbol{x}_{ki}(t_{ki})$ denote this subject's corresponding covariate value.

The partial likelihood to estimate $\boldsymbol{\beta}_k$ in the PWP GP is:

$$L(\beta) = \prod_{k\geq 1}\prod_{i=1}^{d_k} \frac{\exp(\boldsymbol{\beta}_k' \boldsymbol{x}_{ki}(t_{ki}))}{\sum_{m\in \tilde{R}(u_{ki},k)} \exp(\boldsymbol{\beta}_k' \boldsymbol{x}_m(\tau_m + t_{ki}))}. \quad 4.10$$

Here $\tilde{R}(u,k)$ denotes the set of subjects at risk in stratum k at gap time $u - 0, u \in [u_{k,i-1}, u_{ki})$ $(i = 1, \ldots, d_k + 1)$ where $u_{k0} = 0$ and $u_{k,d_k+1} = \infty$. τ_m is the last event time on subject m prior to entry into stratum k, $\tau_m = 0$ means that there are no further event expected for the subject.

4.6 Model building approach

Each of the models will potentially include many covariates and hence we need a strategy for variable selection. The below steps will be followed to obtain a final multivariable model:

- Univariate analysis - the first step in building the models is to perform a univariate analysis on each covariate considered per model. If a covariate is statistically significant at 10% level of significance under the univariate analysis then this covariate will be fitted in the preliminary multivariable model (note that treatment arm will not be included included at this stage).
- Preliminary multivariable model - all covariates that are significant at 10% level of significance at this stage are noted and included in all the models that have been discussed in this section. Suppose for instance that covariate X is a significant predictor for preliminary multivariable Model A and covariate X is not significant in preliminary multivariable Model B stage, covariate X will be included in the final multivariable analysis for Model A and Model B, regardless of lack significance that covariate X achieved under Model B. This is to make comparison of the models easier. Results for the preliminary multivariable models will not be presented, as this is an interim step, which is not of interest, with regards to interpretting the models.
- Final multivariable model – All models fitted will contain the same covariates that proved to be statistically significant in any of the preliminary multivariable models that were fitted in the previous step, and the treatment arm variable, and other clinical important variables will be added into the final model regardless of statistical significance results obtained in the univariate and preliminary multivariable analysis.

4.7 Model adequacy assessment

The main goal here is to determine how accurately each of the models predict the survival probability. This is done by plotting estimated survival probability from each model with the plot of Kaplan-Meier estimates (estimate for observed survival function) estimating probability of surviving from recurring events, much of the SAS coding here is from Reyes & Thomas (2014).

4.8 Application

4.8.1. Background

The Cox regression model has been fitted in Chapter 3, and the final model that was obtained included treatment arm, age, sex, alcohol consumed, total bilirubin, Hepatitis B Surface

Antigen and Baseline LFT results. The best Cox regression model obtained assumed that there is a time-varying coefficient that explains the association of treatment arm and timing of hepatotoxicity. This information is carried through in building models in this section.

4.8.2. Univariate analysis

The potential covariates to be included in the models and that therefore will be assessed in the univariate analysis, are the treatment arm, age, sex, body mass index (BMI) strata, history of TB, extra pulmonary TB, World Health Organization (WHO) stage, alcohol consumed, CD4 count category, CD8 count, baseline: log viral load, alkaline phosphatase (ALK), total bilirubin (BILI), lactate dehydrogenase (LDH), hepatitis B surface antigen (HBsAg) and Baseline LFT.

Results from the univariate analysis are presented in Figure 4.2 to Figure 4.4 for each model, the hazard ratio and the 95% CI are depicted in a graph, where the dots represent the hazard ratio, color coded red if the hazard ratio is significant at 5% level of significance and color is yellow if the hazard ratio is significant at 10% level of significance in the model, otherwise the color is black. The lines extending from the dot indicate the 95% confidence interval.

Figure 4.2: Univariate Cox model analysis

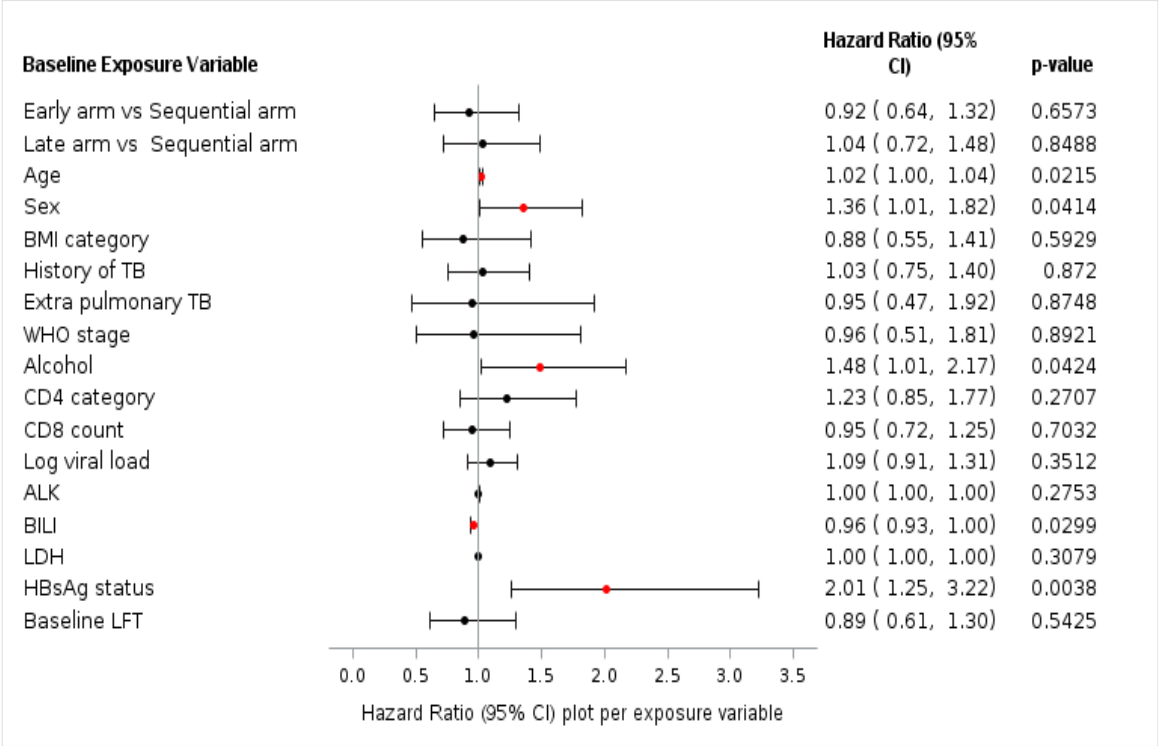


Figure 4.3: Univariate Islam's resolution model

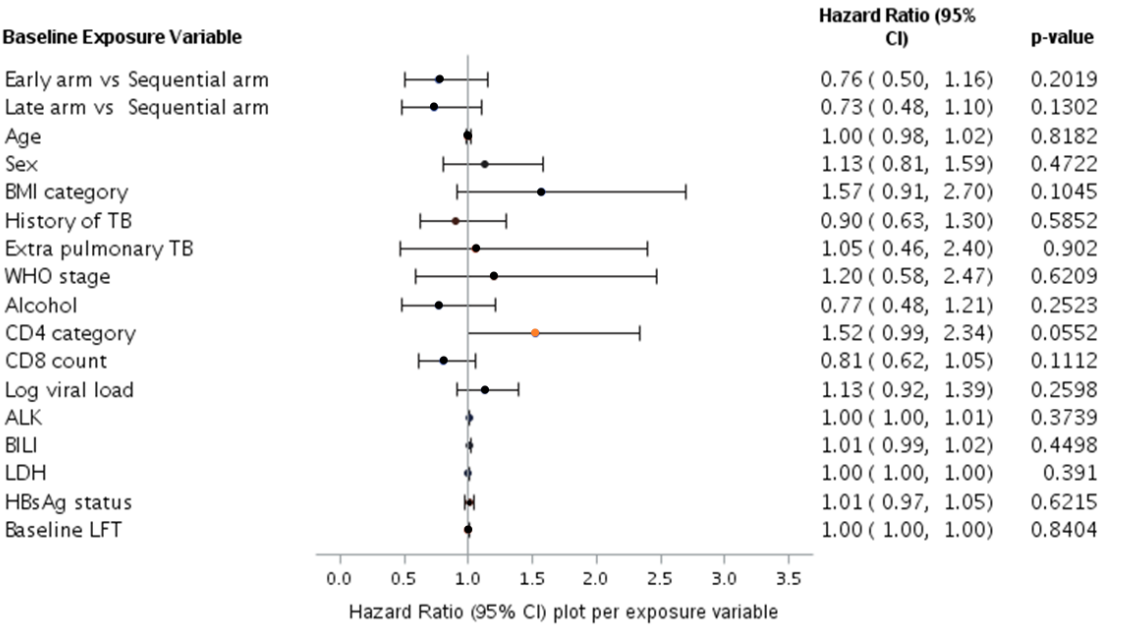


Figure 4.4: Univariate Islam's recurrent model

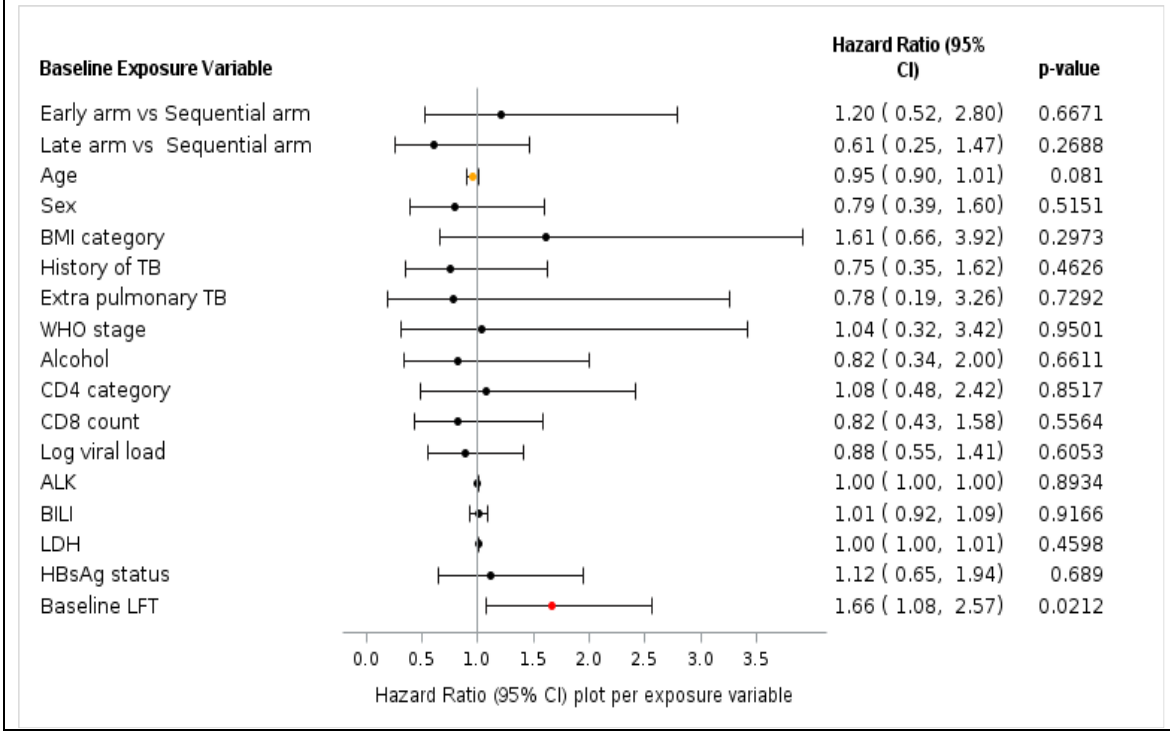


Figure 4.5: Univariate AG model analysis

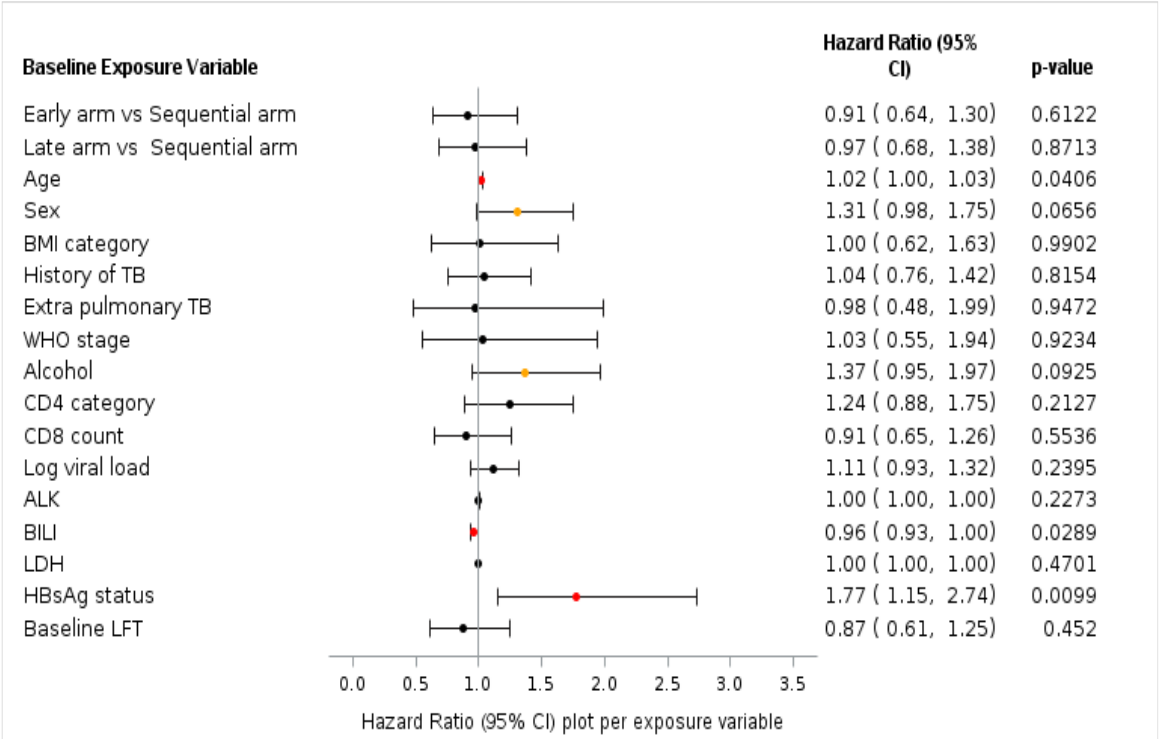


Figure 4.6: Univariate Shared Frailty model

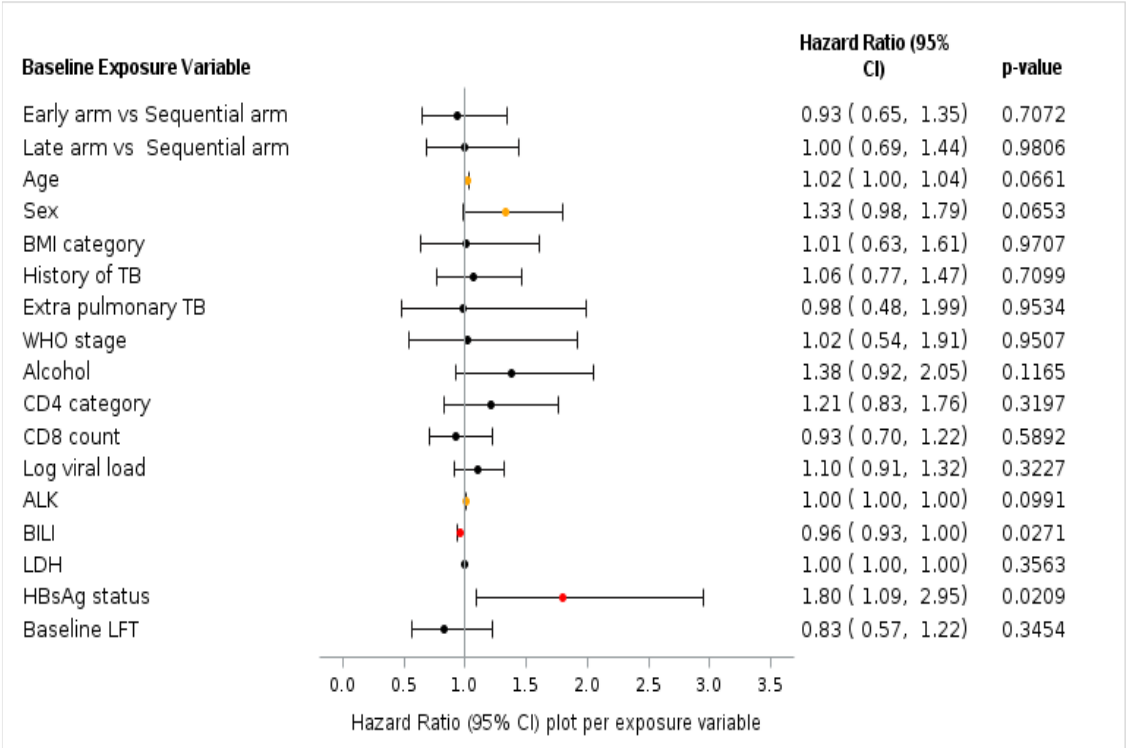


Figure 4.7: Univariate PWP TT model

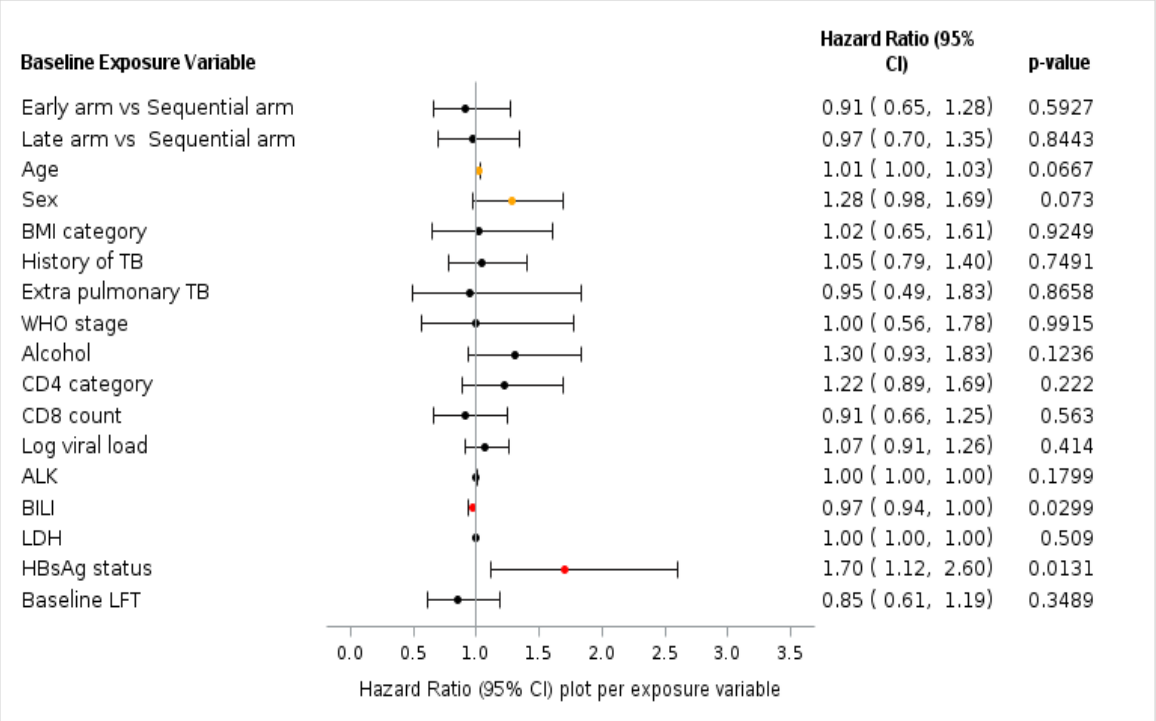
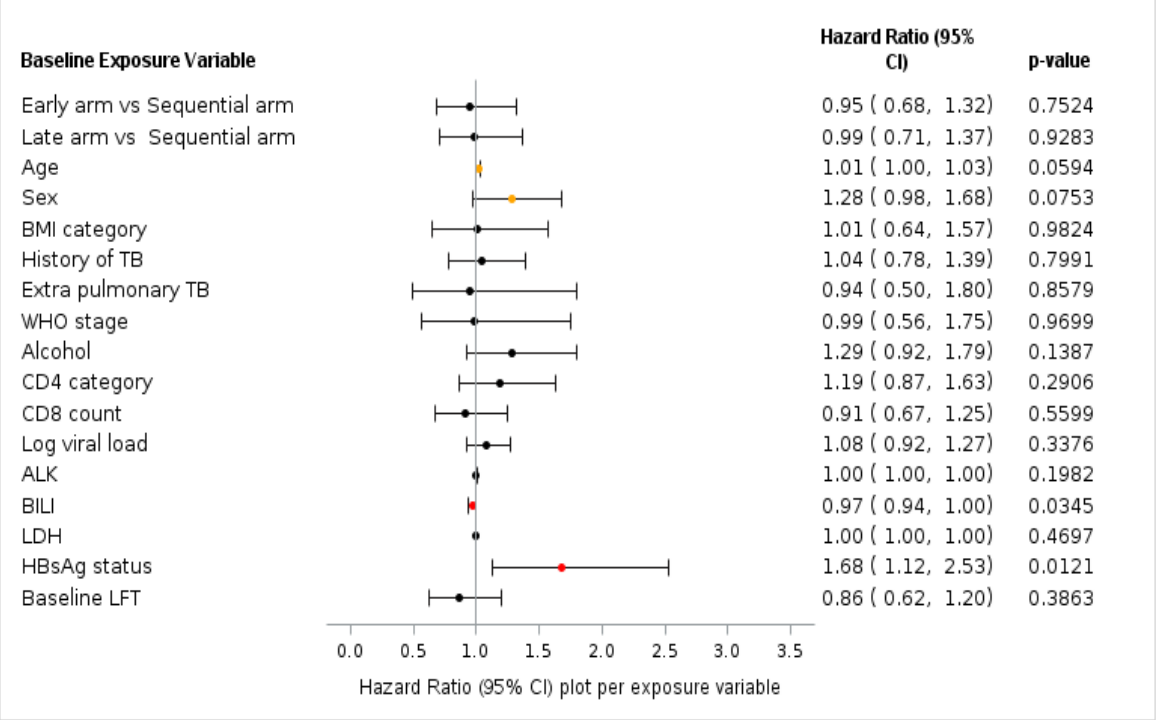


Figure 4.8: Univariate PWP GT model



A summary of variables that are statistically significant on the univariate analysis level (shown in Figure 4.2 and Figure 4.8) is in Table 4.4. Age, sex, alcohol consumed, CD4 category, ALK,

BILI, HBsAg status and baseline LFT were significant enough to be considered in the preliminary multivariable models. However, the level of significance for each variable per model output varied, for instance age is statistically significance at 5% level of significance as a predictor of occurrence of hepatotoxicity according to the Cox and AG model, whereas, it is only statistically significant at 10% level of the significance as a predictor of occurrence of hepatotoxicity for shared frailty model, PWP TT, PWP GT and the recurrent model.

The male sex has an increased risk of developing hepatotoxicity compared to females, the sex variable is statistically significant at 5% level for significance as a predictor of hepatotoxicity occurrence based on the Cox model, whereas, it is only statistically significant at 10% level of significance as a predictor of occurrence of hepatotoxicity for the AG, Shared frailty, PWP TT, and PWP GT model. Alcohol is statistically significant at 5% level for significance as a predictor of occurrence of hepatotoxicity based on the Cox model, whereas, it is only statistically significant at 10% level of the significance as a predictor of occurrence of hepatotoxicity for the AG model.

CD4 category and ALK are statistically significant at 10% level of the significance as a predictor of resolution and occurrence of hepatotoxicity respectively based for the Resolution and Shared frailty univariable model assessment respectively. BILI and HBsAg are statistically significance at 5% level for significance based the Cox, AG, Shared frailty, PWP TT and PWP GT. And lastly, baseline LFT variable is statistically significant at 10% level of the significance based on the recurrent univariate assessment.

Therefore, for each model considered, treatment arm will be included since this is a factor of interest, and it will be included with an assumption that it has time-varying coefficient, since this has been proved in the previous section, where treatment arm fits the model better if it the effect is assumed to be time-varying. The proportional hazards assumption test will be not re-performed for each of the recurrent event models, the results from the proportional hazards assumption from the Cox Model will be assumed for the recurrent event models, this is to faciliatate the comparability of the models.

Secondly, variable: age, sex, alcohol consumed, CD4 category, BILI, HBsAg status and Baseline LFT will be included in the final multivariable models since these variables were significant for at least one of the preliminary multivariable models that were fitted after the univariate analysis. It is important to note that ALK was not significant in the shared frailty preliminary multivariable model, and therefore excluded in the final multivariable model.

Table 4.4: Summary of statistically significant variables from the univariable analysis by significance levels and respective model

Baseline exposure variable	Statistically significant at:	
	5%	10%
Age	• Cox • AG	• Shared frailty • PWP TT • PWP GT • Recurrent
Sex	• Cox	• AG • Shared frailty • PWP TT • PWP GT
Alcohol	• Cox	• AG
CD4 category		• Resolution
ALK		• Shared frailty
BILI	• Cox • AG • Shared frailty • PWP TT • PWP GT	
HBsAg status	• Cox • AG • Shared frailty • PWP TT • PWP GT	
Baseline LFT		• Recurrent

4.8.3. Multivariable models

The multivariable models were considered and compared based on two views, the first view is in Figure 4.9 considers the multistate model as depicted in Figure 4.1, i.e. the Cox, resolution and recurrent model is considered and examined, note that these are the models presented in Islam's recurrent model framework (Islam, 1994). The second part presented in Figure 4.10 is the multivariable analysis that considers the Cox model and compares it to the AG, shared frailty, PWP TT and the PWP GT model. Note that in each figure the hazard ratios are depicted by a dot that is color code red if the variable is statistically significant at 5% level of significance, yellow if it is statistically significant at 10% level of significance and black otherwise, the 95% CI is indicated by lines extending from the dot. And on the right of each figure the covariate that has been assessed is shown, as well as the value of hazard ratio, 95% CI, p-value and the standard error associated with the covariate.

4.8.4. Multivariable models output for Islam's multistate models

Treatment arm in the first 5 months

Cox model

- Patients in the early arm had a 3% decreased relative hazard of developing first hepatotoxicity compared to patients in the sequential arm (HR=0.97, 95% CI (0.52, 1.84)).
- In the late arm patients were 1.65 times more likely to develop the first episode of hepatotoxicity compared to patients in the sequential arm (HR=1.65, 95% CI (0.92, 2.94)) with p-value = 0.0912.

Resolution model

- In the early arm patients had a 22% decreased chance of resolving hepatotoxicity compared to patients in the sequential arm (HR=0.78, 95% CI (0.46, 1.32)).
- Patients in late arm had a 32% decreased chance of resolving hepatotoxicity compared to patients in the sequential arm (HR=0.68, 95% CI (0.41, 1.14)).

Recurrent model

- In the early arm patients were 3.49 times more likely to develop the second episode of hepatotoxicity compared to patients in the sequential arm (HR=3.49, 95% CI (0.33, 36.70)) and the standard error for this estimate is 1.2013.
- Patients in the late arm were 2.66 times more likely to develop the second episode of hepatotoxicity compared to patients in the sequential arm (HR=2.66, 95% CI (0.27, 26.12)) and the standard error for this estimate is 1.1663.

Treatment arm post 5 months

Cox model

- Patients in the early arm had a 7% decreased hazard of developing first hepatotoxicity compared to patients in the sequential arm (HR=0.93, 95% CI (0.53, 1.63)).
- In the late arm patients has 14% decreased hazard of developing first hepaptotoxicity compared to patients in the sequential arm (HR=0.86, 95% CI (0.49, 1.50)).

Resolution model

- Patients in the early arm had a 18% decreased chance of resolving hepatotoxicity compared to patients in the sequential arm (HR=0.82, 95% CI (0.33, 2.05)).

- Patients in the late arm were 1.79 times more likely to resolve hepatotoxicity compared to patients in the sequential arm (HR=1.79, 95% CI (0.65, 4.95)).

Recurrent model

- In the early arm patients had a 16% reduced hazard of developing the second hepatotoxicity episode compared to patients in the sequential arm (HR=0.84, 95% CI (0.24, 2.88)).
- In the late arm patients had a 70% reduced chance of developing the second episode of hepatotoxicity compared to patients in the sequential arm (HR=0.3, 95% CI (0.10, 0.90)) with p-value = 0.0322.

Age is significant predictor for two of the multistate models, for every year increase in the baseline age, the hazard for developing the first hepatotoxicity increased by 3% according to the Cox model (HR=1.03, 95% CI (1.01, 1.05)) with a small p-value = 0.0139. In contrast, for every year increase in the baseline age the hazard for developing recurring hepatotoxicity decreases by 8% according to the recurrent model (HR=0.92, 95% CI (0.86, 1.00)) with a p-value = 0.0422. Age is not predictive in the resolution model.

Alcohol is a significant factor in the Cox model, where subjects who consume alcohol are 1.7 times more likely to develop hepatotoxicity compared to patients who never consume alcohol (HR=1.7, 95% CI (1.04, 2.79)) with a small p-value = 0.0349. According to the resolution model, patients who consume alcohol have a 44% reduced chance of resolving hepatotoxicity, compared to patients who never consume alcohol (HR=0.56, 95% CI (0.33, 0.97)) with a small p-value = 0.0399. Alcohol was not a predictive factor for the recurrent model.

Total bilirubin and HBsAg status are predictive factors for the Cox model only. An unit increase of the total bilirubin reduces the hazard of developing hepatotoxicity by 6% (HR=0.94, 95% CI (0.91, 0.98)) with p-value = 0.0399, and patients with a positive HBsAg status are 2.1 time more likely to develop first hepatotoxicity compared to patients with a positive HBsAg status at baseline (HR=2.1, 95% CI (1.26, 3.48)) with p-value = 0.0042.

Baseline LFT is a signicant factor for the resolution model, patients with abnormal baseline LFTs are 1.57 times more likely to resolve their first hepatotoxicity (HR=1.57, 95% CI (1.02, 2.43)) with p-value = 0.0042.

Sex and CD4 count were not predictive for any of the multi state multivariable models.

4.8.5. Multivariable models for Cox's generalized recurrent models - AG, Shared fraily, PWP TT and PWP GT

Treatment arm in the first 5 months

Cox model

- Patients in the early arm had a 3% decreased hazard of developing first hepatotoxicity compared to patients in the sequential arm (HR=0.97, 95% CI (0.52, 1.84)).
- Patients in the late arm are 1.65 times more likely to develop first hepatotoxicity compared to patients in the sequential arm (HR=1.65, 95% CI (0.92, 2.94)) with p-value = 0.0912.

Recurrent models

- In contrast to the Cox model output, all recurrent models indicated subjects in the early arm have increased hazard of developing hepatotoxicity compared to patients in the sequential arm. AG model - (HR=2.16, 95% CI (0.74, 6.27)), Shared frailty model - (HR=2.21, 95% CI (0.74, 6.56)), PWP TT model - (HR=2.17, 95% CI (0.74, 6.34)), and PWP GT model - (HR=1.69, 95% CI (0.68, 4.21)), none of these estimates were statistically significant. The lowest to highest standard error observed per model is in this order, Cox model (SE= 0.3245), PWP GT model (SE= 0.4661), PWP TT model (SE= 0.5475), AG model (SE= 0.5450) and the shared frailty model (SE= 0.5555).
- Concurrent to the Cox model, all recurrent models indicated patients in the late arm have an increased hazard of developing hepatotoxicity compared to patients in the sequential arm. AG model - (HR=5.70, 95% CI (2.19, 14.83)), Shared frailty model - (HR=6.06, 95% CI (2.28, 16.10)), PWP TT model - (HR=5.70, 95% CI (2.19, 14.83)), and PWP GT model - (HR=3.17, 95% CI (1.44, 6.99)), all estimates were statistically significant across the models. The lowest to highest standard error observed per model is ranked as, AG model (SE= 0.2394), Cox model (SE= 0.2956), PWP GT model (SE= 0.4035), PWP TT model (SE= 0.4874), and the shared frailty model (SE= 0.4984).

Treatment arm post 5 months

Cox model

- In the early arm patients had a 7% decreased hazard of developing first hepatotoxicity compared to patients in the sequential arm (HR=0.93, 95% CI (0.53, 1.63)).
- Patients in the late arm had a 14% decreased hazard of developing first hepatotoxicity compared to patients in the sequential arm (HR=0.86, 95% CI (0.49, 1.50)).

Recurrent models

- Concurrent to the Cox model output, all recurrent models indicated subjects in the early arm have decreased hazard of developing hepatotoxicity compared to patients in the sequential arm. AG model - (HR=0.78, 95% CI (0.49, 1.25)), Shared frailty model - (HR=0.79, 95% CI (0.48, 1.29)), PWP TT model - (HR=0.80, 95% CI (0.50, 1.28)), and PWP GT model - (HR=0.84, 95% CI (0.53, 1.33)), none of these estimates were statistically significant. The lowest to highest standard error observed per model is in this order, PWP GT model (SE= 0.2328), PWP TT model (SE= 0.2398), AG model (SE= 0.2394), shared frailty model (SE= 0.2510) and Cox model (SE= 0.2845).
- Concurrent to the Cox model, all recurrent models indicated subjects in the late arm have decreased hazard of developing hepatotoxicity compared to patients in the sequential arm. AG model - (HR=0.65, 95% CI (0.40, 1.04)), Shared frailty model - (HR=0.64, 95% CI (0.39, 1.07)), PWP TT model - (HR=0.68, 95% CI (0.42, 1.08)), and PWP GT model - (HR=0.67, 95% CI (0.43, 1.05)), the AG, shared frailty and PWP GT model had estimates that were marginally significant. The lowest to highest standard error observed per model is in this order, PWP GT model (SE= 0.2317), PWP TT model (SE= 0.2408), AG model (SE= 0.2420), shared frailty model (SE= 0.2600) and Cox model (SE= 0.2838).

Age and alcohol are significant predictors of hepatotoxicity for the Cox model and all the recurrent models except the PWP GT model. The hazard ratio, the 95% confidence interval and the SE is similar for the Cox model and all the recurrent models, as well as the PWP GT model, despite the lack of significance for this estimate under this model.

CD4 count is not a significant predictor of hepatotoxicity for the Cox model and the PWP GT model and it is a marginal significant predictor for AG, shared frailty and the PWP TT model. The hazard ratio, the 95% confidence interval and the SE are similar for the Cox model and all the recurrent models. The PWP GT had the least SE and the Cox model the highest SE.

Total bilirubin and HBsAG status are significant predictors of hepatotoxicity for the Cox model and all the recurrent models. The hazard ratio, the 95% confidence interval and the SE are for the Cox model and all the recurrent models.

Neither sex nor baseline LFT were predictive in neither the Cox model nor the recurrent models.

Figure 4.9: Multivariable model output for Islam's multistate models

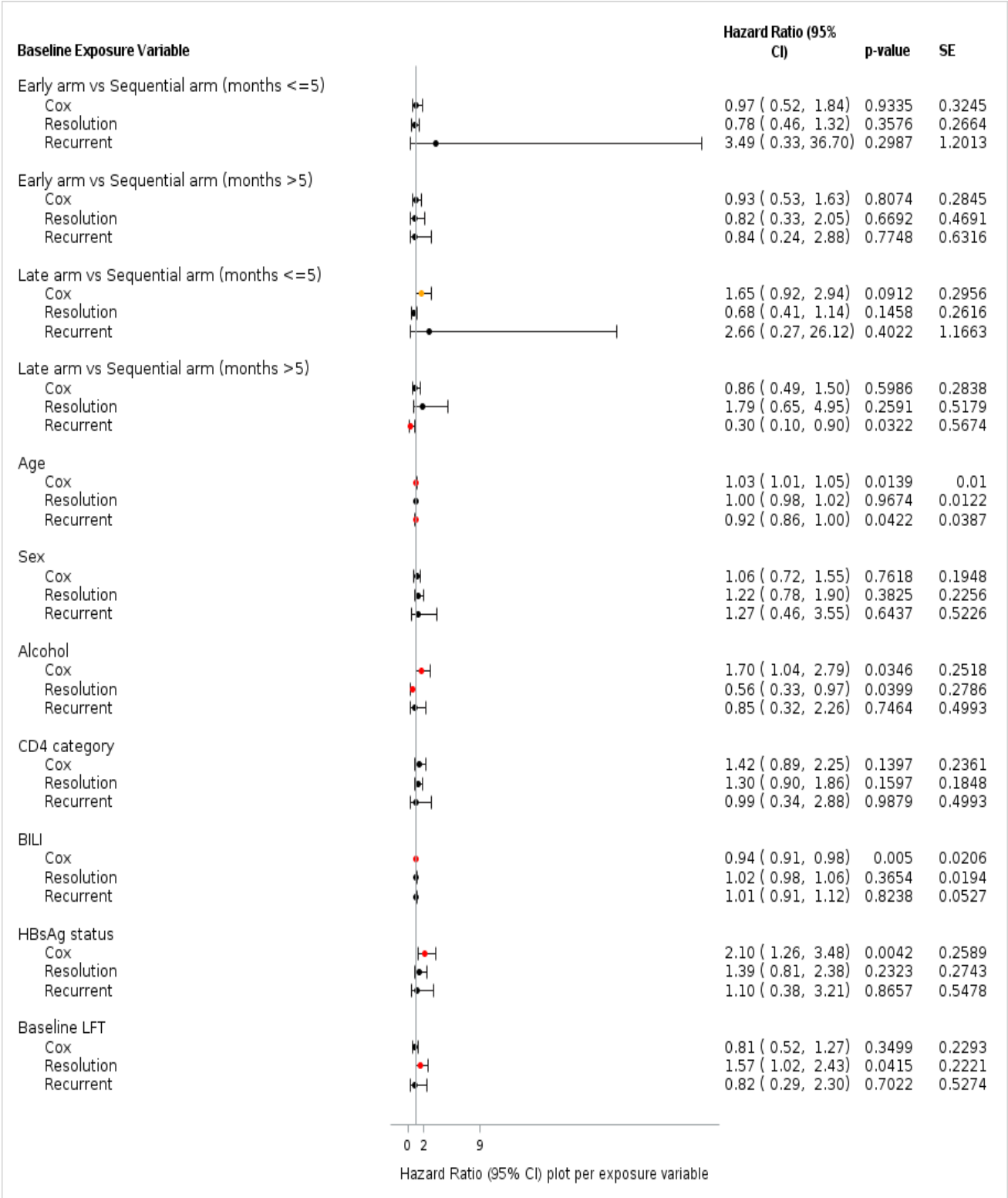
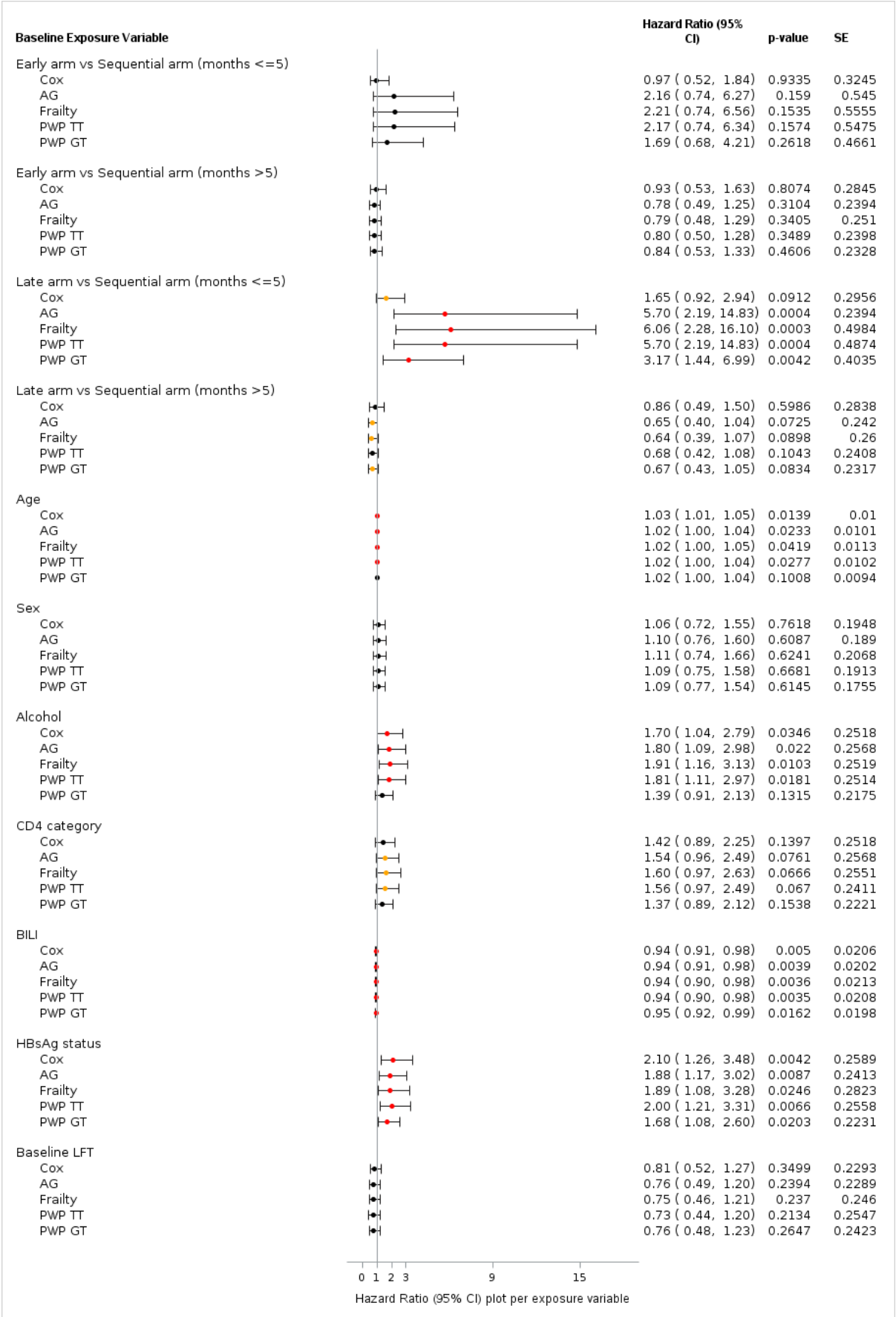


Figure 4.10: Multivariable models for the Cox and Recurrent models



4.8.6. Model validation

In Figure 4.11 to Figure 4.18 the Kaplan- Meier curves are plotted with the estimated survival functions based on covariates that are contained in the final models. The estimated survival function depicts the survival distribution for a patient that is in the sequential arm, with an average age, who is female that never consumes alcohol, had a baseline CD4 + count < 50 cells/mm3, with an average BILI, negative HBsAg status and had normal LFTs at baseline . In Figure 4.11 and Figure 4.12, an estimated survival function derived from the Cox regression model is shown, and it is compared to the observed Kaplan-Meier curve based on the realization of the first hepatotoxicity in - Figure 4.11, and it is compared to the observed Kaplan-Meier curve that is based on the realization of recurring hepatotoxicity in Figure 4.12. In Figure 4.11 the Cox regression model is not convincing that it fits the data well. Coincidentally, in Figure 4.12, the Cox regression model fits the data fairly well.

In Figure 4.13 and Figure 4.14, both the resolution model and the recurrent model fit the data well. The figure in Figure 4.14 displays some volatility, when comparing the KM curve and survival function for recurrent model, this is due to the thinning out of data for this model.

In Figure 4.15 and Figure 4.16, the AG model and the shared frailty model fit is assessed, a good fit of both these models is questionable. In Figure 4.17 and Figure 4.18, observed Kaplan-Meier curve is stratified by the event number, and compared with estimated survival curves produced by the PWP TT and the PWP GT model respectively. The PWP TT models seems to have a better fit for the stratified observed survival time. The fit of the PWP GT model is questionable. It important note that the time function in each Kaplan-Meier plotted in Figure 4.18 is based on a Gap time for the PWP GT model and based on calendar time for the PWP TT model, hence the positioning of the Kaplan-Meier curve is different per figure.

Figure 4.11: Observed Kaplan-Meier based plot on first hepatotoxicity event versus Cox regression output of estimated survival function

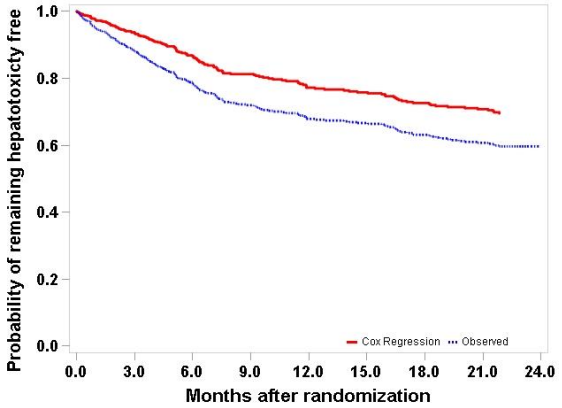


Figure 4.12: Observed Kaplan-Meier plot based on recurring hepatotoxicity event versus Cox regression model's estimated survival function

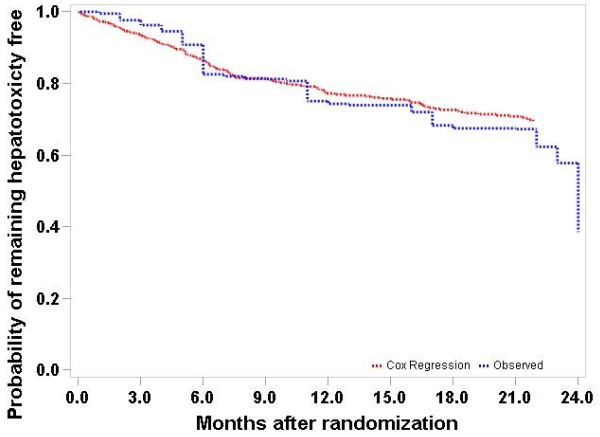


Figure 4.13: Observed Kaplan-Meier plot based on the first resolution of hepatotoxicity versus the resolution model's estimated survival function

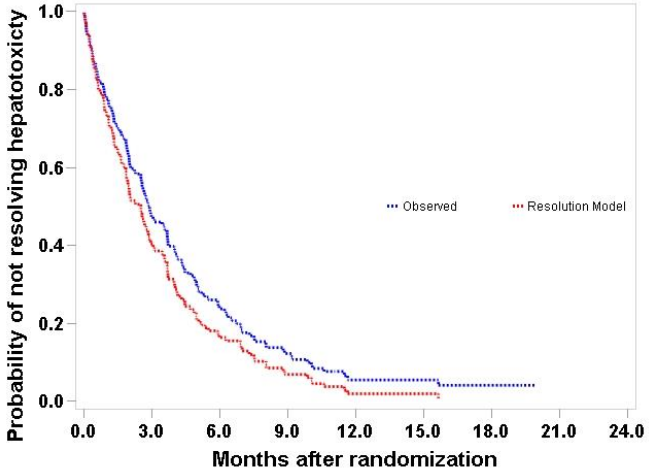


Figure 4.14: Observed Kaplan-Meier plot based on the occurrence of the second hepatotoxicity event versus the repeated episode model's estimated survival function

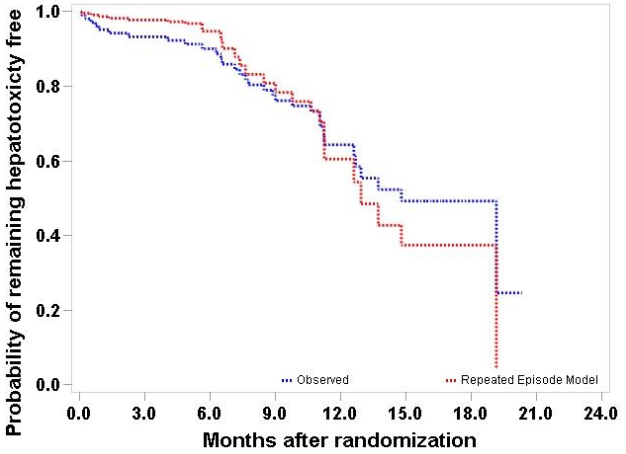


Figure 4.15: Observed Kaplan-Meier plot based on recurring hepatotoxicity event versus AG model's estimated survival function

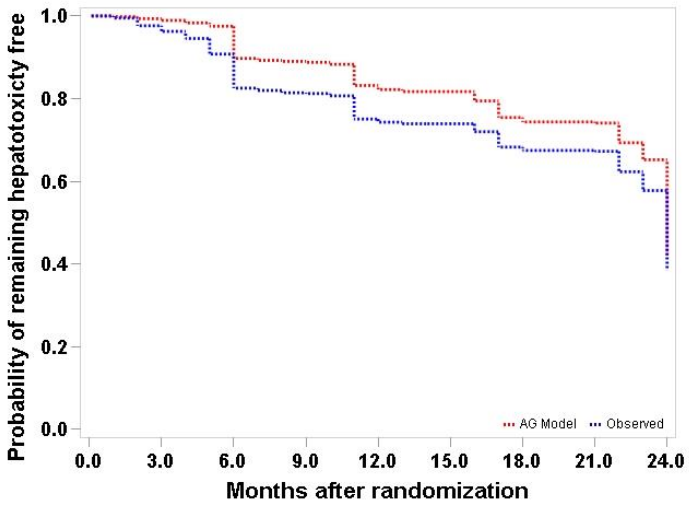


Figure 4.16: Observed Kaplan-Meier plot based on recurring hepatotoxicity event versus shared frailty model's estimated survival function

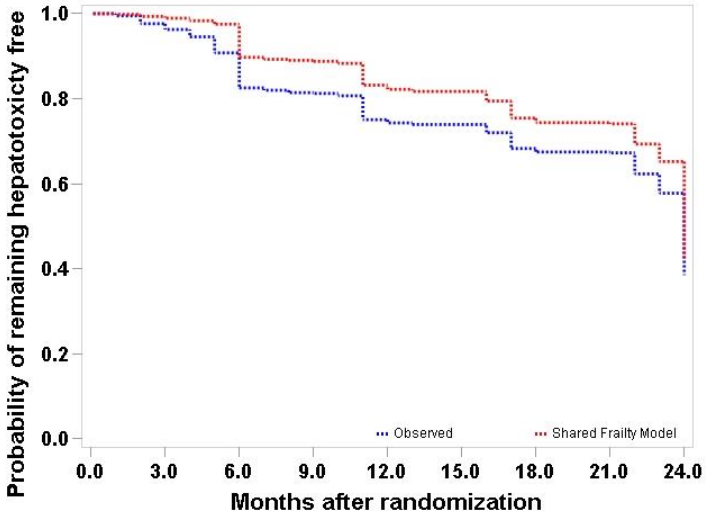


Figure 4.17: Observed Kaplan-Meier plot based on recurring hepatotoxicity event versus PWP TT model's estimated survival function

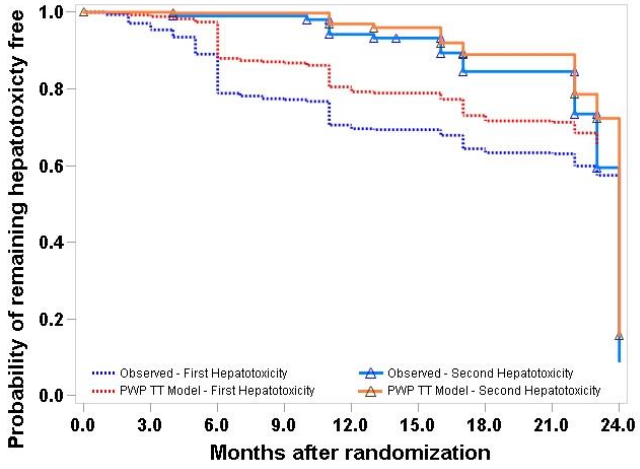
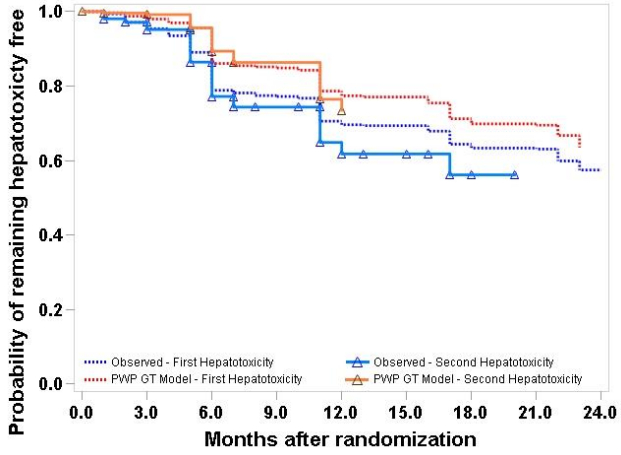


Figure 4.18: Observed Kaplan-Meier plot based on recurring hepatotoxicity event versus PWP GT model's estimated survival function



4.8.7. Model fit statisitcs

The PWP TT model has the lowest AIC statistic amongst Cox's generalized recurrent model. This implies that the PWP TT model fit the recurrent data better, than the other models.

Table 4.5: Akaike information criterion

Cox generalised recurrent models	**AIC**
AG model	1581.644
Shared Frailty model	1580.603
PWP TT	1557.512
PWP GT	1758.841

4.9 Conclusion

If Islam's multistate models are considered in combination (Cox, resolution and recurrent model) a chronological view of the evolution of hepatotoxicity is examined throughout the study duration. Through these multistate models, it can be assessed, how patients in the SAPiT trial transition from study randomization to the first hepatotoxicity (Cox model), and then transition from the first hepatotoxicity to resolving hepatotoxicity (resolution model) and evaluate how patients transition from the first resolution of hepatotoxicity to the second development of hepatotoxicity (recurrent model).

The Cox model and the recurrent model (second hepatotoxicity) are attempting to model the same risk, this is the risk of developing hepatotoxicity in the study. The difference between these two models is the time when this risk is estimated and the risk set used in the model. It is important to note that the risk set considered in the recurrent model are patients who have developed hepatotoxicity at least once in their lifetime, therefore patients in this model may be vulnerable or immune to hepatotoxicity depending on the biological impact hepatotoxicity has on the body of the patient. This suggests that results from this model cannot be generalized to patients who do not have a history of hepatotoxicity, as results from the recurrent model are biased by the history that the patient has with hepatotoxicity.

The estimated hazard ratio per covariate when comparing the Cox model and the recurrent model differs in magnitude for some covariates in the models, in a sense that the hazard ratio output can be below 1 for one model and above 1 for another model. This fundamentaly changes how the association of that factor with the development of hepatotoxicity is interpreted. In general if the

hazard ratio is below 1 for a continuous covariate, for every unit increase in that covariate decreases the hazard of hepatotoxicity, and if the hazard ratio is above 1, then for every unit increase of the continuous covariate increases the hazard of development of hepatotoxicity. For a categorical variable if a hazard ratio is below 1 then the reference category of that covariate has an increased hazard of developing hepatotoixicty compared to the category at hand, if the hazard ratio is above 1 for a categorical variable then the reference category has an decreased hazard of developing hepatotoxicity then the category in question.

For instance the Cox model, indicated that patients in sequential arm (reference category) are more likely to develop first hepatotoxicity compared to patients in the early arm (category in question) in the first 5 months. In contrast, within the same period the recurrent model suggested that patients in the sequential arm have a much lower chance of developing second hepatotoxicity compared to patients in the early arm. This suggests that the risk profile of a patient is not the same when modelling first hepatotoxicity compared to when modelling second hepatotoxicity.

Likewise for variables: age, alcohol consumed and total bilirubin indicate that the risk profile for a patient who develops first hepatotoxicity is different to risk profile of patient who develops second hepatotoxicity. Whereas sex, CD4 count, HBsAg status and the baseline LFT are inconclusive of how the risk changes from the first hepatotoxicity to the second hepatotoxicity.

Considering the resolution model in comparison to the Cox model and recurrent model is akin comparing apples with oranges, since the resolution model assesses timing of resolving hepatotoxicity compared to timing of development of hepatotoxicity (either first or second) hepatotoxicity. Information that is gathered from analyzing the Cox model, the recurrent model alongside a resolution model indicate that, for instance, although increase of a continuous variable X increases the hazard of developing hepatotoxicity it does not automatically mean that the decrease of variable X increases the chance of resolving hepatotoxicity. The same applies if variable X is categorical variable, if category A increases the hazard of developing hepatotoxicity compared to category B of X, that does not always mean that category A decreases the chance of resolving hepatotoxicity compared to category B of X.

However, it is worthwhile to investigate protective factors of resolving hepatotoxicity, since at times it is difficult to prevent hepatotoxicity from occurring, therefore clinicians can refer to the protective factors from the resolution model to determine if the patient has high chance of resolving hepatotoxicity. Naidoo, et al., (2014) details the ‘Changes to antiretroviral drug regimens

during integrated TB-HIV treatment: results of the SAPiT trial', and the outcome from this study was that only 3 patients changed their treatment regimen because of elevated transaminases (liver function test) and hyperlactatemia. Consequentially, it could be recommended that clinicians refrain from interrupting optimal TB treatment or ART, and rather allow time to pass and maybe hepatotoxicity will resolve by itself without treatment interuption, like the majority of hepatotoxicity cases that occurred during the study period.

All of Islam's multistate model can be designated as a piecewise Cox models, since the model form for all three models is the same as that of Cox model, the data is considered 'piece by piece' depending on whether a participant is eligible to make the risk set or not. The Cox model has the largest risk set, the resolution model has the next largest risk set, and the recurrent model has the smallest risk set of the three models. The impact of size of each risk set is apparent when comparing the standard errors. The recurrent model has the highest standard error compared to the Cox model and the resolution model. This translates to a wider confidence interval output from the recurrent model compared to the Cox and the resolution model. Suggesting that estimates from the recurrent model are not very precise. And the results can not really be generalized to another population.

However, modeling the recurrent model alongside the Cox model gives an indication of whether there is event dependence when considering recurring hepatotoxicity or not. As mentioned above assessment of the Cox model and the recurrent model as Islam suggested indicated that risk profile of the patient changes from the first hepatotoxicity to the second hepatotoxicity, this implied that there is some level event dependence.

The AG model and shared frailty models intended to model the overall effect of the risk factors on the occurrence of an event, and this is accomplished through an application of a common baseline hazard, rather than an event specific baseline hazard. And a single coefficient is estimated per risk factor included in the model, rather than estimating event specific coefficient for each risk factor in the model. The difference between the two models is that a random effect is modelled in the shared frailty model to account for heterogeneity (or participant specific excess risk) that comes with each subject in the data set.

On the other hand the PWP TT and PWP GT assumes an event specific baseline hazard and models an event specific coefficient for each covariate in the model, this is, the parameter esimates are allowed to vary from event to event. Based on the description of the models in combination with concluding remarks from Islam's model analysis which suggested event

dependence is evident when modelling the occurrence of second hepatotoxicity, suggesting that the PWP TT and PWP GT models are the more appropriate to models for recurring hepatotoxicity in the SAPiT data compared to the AG and the shared frailty model, since event dependence is accounted for explicitly in the former models.

The model estimates are often closely comparable for the AG, shared fraily and the PWP TT, while parameter estimates for the PWP GT is slightly different to the rest, and the standard error for the PWP GT is lower than other models. It can be deduced that the estimates are similar for for the three models AG, shared frailty and PWP TT because risk interval for each of the models is the same, it is based on calendar time, this is time since study entry, whereas the risk interval for the PWP GT model is based on time since last event.

The parameter estimates that quantifies the hazard ratio for the treatment arm variable are quite wide for the AG, shared frailty, PWP TT, PWP GT as well as the Islam's recurrent model specifically in the first 5 months of study randomization. This may be because, although they may be reasonable enough first hepatotoxicity events occurring with the first 5 months, there are not many recurring events that happen in that period, the sparse occurrence of the second, third and fourth events causes the standard errors to be high, as a result the confidence intervals are wide, suggesting that the resulting parameter estimates are not precise.

As mentioned above, the two models contending for the best model for recurring hepatotoxicity is between PWP TT and PWP GT, since these models address event dependence directly in its formulae. However, PWP TT has the lowest AIC compared to the PWP GT model, suggesting that the PWP TT is the most suitable model for SAPiT data for recurring hepatotoxicity.

Chapter 5

Discussion

This project used data from the 'Starting Antiretroviral therapy at three Points in Tuberculosis' (SAPiT) trial that was conducted in sub-Saharan Africa. The SAPiT trial was intended to identify an optimal timing of ART initiation in patients who are on TB-treatment. Patients were randomized into three treatment arms, the first arm initiated ART during the intensive phase of TB treatment (early arm), the second arm initiated ART during the continuation phase of TB treatment (late arm) and the third arm initiated ART after the completion of the continuation phase of TB treatment (sequential arm). Reduced mortality rates were observed in the early and late arm compared patients in sequential arm were noted (Abdool Karim, et al., 2010). Modelling hepatotoxicity in this trial is of secondary importance, despite this, a thorough assessment of hepatotoxicity is important, since occurrence of hepatotoxicity may lead to treatment cessation, or treatment regime interruption which may result in a sub-optimial treatment outcome.

In this study, 30% (179/593) of the patients in the trial developed the first occurrence of hepatotoxicity (Grade >= 1). Severe hepatotoxicity was not common, with only just 2% (13/593) occurring in the study. Of the patients who developed the first occurrence of hepatotoxicity 76% (136/179) resolved their first hepatotoxicity. Thereafter, of the patients who resolved hepatotoxicity, 24% (32/136) of them developed a second event of hepatotoxicity. Incidence rate comparison were similar across the treatment arms, when assessing the occurrence of first hepatotoxicity, severe hepatotoxicity, resolved hepatotoxicity and recurring hepatoxicity (Table 2.2).

The occurrence of first hepatotoxicity is in line with other studies conducted in the sub-Saharan Africa, like the Ethopian study which revealed that 24% of patients that were been treated for HIV and TB, developed hepatotoxicity (Yimer, et al., 2014). In contrast to the current study, the same study revealed that 16% of the population developed severe hepatotoxicity. This could be due to misaligned hepatotoxicity classification. For instance, the upper limit of normal (ULN) used for ALT in the Ethopian study was lower compared to the current study, and the grading threshold was different for patients with baseline abnormal LFTs compared to the current study. More

similar to this study was a study conducted in South Africa which revealed that 4.6% of the population developed severe hepatotoxicity (Hoffmann, et al., 2007).

The lack of a unanimous definition for hepatotoxicity has been an obstacle when studying hepatotoxicity in the past; most studies relied on their country specific national health policies to dictate the definition of hepatotoxicity. As a result the hepatotoxicity definitions from different countries were not comparable, this hindered precise comparisons of the incidence and factors associated with hepatotoxicity across regions. Aithal, et al (2011) propose a uniform definition of hepatotoxicity to facilitate ease of comparison of study results. Additionally, a common name for 'hepatotoxicity' is now 'drug induced liver injury' (DILI) (Yimer, et al., 2014; Naidoo, et al., 2019; Aithal, et al., 2011; Naidoo, et al., 2015).

The manner in which hepatotoxicity is defined in this project was not progressive since this project still relied on the guidance from clinicians that worked on the SAPiT trial to advise on the definition of hepatotoxicity. As noted in earlier chapters the definition was based on the DAIDS tables (U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Allergy and Infectious Diseases, Division of AIDS, 2004), while other studies like Hoffmann, et al (2007) cited that they used definitions as specified in the 'AIDS Clinical Trials Network definitions' and Yimer, et al (2014) cited that they have used definitions as specified by Aithal, et al (2011).

The incidence rate of first hepatotoxicity in patients who had an abnormal liver function at baseline in this study was 39.1 and 38.2 per 100 py in the late arm and the sequential arm respectively, this was much higher than 14.1 incidence rates per 100 py observed in the early arm. Suggesting that if a patient presents themselves with an abnormal liver function before TB-treatment/ART initiation, then initiating ART during the intensive phase of TB might curb their chance of developing hepatotoxicity.

Patients with baseline CD4 count greater than 50 cells per mm^3 had an increased likelihood of developing hepatotoxicity post ART initiation in the early arm 28.4 incidence cases per 100 py, this was higher than 16.4 and 13.1 incidence cases of first hepatotoxicity that occurred in the late and sequential arm respectively. This difference is partly because patients in the early arm did not have much time to experience hepatotoxixity prior ART initiation, hence the higher incidence cases post ART initiation compared to the other treatment arm. However, despite the lack of strong evidence, ART initiation could be delayed until the continuation phase of TB treatment for patients with baseline CD4 count greater than 50 cells per mm^3 as this decreases the chance of developing hepatotoxicity for these patients, without putting the lives of these patients at risk,

since mortality rate in the early and late arm were similar (Abdool Karim, et al., 2010), an added advantage is that the chance of the Immune Reconstitution Inflammatory Syndrome (IRIS) after ART initiation is reduced in the late arm compared to the early arm (Naidoo, et al., 2012).

Furthermore, the Cox regression multivariable model demonstrated that the chance of developing first hepatotoxicity is increased for a patient who is older, consumes alcohol, has low baseline total bilirubin, has a positive baseline Hepatitus B surface antigen status and or if the patient is in the late arm instead of the sequential arm (this risk is favorable for the sequential arm - only in the first 5 months of study randomization).

In line with the current study, the following factors associated with the occurrence of hepatotoxicity were identified in other studies: Hepatitus B surface antigen status (Hoffmann, et al., 2007), Age (Naidoo, Evans, Jong, Mellet, & Berhanu, 2015) and alcohol (Pukenyte, et al., 2007). High 'direct' bilirubin has been identified as a risk factor (Yimer, et al., 2014) as opposed to the low 'total' bilirubin that has been identified in this study.

Shortfall of this study on a clinical perspective, is that herbal and traditional remedies were not considered in this study and some of these are known hepatotoxins (Mills, et al., 2005), and are commonly used among ART patients in Africa (Babb, et al., 2007).

The sequential arm is no longer a recommended treatment regimen for patients co-infected with HIV and TB according to the clinical guidelines in South Africa, given that patients in the early and late arm have reduced mortality rates compared to the sequential arm (Abdool Karim, et al., 2010). However, there were parts of this project that indicated that patients in the early and especially late arm have a higher risk of developing hepatoxicity compared to patients in the sequential arm. Moreover, the late arm had a high chance of developing severe hepatotoxicity than the sequential arm. Therefore, close monitoring of the early arm is required post ART initiation as incidence cases of hepatotoxicity were high in this arm compared to the sequential arm. And close monitoring is recommend for the late arm, particulary in the first 5 month of TB treatment as the risk of hepatotoxicity was high for this arm compared the sequential arm.

Patients who do not consume alcohol and or had an abnormal baseline LFT have an increased chance of resolving their first hepatotoxicity. To the best of my knowledge, no study has gone to the extent of analyzing resolution of hepatotoxicity beyond quantifying just the number of patients that resolved hepatotoxicity by the end of the study. And this is an important topic, since through the assessment of resolution of hepatotoxicity it was apparent that in most cases, hepatotoxicity

is mild with a high resolution rate (76%). Therefore, the study of occurence hepatotoxicity should be coupled with the study of resolving hepatotoxicity, to inform clinicians or treating phyisicians of factors associated with resolving hepatotoxicity so that they can relay these to the patients who have developed hepatotoxicity and this might prevent uneccessary treatment interruptions.

The risk of developing second hepatotoxicity is increased in patients who are younger and are in the sequential arm instead of the late arm (this risk is sustained after 5 months of study randomization). Moreover, the assessment of the overall effects on recurring hepatotoxicity were determined through the AG model and shared frailty model, where in both models, the applied time scale is based on a calendar time interval (time since study entry) and event dependence is addressed by a robust variance estimator. The former model assumes that each of the patient's observations are homogenous and the latter model assumes that each of the patient's observations are heterogenous and therefore accounted for accordingly through a random effect (or frailty term). Both models demonstrated indistinguishable results; patients have an increased hazard of developing recurrent hepatotoxicity if a patient is older, consumes alcohol, has a baseline: CD4 count that is greater than 50 cells per mm^3, low total bilirubin, and positive Hepatitus B surface antigen status, if the patient is in the late arm rather than the sequential arm particulary in the first 5 months of study randomization, and if the patient is in the sequential arm rather than the late arm after 5 months of study randomization. The similar model output from the AG and shared frailty model suggests that the heterogeneity within the subjects' observations is minimal.

The assessment of the overall effects on recurring hepatotoxicity was performed yet again through the PWP TT and the PWP GT model and the time scale applied was based on calendar time interval for the former and based on gap time (time since last event) for the latter model. Assuming that each of the patient's observations are homogenous, however, event dependence was accounted for via stratrification instead of variance correction, and the overall effects were obtained through averaging effect for each event per covariate. The PWP TT model demonstrated that patients have an increased hazard of developing recurrent hepatotoxicity: if the patient is older, consumes alcohol, baseline CD4 count is greater than 50 cell per mm^3, have a low a baseline total bilirubin, the Hepatitus B surface antigen is positive and if the patient is in the late arm rather than the sequential arm particulary in the first 5 months of study randomization. Lastly, the PWP GT model, demonstrated that patients have an increased hazard of developing recurring hepatotoxicity: if they have a low baseline total bilirubin, the Hepatitus B surface antigen is positive, if the patient is in the late arm rather than the sequential arm particulary in the first 5

months of study randomization, and if the patient is in the sequential arm rather than the late arm after 5 months of study randomization.

The study of recurring events has been studied for recurrent sports injuries (Ullah, et al., 2012) and studied for infectious diseases like malaria (Sagara, et al., 2014). To my knowledge recurring events of hepatotoxicity have not been studied in literature to the extent that has been done in this project. And as such, studies are needed given that ART is for the longterm and that there is high prevalence of a common risk factor which is chronic hepatitis B in sub-Saharan Africa (Yimer, et al., 2014), therefore patients in sub-Saharan Africa are continually at risk of developing hepatotoxicity. In addition, the distribution of factors associated with hepatotoxicity may differ from those observed during the first year of therapy, and this has been observed through modelling timing of the second hepatotoxicity event.

Recurrent events model use more information about the subject than a traditional Cox model, which improves the power of these models relative to the Cox model. Although power for each model was not quantified in this study, other studies have reported increased power for recurrent models compared to the Cox model (Ullah, et al., 2012). However, it seems as though recurrent event models produce higher standard errors compared to the traditional Cox model, this reduces the precision of the estimates produced by these models. In addition, the AG, shared frailty and the PWP TT model identified CD4 category as marginal predictive factor for hepatotoxicity, this factor was not identified by Cox model in the current study, however, it has been identified in other studies related to hepatotoxicity (Pukenyte, et al., 2007), indicating that considering the recurring event gives additional information about the significant risk factors associated with hepatotoxicity.

Risk factors associated with the development of hepatotoxicity were not the same for the Cox model compared to the recurrent model that modelled second hepatotoxicity. Suggesting that there is a event dependence in the biological mechanism of recurring hepatotoxicity. This narrowed down the appropriate models for recurring hepatotoxicity to two models, PWP TT and PWP GT. The PWP TT had lower AIC compared to the PWP GT, consequentially, it can be deduced that the best model to model hepatotoxicity is the PWP TT.

A method weakness to be noted is regarding the use of the heavyside function to account for non-proportional hazards. It is important to note that the estimate of the hazard ratio after 5 months is not comparable to estimates from a randomised cohort. Given that, the population at risk used to estimate the hazard ratio will already exclude all high risk patients that have experienced hepatotoxicity earlier in the study (before 5 months).

It is important to note that, there was a high volume of censored patients in this study due to loss to follow up and death as shown in Figure 2.2. And it seems as though the former and latter group were mainly comprised of patients who did not initiate ART. The main purpose of this study was to investigate hepatotoxicity in patients who initiated ART during TB treatment, given this, the high volume of censored patients that did not intiate ART may have influenced the results in this study. Therefore, the assumption of non-informative censoring may have been violated. In this regard, further research of modelling hepatotoxicity under a competing risk framework is proposed. This is possible under Islam's multi-state models (Islam, 1994).

In conclusion, although there is an high rate of hepatotoxicity occurrence in the current study it was usually transient or mild and did not necessisate treatment interruption. However, close monitoring of patients especially in the first 5 months of TB-treatment is recommended. It is also recommended that patients with an abnormal liver function at baseline, and have CD4 count less than 50 cell/L mm^3 should be started on ART during the intensive phase as opposed to delaying their ART initiation to the continuation phase. In contrast patients who are healthier (baseline CD4 count greater than 50 cell/L mm3), should rather initiate ART during the continuation phase of TB treatment, so as to avoid high rates of hepatotoxicity.

Understanding the biological mechanism of an infection makes it clear which reccurent event model to apply, if there is minimal knowledge of the infection like the case of hepatotoxicity in this study, it is worthwhile to consider multistate models recommended by Islam (1994) to help choose the recurrent event model that explains the data. The PWP TT model seemed to be the best model for modelling recurring hepatotoxicity, however more research is still required to support this statement.

Appendix

SAS code used

Kaplan Meier Curves

```
proc lifetest data=mss.y2 plots=(s lls)   ;
time &mon*hepa(0);/* the zero is a censor*/
symbol v=none;
strata Group ;
/*where base_res = 0 ;*/
run;

 data failtemp; set est;
  lagsurv=lag(failure);
  do i=1 to 100; if failure eq . then failure=lagsurv;
lagsurv=lag(failure); end;
  if censor eq 0 then failure2=failure;
  surv=1-lagsurv;
 run;
 goptions reset=all gunit=pct  cback=white
ftitle=times ftext=times htitle=3 htext=3 noborder;
axis1 label=(a=90 h=3 'Probabilty of any Hepatotoxicity')
order = (0 to 1 by 0.20)  value=(h=3)  width=2 minor=none;
axis2 label=(h=3 'Months after randomization') order= (0 to 24 by 3)
value=(h=3)  width=2 minor=none;
legend label=none value=('Early arm' 'Late arm' 'Sequential arm')
MODE = PROTECT position=(outside);
/*ods rtf file='D:\Users\F4733061\Desktop\KM3.rtf';*/
proc gplot data=Failtemp;
plot surv*&mon =group /haxis=axis2 vaxis=axis1 /*nolegend;*/
legend=legend1;
symbol1 i=stepjoin color=red v=none width=1 line=1;
symbol2 i=stepjoin color=blue  width=1 v=none line=2;
symbol3 i=stepjoin color=black width=3 v=none line=20 ;
format &mon 10.0;
run;
quit;
```

Cumulative incidence

```
Data mss.cause_of_death_v2;
set "CAUSE OF DEATH V2_0000"n;
```

```
run;

data death_list;
set mss.cause_of_death_v2;
where died=1;

run;

proc sql;
create table Y7
as select a.*,b._1a_Date_Of_Death as Date_Of_Death
from mss.y2 as a
left join death_list as b
On a.participant_id = b.participant_id;
where b._1a_Date_Of_Death ne "";
quit;


data mss.y2_death;
      set Y7;

      if hepa= 0 and Date_Of_Death ne "" then
            do;
                  Days = Date_Of_Death - FSPECIMEN;

                  if Days ne . then
                        do;
                              Hepa = 2;
                              Event =2;
                              Months = Days/30.5;
                              Monthz = Months;
                              Years = Months/12;
                              Yearz= Years;
                        end;
            end;
RUN;

proc sql;
create table Y8
as select a.*,b._1a_Date_Of_Death as Date_Of_Death
from mss.y3_2ndpos as a
left join death_list as b
On a.participant_id = b.participant_id;
where b._1a_Date_Of_Death ne "";
quit;


data mss.y3_death;
      set Y8;

      if hepa= 0 and Date_Of_Death ne "" then
            do;
```

```
                Days = Date_Of_Death - FSPECIMEN;

                if Days ne . then
                     do;
                          Hepa = 2;
                          Event =2;
                          Months = Days/30.5;
                          Monthz = Months;
                          Years = Months/12;
                          Yearz= Years;
                     end;
          end;
RUN;

data First_Second (keep= participant_id Event monthz group);
set
mss.y2_death
mss.y3_death

/*(keep= participant_id Event monthz group)*/
;

run;


ods graphics on;
proc lifetest data=mss.y2_death plots=cif(test) outcif= cif
/*timelist=0.5 1.0 1.5 2.0 4.0 6.0*/;
time Monthz*Event(0)/eventcode=1;
strata Group / order=internal;
/*format Disease diseaseLabel. Gender genderLabel.;*/
run;

 goptions reset=all gunit=pct  cback=white
ftitle=times ftext=times htitle=3 htext=3 noborder;
axis1 label=(a=90 h=3 'Cumulative Incidence Function')
order = (0 to 0.5 by 0.1)  value=(h=3)  width=2 minor=none;
axis2 label=(h=3 'Months after randomization') order= (0 to 24 by 3)
value=(h=3)  width=2 minor=none;
legend label=none value=('Early arm' 'Late arm' 'Sequential arm')
MODE = PROTECT position=(outside);
/*ods rtf file='D:\Users\F4733061\Desktop\KM3.rtf';*/
proc gplot data=cif;
plot cif*monthz = group /haxis=axis2 vaxis=axis1 /*nolegend;*/
legend=legend1;
symbol1 i=stepjoin color=red v=none width=1 line=1;
symbol2 i=stepjoin color=blue  width=1 v=none line=2;
symbol3 i=stepjoin color=black width=3 v=none line=20 ;
format monthz 10.0;
run;
quit;
```

```
/*First and recurring event*/


proc lifetest data=first_second plots=cif(test) outcif= cif
/*timelist=0.5 1.0 1.5 2.0 4.0 6.0*/;
time Monthz*Event(0)/eventcode=1;
strata Group / order=internal;
/*format Disease diseaseLabel. Gender genderLabel.;*/
run;

 goptions reset=all gunit=pct  cback=white
ftitle=times ftext=times htitle=3 htext=3 noborder;
axis1 label=(a=90 h=3 'Cumulative Incidence Function')
order = (0 to 0.5 by 0.1)  value=(h=3)  width=2 minor=none;
axis2 label=(h=3 'Months after randomization') order= (0 to 24 by 3)
value=(h=3)  width=2 minor=none;
legend label=none value=('Early arm' 'Late arm' 'Sequential arm')
MODE = PROTECT position=(outside);
/*ods rtf file='D:\Users\F4733061\Desktop\KM3.rtf';*/
proc gplot data=cif;
plot cif*monthz = group /haxis=axis2 vaxis=axis1 /*nolegend;*/
legend=legend1;
symbol1 i=stepjoin color=red v=none width=1 line=1;
symbol2 i=stepjoin color=blue  width=1 v=none line=2;
symbol3 i=stepjoin color=black width=3 v=none line=20 ;
format monthz 10.0;
run;
quit;
```

Cox proportional hazard model

```
/*Does not outputs Hazard ratios*/

Proc PHREG data=MSS.NEW_STACKHEPA covm covs /*Plot(overlay) =
cumhaz*/;
      class
      Group (ref= 'Sequential Arm')
     gender(ref='0')
      alcohol2(ref='Not a consumer')
      cd4cat (ref='0')
      hbv (ref='0')
      res (ref='0')

/param=ref order=internal
      ;
      model Months*Hepa(0) =
      Group
```

```
      age
     gender
      Alcohol2
     cd4cat
      bil
      hbv
      res
      Alcohol2*hbv
      bil*hbv

      / risklimits ties=efron;
      /*GRage = age_C*months;*/
      Baseline  out=_null_;
/*    strata res ;*/
run;

/*Outputs Hazard ratios*/

DATA cox_stack2;
      SET mss.new_stackhepa;
      Format groupy1 1. groupy2 1.;

      if group = "Early Arm" then
            groupy1 = "1";
      else groupy1 = "0";

      if group = "Late Arm" then
            groupy2 = "1";
      else groupy2 = "0";

      Format alc1 1. alc2 1.;

      if alcohol2 = 'Not a consumer' then
            alc2 = "0";
      else alc2 = "1";

      format months 30.;

      tstop =months;
      if tstop < =5 then
            gt1 = 1;
      else gt1 = 0;

      if tstop > 5 then
            gt2 = 1;
      else gt2 = 0;

      if bil < =8.43 then
            bilcat = 1;
      else bilcat = 0;
RUN;
```

```
Proc PHREG data=cox_stack2 covm covs /*Plot(overlay) = cumhaz*/;
      class
      Group (ref= 'Sequential Arm')
     gender(ref='0')
      alcohol2(ref='Not a consumer')
      cd4cat (ref='0')
      hbv (ref='0')
      res (ref='0')

/param=ref order=internal
      ;
      model Months*Hepa(0) =
      Group
      age
     gender
      Alcohol2
     cd4cat
      bil
      hbv
      res
/*    Alc_hbv*/
/*    Bil_hbv*/

      / risklimits ties=efron;
/*    Alc_hbv = Alc2*hbv;*/
/*    Bil_hbv = bil*hbv;*/
      Baseline  out=_null_;
/*    strata res ;*/
run;
```

Cox-Snell residuals

```
Proc PHREG data=MSS.NEW_STACKHEPA covm covs /*Plot(overlay) =
cumhaz*/;
      class
      Group (ref= 'Sequential Arm')
     gender(ref='0')
      alcohol2(ref='Not a consumer')
      cd4cat (ref='0')
      hbv (ref='0')
      res (ref='0')

/param=ref order=internal
      ;
      model Months*Hepa(0) =
      Group
      age
     gender
      Alcohol2
     cd4cat
```

```
      bil
      hbv
      res
/*    Alcohol2*hbv*/
/*    bil*hbv    */
           /alpha=0.05 risklimits ties=efron;
      ID participant_id;
/*    where tstart < tstop;*/
  output out = figure11_1
LOGSURV = h
/method = ch;
/*  -logsurv is the cox-snell residual*/
run;

data figure11_1a;
  set figure11_1;
  h = -h;
  cons = 1;
run;
/* Creating gradient 1 */
proc phreg data = figure11_1a ;
  model  h*hepa(0) = cons;
  output out = figure11_1b logsurv = ls /method = ch;
run;
data figure11_1c;
  set figure11_1b;
    haz = - ls;
run;
proc sort data = figure11_1c;
 by h;
run;
title "Cox Snell Residual";
axis1 order = (0 to 3 by .5) minor = none;
axis2 order = (0 to 3 by .5) minor = none label = ( a=90);
symbol1 i = stepjl c= blue;
symbol2 i = join c = red l = 3;
proc gplot data = figure11_1c;
  plot haz*h =1 h*h =2 /overlay haxis=axis1 vaxis= axis2;
  label haz = "Estimated Cumulative Hazard Rates";
  label h = "Cox-Snell Residual";
run;
quit;
```

Martingale Residual

```
proc phreg data = COX_STACK1;
 class
    Treatment_Arm (ref='Sequential Arm')
```

```
/*    age1(ref="(<=34.5)") */
      gender1(ref="Female")
      CD4_category(ref="CD4<0.05x10^9")
      hbv1(ref="Negative")
      alcohol (ref ='Not a consumer')
    Baseline_LFT (ref ="Normal")
;
model months*hepa(0)=
    Treatment_Arm
      age
      gender1
      alcohol
      CD4_category
      hbv1
      bil
    Baseline_LFT
    alcohol*hbv1
      hbv1*bil

           /alpha=0.05 risklimits ties=efron;
      ID participant_id;
/*    where tstart < tstop;*/
output out=resid resmart=mart resdev=dev;
run;
proc gplot data=resid;
/*title "Schoenfeld Residuals for Treatment arm ";*/
plot (mart)*Treatment_arm
/ CFRAME=white OVERLAY VAXIS=axis1 HAXIS=axis2 FRAME VREF=0 VMINOR=0
HMINOR=0
CAXIS = blue NAME='plot3';
symbol value=dot i=sm60s h=0.5 w=3;
axis1 label =(a=90 r=0 f='Arial''Schoenfeld Residuals')
value=(f='Arial' );
axis2 label=( f='Arial' 'Age')value=(f='Arial') ;
RUN;
QUIT;


proc sort data= resid out=a;
by group;
run;

ods graphics on;
ods graphics off;
options nogstyle;
title ' Treatment Arm';
   proc boxplot data=a;
/*   by group;*/
   id Participant_ID;
      plot mart*group
/
/*            nohlabel*/
```

```
         boxstyle      = schematic
         boxwidthscale = 1
/*         bwslegend*/
            cboxfill=white
            cboxes=blue
            font='Arial'
            CLABEL=black
            BOXWIDTH=10

            HEIGHT= 4.5
            CLIPSYMBOL=Red
/*          haxis=axis2*/
            idcolor=red
            idsymbol=diamond
;
symbol1 /*color=blue*/ w=2.5 h=2.5 height=2 value=dot;
 label group = 'Treatment Arm';

   run;

   options gstyle;
goptions reset=symbol ;

/*Gender*/
proc sort data= resid out=ab;
by gender1;
run;
options nogstyle;
/*title ' Treatment Arm';*/
   proc boxplot data=ab;
/*   by group;*/
   id Participant_ID;
      plot mart*gender1
/
/*            nohlabel*/
         boxstyle      = schematic
         boxwidthscale = 1
/*         bwslegend*/
            cboxfill=white
            cboxes=blue
            font='Arial'
            CLABEL=black
            BOXWIDTH=10
            HEIGHT= 5
            CLIPSYMBOL=Red
/*          haxis=axis2*/
            idcolor=red
            idsymbol=diamond
;
symbol1 /*color=blue*/ w=2.5 h=2.5 height=2 value=dot;
 label gender1 = 'Sex';
```

```
   run;

   options gstyle;
goptions reset=symbol ;

proc gplot data=resid;
/*title "Schoenfeld Residuals for Treatment arm ";*/
plot (mart)*age
/ CFRAME=white OVERLAY VAXIS=axis1 HAXIS=axis2 FRAME VREF=0 VMINOR=0
HMINOR=0
CAXIS = blue NAME='plot3' ;
symbol1 color=black interpol= value=dot i=sm60s h=0.5 w=3;
/*symbol2 color=red value=line i=sm60s h=0.5 w=3;*/
axis1 label =(a=90 r=0 f='Arial''Martingale Residuals')
value=(f='Arial' );
axis2 label=( f='Arial' 'Age')value=(f='Arial') ;
RUN;
QUIT;

/**/
```

Deviance residual

```
proc phreg data = COX_STACK1;
 class
    Treatment_Arm (ref='Sequential Arm')
/*    age1(ref="(<=34.5)") */
      gender1(ref="Female")
      CD4_category(ref="CD4<0.05x10^9")
      hbv1(ref="Negative")
      alcohol (ref ='Not a consumer')
    Baseline_LFT (ref ="Normal")
;
model months*hepa(0)=
    Treatment_Arm
      age
      gender1
      alcohol
      CD4_category
      hbv1
      bil
    Baseline_LFT
    alcohol*hbv1
      hbv1*bil

            /alpha=0.05 risklimits ties=efron;
      ID participant_id;
/*    where tstart < tstop;*/
output out=resid resmart=mart resdev=dev;
run;
```

```
proc gplot data=resid;
/*title "Schoenfeld Residuals for Treatment arm ";*/
plot (mart)*Treatment_arm
/ CFRAME=white OVERLAY VAXIS=axis1 HAXIS=axis2 FRAME VREF=0 VMINOR=0
HMINOR=0
CAXIS = blue NAME='plot3';
symbol value=dot i=sm60s h=0.5 w=3;
axis1 label =(a=90 r=0 f='Arial''Schoenfeld Residuals')
value=(f='Arial' );
axis2 label=( f='Arial' 'Age')value=(f='Arial') ;
RUN;
QUIT;


proc sort data= resid out=a;
by group;
run;

ods graphics on;
ods graphics off;
options nogstyle;
title ' Treatment Arm';
   proc boxplot data=a;
/*   by group;*/
   id Participant_ID;
      plot dev*group
/
/*            nohlabel*/
         boxstyle      = schematic
         boxwidthscale = 1
/*         bwslegend*/
           cboxfill=white
           cboxes=blue
           font='Arial'
           CLABEL=black
           BOXWIDTH=10

           HEIGHT= 4.5
           CLIPSYMBOL=Red
/*         haxis=axis2*/
           idcolor=red
           idsymbol=diamond
;
symbol1 /*color=blue*/ w=2.5 h=2.5 height=2 value=dot;
 label group = 'Treatment Arm';

   run;

   options gstyle;
goptions reset=symbol ;
```

```
/*Gender*/
proc sort data= resid out=ab;
by gender1;
run;
options nogstyle;
/*title ' Treatment Arm';*/
   proc boxplot data=ab;
/*   by group;*/
   id Participant_ID;
      plot dev*gender1
/
/*            nohlabel*/
         boxstyle      = schematic
         boxwidthscale = 1
/*         bwslegend*/
           cboxfill=white
           cboxes=blue
           font='Arial'
           CLABEL=black
           BOXWIDTH=10
           HEIGHT= 5
           CLIPSYMBOL=Red
/*         haxis=axis2*/
           idcolor=red
           idsymbol=diamond
;
symbol1 /*color=blue*/ w=2.5 h=2.5 height=2 value=dot;
 label gender1 = 'Sex';

   run;

   options gstyle;
goptions reset=symbol ;

proc gplot data=resid;
/*title "Schoenfeld Residuals for Treatment arm ";*/
plot (dev)*age
/ CFRAME=white OVERLAY VAXIS=axis1 HAXIS=axis2 FRAME VREF=0 VMINOR=0
HMINOR=0
CAXIS = blue NAME='plot3' ;
symbol1 color=black interpol= value=dot i=sm60s h=0.5 w=3;
/*symbol2 color=red value=line i=sm60s h=0.5 w=3;*/
axis1 label =(a=90 r=0 f='Arial''Martingale Residuals')
value=(f='Arial' );
axis2 label=( f='Arial' 'Age')value=(f='Arial') ;
RUN;
QUIT;

Testing for non proportional hazard

Proc PHREG data=MSS.NEW_STACKHEPA covm covs /*Plot(overlay) =
cumhaz*/;
```

```
      class
      Group (ref= 'Sequential Arm')
      gender(ref='0')
     alcohol2(ref='Not a consumer')
      cd4cat (ref='0')
      hbv (ref='0')
      res ( ref='0')
      ;
      model Months*Hepa(0) =
      Group
      age
      gender
      Alcohol2
      cd4cat
      bil
      hbv
     res
/*    alcohol2*hbv*/
/*    hbv*bil*/
      group*fun
      gender*fun
      age*fun
      alcohol2*fun
      cd4cat*fun
      bil*fun
      hbv*fun
      res*fun


      /*         a b*/
      / risklimits ties=efron;
      /*GRage = age_C*months;*/
      Baseline  out=_null_;
      where months not in (0 .);
      fun=log(months);

      /*         a=   group*fun;*/
      /*         b=gender*fun;*/
      /*         age*fun*/
      /*         alcohol1*fun*/
      /*         cd4cat*fun*/
      /*         bil*fun*/
      /*         hbv*fun*/
      /*     hazardratio 'Group' Group /cl=both;*/
      /*     hazardratio 'Gender' Gender / cl=both;*/
      /*     hazardratio 'Age' Age / diff=ref cl=both*/
      ;
run;
```

Log-log test

```
/*proc summary data=cox_stack nway missing;*/
/*var viral_load baseast basealt alk bil ldh  ;*/
/*output out=ab MEDIAN=;*/
/*run;*/


data cox_stack1;
      set mss.new_stackhepa;
      Format Treatment_Arm $20.;

    Treatment_Arm = group;


      Format age1 $15.;


    if age > 34.5 then age1 = "Age >34.5";
    else age1 = "Age <=34.5";


      Format CD4_category  $30.;

      if cd4cat = "1" then
            CD4_category = "CD4<0.05x10^9";
      else if cd4cat = "0" then
            CD4_category = "CD4>=0.05x10^9";
      Format hbv1 $30.;

if hbv = "1" then
            hbv1 = "Positive";
      else if hbv = "0" then  hbv1 = "Negative";
      Format Baseline_Bilirubin $30.;


      if bil < 7 then
            Baseline_Bilirubin = "Bilirubin <7";
      else Baseline_Bilirubin = "Bilirubin >=7";


      drop age gender    HBV   bil ;
      Rename age1 =Age gender1 =Sex  alcohol2 =alcohol
            hbv1= HBV  months= time  ;


      format Baseline_LFT $15.;

      if res = 1 then Baseline_LFT = "Abnormal";
    else if res = 0 then Baseline_LFT = "Normal";
```

```
     ;
run;

%Macro lls(VAR);
ods graphics on;
/*%let var = Treatment_arm;*/
proc lifetest data=cox_stack1 method=km OUTSURV=CURVES&VAR. ;
time Time*HEPA(0);
STRATA &VAR.  ;
/*where CD4_category = "CD4>=0.05x10^9";*/
RUN;

DATA CAT&VAR.;
SET CURVES&VAR.;
LS=-LOG(SURVIVAL);
LLS=LOG(-LOG(SURVIVAL));
Neg_LLS=-LOG(-LOG(SURVIVAL));
/*    if Treatment_Arm = "Early arm" then lls_early = neg_lls;*/
/*    else lls_early = .;*/
/*    if    Treatment_Arm = "Late arm" then lls_late = neg_lls;*/
/*    else lls_late = .;*/
/*    if  Treatment_Arm = "Sequential arm" then lls_seq = neg_lls;*/
/*    else lls_seq = .;*/
if Neg_LLS > -0.000000001;
RUN;
/*SYMBOL COLOR=BLUE;*/
/*SYMBOL2 COLOR=RED;*/
/*SYMBOL2 COLOR=GREEN;*/

symbol color=black  width=2 v=none line=1  ;
symbol2 color=blue  width=2 v=none line=1;
symbol3 color=red width=2 v=none line=1 ;

proc gplot data=CAT&VAR.;
plot Neg_LLS*time=&VAR.
/ CFRAME=white OVERLAY VAXIS=axis1 HAXIS=axis2 FRAME VREF=0 VMINOR=0
HMINOR=0
CAXIS = black NAME='plot3'    ;
symbol color=black  width=2 v=none line=1  ;
symbol2 color=blue  width=2 v=none line=1;
symbol3 color=red width=2 v=none line=1 ;
axis1 label =(a=90 r=0 f='Arial''Schoenfeld Residuals')
value=(f='Arial' );
/*axis2 label=( f='Arial' 'Age')value=(f='Arial') ;*/
;
/*where Treatment_Arm in ( "Early arm"  "Late arm"  "Sequential arm")
;*/
run;
quit;

%Mend;
```

```
%lls(Baseline_LFT);
```

Cox regression model with non-proportional hazard

```
DATA cox_stack2;
      SET mss.new_stackhepa;
      Format groupy1 1. groupy2 1.;

      if group = "Early Arm" then
            groupy1 = "1";
      else groupy1 = "0";

      if group = "Late Arm" then
            groupy2 = "1";
      else groupy2 = "0";

      Format alc1 1. alc2 1.;

      if alcohol1 = "Occasionally" then
            alc1 = "1";
      else alc1 = "0";

      if alcohol1 = "Frequently" then
            alc2 = "1";
      else alc2 = "0";


      format months 30.;

      tstop =months;
      if tstop < =5 then
            gt1 = 1;
      else gt1 = 0;

      if tstop > 5 then
            gt2 = 1;
      else gt2 = 0;

      if bil < =8.43 then
            bilcat = 1;
      else bilcat = 0;
RUN;

%let mon = 4;
proc phreg data = cox_stack2 covm covs(aggregate);
      class
            gender(ref='0')
            cd4cat(ref='0')
```

```
            alcohol2 (ref='Not a consumer')
/*          cd8(ref='0') */
            hbv(ref='0')
         res (ref='0')
      ;
      model Months*hepa(0)=

            groupy11
            groupy12
            groupy21
            groupy22
            age
            gender
            alcohol2
            cd4cat
            bil
            hbv
            res



            /alpha=0.05 risklimits ties=efron;

      if Months <= &mon. then
            groupy11 = groupy1;
      else groupy11 = 0;

      if Months > &mon. then
            groupy12 = groupy1;
      else groupy12 = 0;

      if Months <= &mon. then
            groupy21 = groupy2;
      else groupy21 = 0;

      if Months > &mon. then
            groupy22 = groupy2;
      else groupy22 = 0;

      where months>0;

      id participant_id;
run;
```

Resolution multivariable model

```
%let mon = 5;
proc phreg data = cox_stack2_res covm covs(aggregate);
      class
```

```
            gender(ref='0')
            cd4cat(ref='0')
            alcohol1 (ref="Never")
/*          cd8(ref='0') */
            hbv(ref='0')
      ;
      model Months*reso(0)=

            groupy11
            groupy12
            groupy21
            groupy22
            age
            gender
            alcohol1
            cd4cat
            bil
            hbv



            /alpha=0.05 risklimits ties=efron;

      if Months <= &mon. then
            groupy11 = groupy1;
      else groupy11 = 0;

      if Months > &mon. then
            groupy12 = groupy1;
      else groupy12 = 0;

      if Months <= &mon. then
            groupy21 = groupy2;
      else groupy21 = 0;

      if Months > &mon. then
            groupy22 = groupy2;
      else groupy22 = 0;

      where months>0;

      id participant_id;
run;
```

Recurrent multivariable model

```
%let mon = 5;
proc phreg data = cox_stack2_r covm covs(aggregate);
      class
            gender(ref='0')
```

```
            cd4cat(ref='0')
            alcohol1 (ref="Never")
/*          cd8(ref='0') */
            hbv(ref='0')
      ;
      model Months*hepar(0)=

            groupy11
            groupy12
            groupy21
            groupy22
            age
            gender
            alcohol1
            cd4cat
            bil
            hbv



            /alpha=0.05 risklimits ties=efron;

      if Months <= &mon. then
            groupy11 = groupy1;
      else groupy11 = 0;

      if Months > &mon. then
            groupy12 = groupy1;
      else groupy12 = 0;

      if Months <= &mon. then
            groupy21 = groupy2;
      else groupy21 = 0;

      if Months > &mon. then
            groupy22 = groupy2;
      else groupy22 = 0;

      where months>0;

      id participant_id;
run;
```

Andersen Gill model

```
DATA mss.ag_stack2;
      SET mss.ag_stack;
      Format groupy1 1. groupy2 1.;

      if group = "Early Arm" then
```

```
            groupy1 = "1";
      else groupy1 = "0";

      if group = "Late Arm" then
            groupy2 = "1";
      else groupy2 = "0";



      format months 30.;

      tstop =months;
      if tstop < =5 then
            gt1 = 1;
      else gt1 = 0;

      if tstop > 5 then
            gt2 = 1;
      else gt2 = 0;

      if bil < =8.43 then
            bilcat = 1;
      else bilcat = 0;
RUN;


proc phreg data = mss.ag_stack2 covm covs(aggregate);
      class
            gender(ref='0')
         alcohol1 (ref = 'Never')
            cd4cat(ref='0')
            hbv(ref='0')
      ;
      model (tstart, tstop)*hepa(0)=

            groupy11
            groupy12
            groupy21
            groupy22
            age
            gender
         alcohol1
            cd4cat
            bil
            hbv
            interval

            /alpha=0.05 risklimits;

      if time <= 5  then
            groupy11 = groupy1;
```

```
      else groupy11 = 0;

      if time > 5 then
            groupy12 = groupy1;
      else groupy12 = 0;

      if time <= 5 then
            groupy21 = groupy2;
      else groupy21 = 0;

      if time > 5 then
            groupy22 = groupy2;
      else groupy22 = 0;
      id participant_id;
run;
```

Shared frailty model

```
%let mon = 5;

proc phreg data = mss.ag_stack2   covm  covs(aggregate);
      class
          Participant_ID
            gender(ref='0')
            cd4cat(ref='0')
            alcohol1 (ref="Never")
            hbv(ref='0')
      ;
      model (tstart, tstop)*hepa(0)=

            groupy11
            groupy12
            groupy21
            groupy22
            age
            gender
            alcohol1
            cd4cat
            bil
            hbv



            /alpha=0.05 risklimits ties=efron;

      if tstop <= &mon. then
            groupy11 = groupy1;
      else groupy11 = 0;
```

```
      if tstop > &mon. then
            groupy12 = groupy1;
      else groupy12 = 0;

      if tstop <= &mon. then
            groupy21 = groupy2;
      else groupy21 = 0;

      if tstop > &mon. then
            groupy22 = groupy2;
      else groupy22 = 0;

      id participant_id;
      random participant_id;
      where tstart < tstop or time ne 0;;
run;
```

PWP TT

```
%let mon = 5;

proc phreg data = mss.ag_stack2  covs(aggregate);
      class
            gender(ref='0')
            cd4cat(ref='0')
            alcohol1 (ref="Never")
            hbv(ref='0')
      ;
      model (tstart, tstop)*hepa(0)=

            groupy11
            groupy12
            groupy21
            groupy22
            age
            gender
            alcohol1
            cd4cat
            bil
            hbv



            /alpha=0.05 risklimits ties=efron;

      if tstop <= &mon. then
            groupy11 = groupy1;
      else groupy11 = 0;
```

```
      if tstop > &mon. then
            groupy12 = groupy1;
      else groupy12 = 0;

      if tstop <= &mon. then
            groupy21 = groupy2;
      else groupy21 = 0;

      if tstop > &mon. then
            groupy22 = groupy2;
      else groupy22 = 0;

      id participant_id;
      where tstart < tstop or time ne 0;
      strata interval;
run;
```

PWP GT

```
%let mon = 5;

proc phreg data = mss.ag_stack2  covs(aggregate);
      class
            gender(ref='0')
            cd4cat(ref='0')
            alcohol1 (ref="Never")
            hbv(ref='0')
      ;
      model time*hepa(0)=

            groupy11
            groupy12
            groupy21
            groupy22
            age
            gender
            alcohol1
            cd4cat
            bil
            hbv



            /alpha=0.05 risklimits ties=efron;

      if time <= &mon. then
            groupy11 = groupy1;
      else groupy11 = 0;

      if time > &mon. then
```

```
            groupy12 = groupy1;
      else groupy12 = 0;

      if time <= &mon. then
            groupy21 = groupy2;
      else groupy21 = 0;

      if time > &mon. then
            groupy22 = groupy2;
      else groupy22 = 0;

      id participant_id;
      where tstart < tstop or time ne 0;
      strata interval;
run;
```

Forest Plot for recurrent model

```
%let dpi=100;
%let w=9in;
%let h=14in;

/*"|| "(" || put(l) || ", " ||put(h) ||")";*/

/*--Leading blanks in the "Baseline Exposure Variable"n variable must
be non--blank spaces  --*/
/*--Use character value 'A0', or copy from Windows System Character
Map--*/
/*--Regular leading blanks will be stripped, losing the indentation
--*/
data forest;
      input index Indent "Baseline Exposure Variable"n $6-47 "p-value"n
Hazard  Low  High;
      format "p-value"n 7.4 ;
      datalines;

10 0 Early arm vs Sequential arm (months <=5)..... . . . .
11 2 ..Cox ..................................... 0.9335 0.973 0.515
1.839
12 2 ..AG ...................................... 0.1590 2.155 0.740
6.271
13 2 ..Frailty.................................. 0.1535 2.210 0.744
6.564
14 2 ..PWP TT................................... 0.1574 2.169 0.742
6.341
15 2 ..PWP GT................................... 0.2618 1.687 0.677
4.206
16 0 ........................................... . . . .
```

17 0 Early arm vs Sequential arm (months >5)..... . . . .
18 2 ..Cox 0.8074 0.933 0.534
1.629
19 2 ..AG 0.3104 0.784 0.491
1.254
20 2 ..Frailty..... 0.3405 0.787 0.481
1.287
21 2 ..PWP TT..... 0.3489 0.799 0.499
1.278
22 2 ..PWP GT..... 0.4606 0.842 0.534
1.329
23 0 ....
24 0 Late arm vs Sequential arm (months <=5)..... . . . .
25 2 ..Cox 0.0912 1.648 0.923
2.940
26 2 ..AG 0.0004 5.699 2.191
14.828
27 2 ..Frailty..... 0.0003 6.062 2.282
16.101
28 2 ..PWP TT..... 0.0004 5.703 2.194
14.825
29 2 ..PWP GT..... 0.0042 3.170 1.437
6.991
30 0 ....
31 0 Late arm vs Sequential arm (months >5)..... . . . .
32 2 ..Cox 0.5986 0.861 0.494
1.502
33 2 ..AG 0.0725 0.648 0.403
1.041
34 2 ..Frailty..... 0.0898 0.643 0.387
1.071
35 2 ..PWP TT..... 0.1043 0.676 0.422
1.084
36 2 ..PWP GT..... 0.0834 0.670 0.425
1.054
37 0 ....
38 0 Age..... . . . .
39 2 ..Cox 0.0139 1.025 1.005
1.045
40 2 ..AG 0.0233 1.023 1.003
1.043
41 2 ..Frailty..... 0.0419 1.023 1.001
1.046
42 2 ..PWP TT..... 0.0277 1.023 1.002
1.043
43 2 ..PWP GT..... 0.1008 1.016 0.997
1.035
44 0 ....
45 0 Sex..... . . . .
46 2 ..Cox 0.7618 1.061 0.724
1.554

47 2 ..AG 0.6087 1.102 0.761
1.596
48 2 ..Frailty..... 0.6241 1.107 0.738
1.660
49 2 ..PWP TT..... 0.6681 1.085 0.746
1.579
50 2 ..PWP GT..... 0.6145 1.092 0.774
1.541
51 0 ....
52 0 Alcohol..... . . . .
53 2 ..Cox 0.0346 1.702 1.039
2.789
54 2 ..AG 0.0220 1.801 1.088
2.979
55 2 ..Frailty..... 0.0103 1.908 1.164
3.126
56 2 ..PWP TT..... 0.0181 1.812 1.107
2.965
57 2 ..PWP GT..... 0.1315 1.388 0.906
2.126
58 0 ....
59 0 CD4 category..... . . . .
60 2 ..Cox 0.1397 1.417 0.892
2.251
61 2 ..AG 0.0761 1.543 0.955
2.491
62 2 ..Frailty..... 0.0666 1.597 0.969
2.633
63 2 ..PWP TT..... 0.0670 1.555 0.970
2.494
64 2 ..PWP GT..... 0.1538 1.373 0.888
2.122
65 0 ....
66 0 BILI..... . . . .
67 2 ..Cox 0.0050 0.944 0.907
0.983
68 2 ..AG 0.0039 0.943 0.907
0.981
69 2 ..Frailty..... 0.0036 0.940 0.901
0.980
70 2 ..PWP TT..... 0.0035 0.941 0.904
0.980
71 2 ..PWP GT..... 0.0162 0.954 0.917
0.991
72 0 ....
73 0 HBsAg status..... . . . .
74 2 ..Cox 0.0042 2.097 1.262
3.483
75 2 ..AG 0.0087 1.883 1.174
3.021
76 2 ..Frailty..... 0.0246 1.886 1.084
3.280

```
77 2 ..PWP TT.............................. 0.0066 2.003 1.213
3.306
78 2 ..PWP GT.............................. 0.0203 1.678 1.084
2.599
79 0 ..................................... . . . .
80 0 Baseline LFT.......................... . . . .
81 2 ..Cox ................................. 0.3499 0.807 0.515
1.265
82 2 ..AG .................................. 0.2394 0.764 0.488
1.196
83 2 ..Frailty.............................. 0.2370 0.748 0.462
1.211
84 2 ..PWP TT.............................. 0.2134 0.728 0.442
1.200
85 2 ..PWP GT.............................. 0.2647 0.763 0.475
1.227
86 0 ..................................... . . . .
```

```
;
run;

/*--Replace '.' in "Baseline Exposure Variable"n with blank--*/
data forest2;
  set forest;
  format h $5. lo $5. hi $5. ;
      zero=0;
      h=put(hazard,5.2);
      lo=put(low,5.2);
      hi=put(high,5.2);
      "Hazard Ratio (95% CI)"n= put(h, $5.) ||" " || "(" || put(lo,$5.)
|| ", " ||put(hi,$5.) ||")";
      if hazard =. then "Hazard Ratio (95% CI)"n = "";

  "Baseline Exposure Variable"n=translate("Baseline Exposure
Variable"n, ' ', '.');
/*  val=mod(_N_-1, 6);*/
/*  indent=ifn(indent eq 2, 1, 0);*/
/*  if val eq 1 or val eq 2 or val eq 3 then ref="Baseline Exposure
Variable"n;*/
  run;

/*--Create font with smaller fonts for axis label, value and data--*/
/*--Define templage for Forest Plot--*/
/*--Template uses a Layout Lattice of 6 columns--*/
%let title =Forest plot of the univariable Cox Model Hazard Ratios
(95% CI) ;

proc template;
      define statgraph Forest;
            dynamic _show_bands _color _thk _font;
```

```
            /*          STYLE  DATA /*/
            /*                              FONT_FACE = "Arial, Helvetica,
sans-serif"*/
            /*                              FONT_SIZE = 5*/
            /*                              FONT_WEIGHT = bold*/
            /*                              FONT_STYLE = italic*/
            /*                              FOREGROUND = cx002288*/
            /*                              BACKGROUND = cxe0e0e0;*/
            begingraph;

                  /*               entrytitle 'Univariable Cox Model';*/
                  discreteattrmap name='text';
                  value '0' / textattrs=(weight=bold size=14);
                  value other;
                  enddiscreteattrmap;
                  discreteattrvar attrvar=type var=indent
attrmap='text';
                  rangeattrmap name="ResponseRange";
                  range min-0.05    / rangeAltColor=RED;    /* or use the
OVER or UNDER keyword */
                  range 0.05-0.1 / rangeAltColor=Orange;     /* color for
missing values */
                  range other / rangeAltColor=black;     /* color for
missing values */
                  endrangeattrmap;
                  rangeattrvar var="p-value"n
/* specify response variable in data set */
                  attrmap="ResponseRange"         /* specify custom color
ramp */
                  attrvar=RangeVar;               /* alias for this
variable/ramp combination */
                  layout lattice / columns=4 columnweights=( 0.45 0.4
0.17 0.09 );

                        /*--Column headers--*/
                        /*      sidebar / align=top;*/
                        /*        layout lattice / rows=2 columns=4
columnweights=(0.2 0.25 0.25 0.3);*/
                        /*          entry textattrs=(size=8) halign=left
""Baseline Exposure Variable"n";*/
                        /*          entry textattrs=(size=8) halign=left
" No.of Patients (%)";*/
                        /*          entry textattrs=(size=8) halign=left
"Hazard Ratio";*/
                        /*          entry halign=center
textattrs=(size=8) "4-Yr Cumulative Event Rate" ;*/
                        /*          entry " "; */
                        /*          entry " "; */
                        /*          entry " "; */
                        /*          entry halign=center
textattrs=(size=8) "Medical Therapy";*/
```

```
                        /*        endlayout;*/
                        /*      endsidebar;*/
                        /*--First "Baseline Exposure Variable"n column,
shows only the Y2 axis--*/
/*                      layout overlay / walldisplay=none
xaxisopts=(display=none) */
/*                            yaxisopts=(reverse=true display=none */
/*                            tickvalueattrs=(weight=bold));*/
/*                            referenceline y=index /
lineattrs=(thickness=_thk color=_color);*/
/*                            axistable y=index value=index*/
/*                                  / /*indentweight=indent
textgroup=type*/*/
/*                            valueattrs= (family='arial' size = 10)*/
/*                                  labelALIGN =left labelATTRS =
(weight=bold);*/
/*                      endlayout;*/
;
                        /*--First "Baseline Exposure Variable"n column,
shows only the Y2 axis--*/
                        layout overlay / walldisplay=none
xaxisopts=(display=none)
                              yaxisopts=(reverse=true display=none
                              tickvalueattrs=(weight=bold));
                              referenceline y=index/
lineattrs=(thickness=_thk color=_color);
                              axistable y=index value="Baseline Exposure
Variable"n
                                    / indentweight=indent
/*textgroup=type*/
                              valueattrs= (family='arial' size = 10)
                                    labelALIGN =left labelATTRS =
(weight=bold);
                        endlayout;

                        /*--Third column showing odds ratio graph--*/
                        layout overlay / xaxisopts =(label='Hazard Ratio
(95% CI) plot per exposure variable'
                              linearopts=(tickvaluepriority=true
                              tickvaluelist=(0.0 0.5 1.0 1.5 2.0 2.5 3.0
6.0 9.0 12.0 15.0 18.0)) tickvalueattrs=(size=8 family='arial'))
                              yaxisopts=(reverse=true display=none)
walldisplay=none;
                              referenceline y=index /
lineattrs=(thickness=_thk color=_color);
                              scatterplot y=index x=hazard /
xerrorlower=low xerrorupper=high  errorbarattrs=(color=blr)
                                    markerattrs=(symbol=circlefilled
color=blr) markercolorgradient=RangeVar;
                              referenceline x=1;
                        endlayout;
```

```
                        /*--Second column showing Count and percent--*/
                        layout overlay / xaxisopts=(display=none)
                              yaxisopts=(reverse=true display=none)
walldisplay=none;
                              referenceline y=index /
lineattrs=(thickness=_thk color=_color);
                              axistable y=index value="Hazard Ratio (95%
CI)"n /
                                    valueattrs= (family='arial' size = 10)
                                    labelALIGN =left labelATTRS =
(weight=bold);
                        endlayout;

                        /*--Sixth column showing P-Values--*/
                        layout overlay / x2axisopts=(display=(tickvalues)
offsetmin=0.25 offsetmax=0.25)
                              yaxisopts=(reverse=true display=none)
walldisplay=none;
                              referenceline y=index /
lineattrs=(thickness=_thk color=_color);
                              axistable y=index value="p-value"n /
                                    display=(label) labelposition=min
valueattrs= (family='arial' size = 10)
                                    labelALIGN =left labelATTRS =
(weight=bold);
                        endlayout;

                        /*--Fourth column showing PCIGroup--*/
                        /*       layout overlay /
x2axisopts=(display=(tickvalues) offsetmin=0.25 offsetmax=0.25) */
                        /*          yaxisopts=(reverse=true
display=none) walldisplay=none;*/
                        /*         referenceline y=ref /
lineattrs=(thickness=_thk color=_color);*/
                        /*         axistable y="Baseline Exposure
Variable"n value=PCIGroup / display=(label) labelposition=max;*/
                        /*       endlayout;*/
                        /*--Fifth column showing Group--*/
                        /*       layout overlay /
x2axisopts=(display=(tickvalues) offsetmin=0.25 offsetmax=0.25) */
                        /*          yaxisopts=(reverse=true
display=none) walldisplay=none;*/
                        /*         referenceline y=ref /
lineattrs=(thickness=_thk color=_color);*/
                        /*         axistable y="Baseline Exposure
Variable"n value=group / display=(label) labelposition=max;*/
                        /*       endlayout;*/
                  endlayout;

                  /*              entryfootnote halign=left
textattrs=(size=7) */
```

```
                /*                    'The p-value is from the test
statistic for testing the interaction between the '*/
                /*                    'treatment and any "Baseline
Exposure Variable"n variable';*/
                /*     entryfootnote halign=left 'This graph uses the
new AXISTABLE plot to display the textual columns';*/
           endgraph;
      end;
run;/*--Need format to show missing as blank--*/

proc format;
  value misblank
    . = ' ';
run;

/*----Create Graph-----*/
ods listing style=htmlblue gpath=&graphs image_dpi=&dpi sge=on;
ods graphics / reset noscale width=&w height=&h
imagename='GTL_ForestPlot';
proc sgrender data=Forest2 template=Forest;
format "p-value"n misblank7.2;
dynamic _color='white' _thk=16 ;
run;

ods listing sge=off;
```

Model validation for shared frailty model

```
%macro coxtvc1(data = , y = , x = , tvvar = , nontvvar = , covs = ,
ests = ,
    modopts = , procopts = , addstmts = , out = SurvEsts);


*********************************************************************
****
    * Fit the model, if applicable;
    %if (&ESTS = ) %then %do;
        *Run model and output estimates;
        proc phreg data = &DATA &PROCOPTS outest = _ests_;
            &ADDSTMTS.;
            model &Y = &X / &MODOPTS.;

            *Add time-varying variables in case not previously
defined;
            %VARDEFN;
      id participant_id;
      random participant_id;
            run;
```

```
        %local ESTS;
        %let ESTS = _ests_;
        %end;



*********************************************************************
****
    * Create COVS dataset
    *   1. Variables of type (B) and (C) will have the values
specified by the
    *       user in COVS, if applicable.
    *   2. If not specified, variables of type (B) and (C) will take
on their
    *       mean or reference value, if applicable. This is done
using
    *       PHREG to get the same effect as BASELINE statement. To
get
    *       correct averages, we use one-record-per-patient. In
addition, as
    *       this procedure is only being used to obtain averages, the
time-
    *       varying variables are not needed in the model at this
point. They
    *       will be included in the actual modeling. ;

    *Determine minimum value of start time;
    proc means data = &DATA noprint;
        var %scan(&Y, 1, "()*,");
        output out = _means_ min = min;
        run;

    %local TMMIN;
    data _null_;
        set _means_;
        call symput("TMMIN", trim(left(min)));
        run;

    *Obtain correct value for time-varying covariates when not
specified;
    proc phreg data = &DATA noprint;
        *only considering one-record-per-patient ensures correct
averages;
        where %scan(&Y, 1, "()*,") = &TMMIN;
        &ADDSTMTS.;
        model &Y = &TVVAR &NONTVVAR;

        *obtain correct values;
                id participant_id;
      random participant_id;
        baseline out = _means_;
```

```
        run;

    *Remove death-time from dataset and keep only one record;
    data _means_;
        set _means_;
        keep &TVVAR &NONTVVAR;
        if (_n_ = 1);
        run;

    *If COVS pre-specified, set all missing variables to default
values
    *  obtained from PHREG above;
    %if (&COVS ^= ) %then %do;
        data &COVS.;
            if(_n_ = 1) then set _means_;
            set &COVS;
            run;
        %end;

    *If COVS not pre-specified, set all variables to default value;
    %if (&COVS = ) %then %do;
        %let COVS = _means_;
        %end;

    *Add time-varying variables to COVS dataset;
    %local I NTVVAR;
    data _temp_covs_;
        %let I = 1;
        %do %while(%scan(&X, &I, %str( )) ^= );
            %scan(&X, &I, %str( )) = 0;
            %let I = %eval(&I + 1);
            %end;

        %let I = 1;
        %do %while(%scan(&TVVAR, &I, %str( )) ^= );
            %scan(&TVVAR, &I, %str( )) = 0;
            %let NTVVAR = %eval(&I);
            %let I = %eval(&I + 1);
            %end;
        run;

    data &COVS;
        *Set only variables accounting for time-varying coefficients;
        if (_n_ = 1) then set _temp_covs_(drop = &TVVAR &NONTVVAR);
        set &COVS;
        run;



**********************************************************************
****
```

```
    * Create dataset that contains the unique records of TVVAR
variables;

    *Obtain unique records;
    proc sort data = &COVS (keep = &TVVAR) out = _unique_ nodupkey;
        by &TVVAR;
        run;

    *Determine the number of unique records;
    %local NUNIQUE;
    data _null_;
        set _unique_;
        call symput("NUNIQUE", trim(left(_n_)));
        run;

    *Sort COVS dataset for later use;
    proc sort data = &COVS out = &COVS;
        by &TVVAR;
        run;



**********************************************************************
****
    * Iterate through unique records;

    *Define macros to hold value of interest;
    %do I = 1 %to &NTVVAR;
        %local XTV&I;
        %end;

    %local J;
    %do I = 1 %to &NUNIQUE;
        *Assign to macro variables the unique values for this record;
        data _null_;
            _point_ = &I;
            *grab only record of interest;
            set _unique_ point = _point_;

            %do J = 1 %to &NTVVAR;
                call symput("XTV&J", trim(left(%scan(&TVVAR, &J, %str(
))));
                %end;
            stop;
            run;

        *Restructure time-varying variables to perform integration;
        data &DATA._temp;
            set &DATA.;

            *Reparameterize variables with time-varying coefficients;
```

```
        %do J = 1 %to &NTVVAR;
            %scan(&TVVAR, &J, %str( )) =
              %scan(&TVVAR, &J, %str( )) - &&XTV&J;
            %end;

        %VARDEFN;
        run;

    *Put other variables back to normal;
    data &DATA._temp;
        set &DATA._temp;
        set &DATA.(keep = &TVVAR);
        run;

    *Get only relevant records from COVS dataset;
    data &COVS._temp;
        set &COVS.;
        by &TVVAR;

        retain _point_ 0;
        _any_first_ = 0;
        %do J = 1 %to &NTVVAR;
            _any_first_ =
              max(_any_first_, first.%scan(&TVVAR, &J, %str( )));
            %end;

        if (_any_first_ = 1) then _point_ = _point_ + 1;

        *keep only relevant records;
        if(_point_ = &I) then output;

        *only keep variables in the model;
        keep &X;
        run;

    *Get survival estimates;
    proc phreg data = &DATA._temp inest = &ESTS. noprint;
        &ADDSTMTS.;
        model &Y. = &X / maxiter = 0;
  id participant_id;
random participant_id;
        baseline out = &OUT._temp covariates = &COVS._temp
          survival = Shat / method = emp;
        run;

    *Update output dataset;
    %if (&I = 1) %then %do;
        data &OUT;
            set &OUT._temp;
            run;
        %end;
```

```
        %if (&I ^= 1) %then %do;
            data &OUT;
                set &OUT
                    &OUT._temp;
                run;
            %end;
        %end;

    %exit:
    %mend coxtvc1;



/**Creating counting process data */
%let SURV= cox_stack2;
%Let time = tstop;
%cpdata(data = &SURV., time = tstop, event = Hepa(0), outdata =
SURV2);
/*proc print data = SURV2; run;*/



data covs;
     age    = 34.018711019;
     bil    = 8.70;
     gender2 =0;
     cd4cat1 =0;
     alcohol =0;
     hbv1=0;
     res1=0;
     gr = 3;
run;


%macro vardefn;

     if tstop1 <= &mon. then
          groupy11 = groupy1;
     else groupy11 = 0;

     if tstop1 > &mon. then
          groupy12 = groupy1;
     else groupy12 = 0;

     if tstop1 <= &mon. then
          groupy21 = groupy2;
     else groupy21 = 0;

     if tstop1 > &mon. then
          groupy22 = groupy2;
     else groupy22 = 0;
```

```
%mend;



%coxtvc1(data = SURV2,
    y = (tstop0, tstop1)*hepa(0),
    x =
   groupy11 groupy12 groupy21 groupy22
/*gr1_lt5 gr1_ge5 gr2_lt5 gr2_ge5*/

   age gender2 alcohol cd4cat1 bil hbv1 res1,
    tvvar = gr ,
    nontvvar = age gender2 alcohol cd4cat1 bil hbv1 res1 ,
    covs = covs,
    addstmts = %str(class participant_id gr gender2 Alcohol cd4cat1
hbv1 res1;)
);

data mss.a;

set survests;
run;
/*%let mon = months ;*/
ods output ProductLimitEstimates=est;
proc lifetest data=mss.ag_stack plots=(s lls)    ;
time tstop*hepa(0);/* the zero is a censor*/
symbol v=none;
/*strata Group ;*/
/*where base_res = 0 ;*/
run;

 data failtemp; set est;
  lagsurv=lag(failure);
  do i=1 to 100; if failure eq . then failure=lagsurv;
lagsurv=lag(failure); end;
  if censor eq 0 then failure2=failure;
  surv=1-lagsurv;
 run;

data Failtemp1;
set Failtemp;
format monthz 10.1;
Monthz=tstop;
run;

proc sort data= Failtemp1 out= Failtemp2 (Keep=surv Monthz) nodupkey;
by descending surv Monthz ;
run;
```

```
data Failtemp3;
set Failtemp2;

Rename monthz= tstop1 surv=shat;
run;

Data Mss.Surv_Frailty;
set
survests (in=a)
Failtemp3 (in=b)
;
format Model $30.;

If a then Model = "Shared Frailty Model";
If b then Model = "Observed";


run;
Proc sort data= Mss.Surv_Frailty;
by Model tstop1;
run;

title "Kaplan Meier and Predicted surival plots";
axis1 order = (0 to 24 by 3) minor = none;
axis2 order = (0 to 1 by 0.2) minor = none label = ( a=90);
symbol1 i=stepjoin color=red v=none width=1 line=1;
symbol2 i=stepjoin color=blue  width=1 v=none line=2;
legend label=none /*value=('Early arm' 'Late arm' 'Sequential arm')*/
MODE = PROTECT position=(Top right inside);


proc gplot data= Mss.Surv_Frailty;
      plot (Shat)*tstop1 = Model / overlay haxis=axis1 vaxis= axis2
legend=legend1
      ;
      label Shat = "Probability of remaining hepatotoxicty free";
      label tstop1 = "Months after randomization";
run;
```

References

Abdool Karim, S. S., Naidoo, K., Grobler, A., Padayatchi, N., Baxter, C., Gray, A., et al. (2010). Timing of initiation of antiretroviral drugs during tuberculosis therapy. *New England Journal of Medicine*, 362(8), 697-706.

Aithal, G. P., Watkins, P. B., Andrade, R. J., Larrey, D., Molokhia, M., Takikawa, H., et al. (2011). Case definition and phenotype standardization in drug-induced liver injury. *Clinical Pharmacology & Therapeutics*, 806-815.

Allison, P. D. (1995). *Survival analysis using SAS: a practical guide.* SAS Institute. Inc., Cary, NC.

Allison, P. D., Hancock, G. R., & Mueller, R. O. (2010). The reviewer's guide to quantitative methods in the social sciences: by Gregorgy R. Hancock and Raplh O. Mueller. Routledge. Chap. Survival Analysis. New York.

Altman, D. G., De Stavola, B. L., Love, S. B., & Stepniewska, K. A. (1995). Review of survival analyses published in cancer journals. *British Journal of Cancer,* 72(2), 511.

Amorim, L., & Cai, J. (2014). Modelling recurrent events: A tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 10.1093/ije/dyu222.

Andersen, P. K., & Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, Vol 10 No. 4, 1100-1120.

AVERT. (2019, January 2019). *Global information and education on HIV and AIDS.* Retrieved from HIV AND AIDS IN SOUTH AFRICA: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/south-africa

Babb, D. A., Pemba, L., Seatlanyane, P., Charalambous, S., Churchyard, G. J., & Grant, A. D. (2007). Use of traditional medicine by HIV-infected individuals in South Africa in the era of antiretroviral therapy. *Psychology, health & medicine, 12(3),* 314-320.

Becker, S. (2004). Liver toxicity in epidemiological cohorts. *Clinical Infectious Diseases* , 38 (Supplement 2), S49-S55.

Binquet, C., Abrahamowicz, M., Astruc, K., Faivre, J., Bonithon-Kopp, C., & Quantin, C. (2009). Flexible statistical models provided new insights into the role of quantitative prognostic factors for mortality in gastric cancer. *Journal of clinical epidemiology*, 62(3), 232-240.

Blanc, F. X., Sok, T., Laureillard, D., Borand, L., Rekacewicz, C., Nerrienet E, et al. (2011). Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis. *The New England Journal*, 365:1471–1481.

Boston University School of Public Health. (2016, June). *Boston University School of Public Health*. Retrieved from MPH Online Learning Modules/Measures of Disease Frequency: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/QuantCore/PH717_Frequency/PH717_Frequency_print.html

Boston University School of Public Health. (2018, March 19). *Master of Public health - Modules*. Retrieved from Measure of Association: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_Association/EP713_Association5.html

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368), 829-836.

Collet, D. (2003). *Modelling Survival Data in Medical Research, Second Edition (Chapman & Hall/CRC Texts in Statistical Science).* Boca Raton London New York Washington D.C.: Chapman & Hall/CRC.

Collett, D. (1994). *Modelling survival data. In Modelling Survival Data in Medical Research .* US: Springer.

Cox, D. R., & Snell, J. E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 248-275.

David, C. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34, 187-220.

de Lima, M. d., & de Melo, H. R. (2012). Hepatotoxicity Induced by Antituberculosis Drugs Among Patients Coinfected With HIV and Tuberculosis. *Cad Saude Publica*, 28(4):698-708.

Degu, W. A., Lindquist, L., Aderaye, G., Aklillu, E., Wold, A. H., Ali, G. Y., et al. (2012). Randomized Clinical Trial to Determine Efficacy and Safety of ART 1 Week after TB Therapy in Patients with CD4 Counts <200 Cells/µL. *PLoS ONE*, doi:10.1371/journal.pone.0122587.

den Brinker, M., Wit, F. W., Wertheim-van Dillen, P. M., Jurriaans, S., Weel, J., van Leeuwen, R., et al. (2000). Hepatitis B and C virus co-infection and the risk for hepatotoxicity of highly active antiretroviral therapy in HIV-1 infection. *Aids*, 14(18), 2895-2902.

den Brinker, M., Wit, F., Wertheim-van Dillen, P., Jurriaans, S., Weel, J., van Leeuwen, R., et al. (2000). Hepatitis B and C virus co-infection and the risk for hepatotoxicity of highly active antiretroviral therapy in HIV-1 infection. *AIDS (London, England)*, 14(18):2895-2902.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American statistical Association*, 72(359), 557-565.

Gisolf, E., Dreezen, C., Danner, S., Weel, J., & Weverling, G. (2000). Risk Factors for Hepatotoxicity in HIV-1—Infected Patients Receiving Ritonavir and Saquinavir with or without Stavudine. . *Clinical infectious diseases, 31(5)*, 1234-1239.

Gooley, T. A., Leisenring, W., Crowley, J., & Storer, B. E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in medicine,*, 18(6), 695-706.

Havlir, D. V., Kendall, M. A., Ive, P., Kumwenda, J., Swindells, S., Qasba, S. S., et al. (2011). Timing of ntiretroviral therapy for HIV-1 infection and tuberculosis. *The New England Journal of*, 365:1482–1491.

Hoffmann, C. J., Charalambous, S., Thio, C. L., Martin, D. J., Pemba, L., Fielding, K. L., et al. (2007). Hepatotoxicity in an African antiretroviral therapy cohort: the effect of tuberculosis and hepatitis B. *Aids*, 21(10), 1301-1308.

Hosmer Jr, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: regression modeling of time-to-event data.* Wiley-Interscience.

Islam, M. A. (1994). Multistate Survival Models for Transitions and Reverse Transitions: An Application toContraceptive Use Data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 441-455.

Jewell, N. P. (2003). Statistics for epidemiology. In *Statistics for epidemiology.* Chapman and Hall/CRC.

Kalbfleisch, J. D., & Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 267-278.

Kay, R. (1982). The analysis of transition times in multistate stochastic processes using proportional hazard regression models. *Communications in Statistics - Theory and Methods*, 11:15, 1743-1756.

Klein, J. P., & Moeschberger, M. L. (1997). *Survival analysis: techniques for censored and truncated data.*

Kleinbaum, D. G., & Klein, M. (2005). *Survival Analysis: A Self-learning Text.* Springer.

Kwara, A., Flanigan, T. P., & Carter, E. J. (2005). Highly active antiretroviral therapy (HAART) in adults with tuberculosis: current status. *The International Journal of Tuberculosis and Lung Disease*, 248-257.

Labarga, P., Soriano, V., Vispo, M. E., Pinilla, J., Martín-Carbonero, L., Castellares, C., et al. (2007). Hepatotoxicity of antiretroviral drugs is reduced after successful treatment of chronic hepatitis C in HIV-infected patients. *Journal of Infectious Diseases, 196(5),* 670-676.

Lin, D. Y., & Wei, L. J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association*, 84(408), 1074-1078.

MacMahon, B., & Pugh, T. F. (1970). *Epidemiology: principles and methods.*

Madadizadeh, F., Ghanbarnejad, A., Ghavami, V., Bandamiri, M. Z., & Mohammadianpanah, M. (2017). Applying additive hazards models for analyzing survival in patients with colorectal cancer in Fars Province, Southern Iran. *Asian Pacific journal of cancer prevention: APJCP*, 18(4), 1077.

Mantel, N., & Haenszel, W. (1959, April). Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. *JNCI: Journal of the National Cancer Institute, 22*(4), 719–748. Retrieved from https://doi.org/10.1093/jnci/22.4.719

Marshall, A., Altman, D. G., Royston, P., & Holder, R. L. (2010). Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC medical research methodology*, 10(1), 7.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research, 18(2),* 195-222.

Mills, E., Cooper, C., Seely, D., & Kanfer, I. (2005). African herbal medicines in the treatment of HIV: Hypoxis and Sutherlandia. An overview of evidence and pharmacology. *Nutrition journal*, 4(1), 19.

Nagu, T. J., Kanyangarara, M., Hawkins, C., Hertmark, E., Chalamila, G., Spiegelman, D., et al. (2012). Elevated alanine aminotransferase in antiretroviral-naïve HIV-infected African patients: magnitude and risk factors. . *HIV medicine*, 13(9), 541–548. .

Naidoo, A., Naidoo, K., Yende-Zuma, N., Gengiah, T. N., Padayatchi, N., Gray, A. L., et al. (2014). Changes to antiretroviral drug regimens during integrated TB-HIV treatment: results of the SAPiT trial. *Antiviral therapy*, 19(2), 161.

Naidoo, K., Hassan-Moosa, R., Mlotshwa, P., Yende-Zuma, N., Govender, D., Padayatchi, N., & Abdool-Karim, S. S. (2019). High Rates of Drug-induced Liver Injury in People Living With HIV Coinfected With Tuberculosis (TB) Irrespective of Antiretroviral Therapy Timing During Antituberculosis Treatment: Results From the Starting Antiretroviral Therapy at Three Points in TB Treatment. *Clinical Infectious Diseases*.

Naidoo, K., Yende-Zuma, N., Padayatchi, N., Naidoo, K., Jithoo, N., Nair, G., et al. (2012). The immune reconstitution inflammatory syndrome after antiretroviral therapy initiation in patients with tuberculosis: findings from the SAPiT trial. *Annals of internal medicine*, 157(5), 313-324.

Naidoo, S., Evans, D., Jong, E., Mellet, K., & Berhanu, R. (2015). Outcomes of TB/HIV co-infected patients presenting with antituberculosis drug-induced liver injury. *South African Medical Journal*, 393-396.

Olawumi, H., Olanrewaju, D., Shittu, A., Durotoye, I., Akande, A., & Nyamngee, A. (2014). Effect of Hepatitis-B Virus Co-Infection on CD4 Cell Count and Liver Function of HIV Infected Patients. *Ghana medical journal*, 96-100.

Ozga, A. K., Kieser, M., & Rauch, G. (2018). A systematic comparison of recurrent event models for application to composite endpoints. . *BMC medical research methodology*, 18(1), 2.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V., & Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 541-554.

Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, Pages 373–379.

Pukenyte, E., Lescure, F. X., Rey, D., Rabaud, C., Hoen, B., Chavanet, P., et al. (2007). Incidence of and risk factors for severe liver toxicity in HIV-infected patients on anti-tuberculosis treatment. *The International Journal of Tuberculosis and Lung Disease, 11(1),* 74-84.

Reyes, E., & Thomas, L. (2014). Tutorial: Survival Estimation for Cox Regression Models with Time-Varying Coefficients using SAS and R. *Journal of Statistical Software*.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). Modern epidemiology (Vol. 3). In K. J. Rothman, S. Greenland, & T. L. Lash, *Modern epidemiology* (pp. 33-48). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Sagara, I., Giorgi, R., Doumbo, O. K., Piarroux, R., & Gaudart, J. (2014). Modelling recurrent events: comparison of statistical models with continuous and discontinuous risk intervals on recurrent malaria episodes data. *Malaria Journal*.

SAS Institute Inc., Cary, NC, USA. (2019).

Shu, C. C., Lee, C. H., Lee, M. C., Wang, J. Y., Yu, C. J., & Lee, L. N. (2013). Hepatotoxicity due to first-line anti-tuberculosis drugs: a five-year experience in a Taiwan medical centre. *The International Journal of Tuberculosis and Lung Disease,* 17(7), 934-939.

Sulkowski, M. S., Thomas, D. L., Chaisson, R. E., & Moore, R. D. (2000). Hepatotoxicity associated with antiretroviral therapy in adults infected with human immunodeficiency virus and the role of hepatitis C or B virus infection. *JAMA: the journal of the American Medical Association*, 283(1), 74-80.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: extending the Cox model.* Springer Science & Business Media.

Therneau, T., & Grambsch, P. (2000). *Multiple Events per Subject. In: Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health.* New York, NY: Springer.

Tostmann, A., Boeree, M. J., Aarnoutse, R. E., De Lange, W., Van Der Ven, A. J., & Dekhuijzen, R. (2008). Antituberculosis drug-induced hepatotoxicity: Concise up-to-date review. *Journal of gastroenterology and hepatology*, 23(2), 192-202.

U.S. Department of Health and Human Services. (2019, February). *AIDSINFO*. Retrieved from Side Effects of HIV Medicines. Fact Sheets. HIV and Hepatotoxicity: https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/22/67/hiv-and-hepatotoxicity

U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Allergy and Infectious Diseases, Division of AIDS. (2004, December). *Division of AIDS (DAIDS) Table for Grading the Severity of Adult and Pediatric Adverse Events, Corrected Version 2.1*.

Ullah, S., Gabbett, T., & Finch, C. (2012). Statistical modelling for recurrent events: An application to sports injuries. *British journal of sports medicine*, 48. 10.1136/bjsports-2011-090803.

Vandormael, A., Dobra, A., Bärnighausen, T., de Oliveira, T., & Tanser, F. (2017). Incidence rate estimation, periodic testing and the limitations of the mid-point imputation approach. . *International journal of epidemiology*, 47(1), 236-245.

Wilson, M. G. (2013). Assessing model adequacy in proportional hazards regression. *Paper presented at the Statistics and Data Analysis, SAS Global Forum.* Indianapolis IN, USA.

World Health Organization. (2018, September 18). *Home/Newsroom/Fact sheets/Detail/Tuberculosis*. Retrieved from World Health Organization: https://www.who.int/en/news-room/fact-sheets/detail/tuberculosis

Yashin, A. I., Vaupel, J. W., & Iachine, I. A. (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, 5:2, 145-159.

Yimer, G., Gry, M., Amogne, W., Makonnen, E., Habtewold, A., Petros, Z., & Aklillu, E. (2014). Evaluation of patterns of liver toxicity in patients on antiretroviral and anti-tuberculosis drugs: a prospective four arm observational study in Ethiopian patients. *PLoS One*, 9 (4).