MODELING LENGTH OF HOSPITAL STAY FOR TUBERCULOSIS TREATED IN-PATIENTS AT QUEEN ELIZABETH CENTRAL HOSPITAL: A COMPETING RISK PERSPECTIVE

MASTER OF SCIENCE (BIOSTATISTICS) THESIS

HALIMA SUMAYYA TWABI

UNIVERSITY OF MALAWI

CHANCELLOR COLLEGE

FEBRUARY, 2016

MODELING LENGTH OF HOSPITAL STAY FOR TUBERCULOSIS TREATED IN-PATIENTS AT QUEEN ELIZABETH CENTRAL HOSPITAL: A COMPETING RISK PERSPECTIVE

MASTER OF SCIENCE (BIOSTATISTICS) THESIS

By

HALIMA SUMAYYA TWABI
Bachelor of Science- University of Malawi

Thesis submitted to the Department of Mathematical Sciences, Faculty of Science, in Partial fulfilment of the requirements for the degree of Master of Science (Biostatistics)

UNIVERSITY OF MALAWI
CHANCELLOR COLLEGE

FEBRUARY, 2016

DECLARATION

I the undersigned hereby declare that this thesis/dissertation is my own original work which has not been submitted to any other institution for similar purposes. Where other people's work has been used acknowledgements have been made.

HALIMA SUMAYYA TWABI

Full Legal Name

Signature

Date

CERTIFICATE OF APPROVAL

The undersigned certify that this thesis represents the student's own work and effort and has been submitted with our approval.

Signature:_____Date:_____

Mavuto Mukaka, PhD (Senior Lecturer)

Main Supervisor

Signature: _____Date:_____

Jimmy Namangale, PhD (Associate Professor)

Co-supervisor

Signature: _____Date:_____

Tsirizani Kaombe, MSc (Lecturer)

Programme Coordinator

DEDICATION

To my late Aunt Zahra, late Grandfather Imran and late Grandma Apwaja.

ACKNOWLEDGMENTS

I would like to thank whole-heartedly my supervisor Dr Mavuto Mukaka, for his guidance, support, constructive ideas, comments and his time and knowledge on my thesis work. I really appreciate the encouragement and his critiques on my work which has helped me structure the entire thesis with understanding.

I would also like to thank Dr Danielle Cohen for her constructive comments on the clinical aspects of the research. Without her help, this work might have been incomplete and would not have been approved by the research ethics committee. I appreciate her assistance

I am also indebted to Dr Ingrid Peterson and the entire SPINE team for providing data used in the research. Her quick response to my misunderstandings and guidance on SPINE data was marvelous.

I would also like to thank Dr Namangale, Mr. Tsirizani Kaombe for their assistance when I needed help in various aspects of the thesis work. Their comments on my thesis have helped me understand some aspects that were challenging to grasp.

I sincerely thank God for guiding me through this work Special thanks to mum, dad, grandmother and my loving husband Shaffi for your support throughout my work. Thank you for giving me strength and encouragement.

ABSTRACT

A retrospective cohort study was done on adult TB in-patients database from Queen Elizabeth Central Hospital (QECH) SPINE database to identify factors explaining time to discharge from hospital while accounting for a competing event: death. The study aimed to apply and compare competing risk models on TB data. Semi-parametric Cause-specific hazards (CSH) and Sub-distribution hazard (SDH) models were applied to model the effect of HIV status, age, and Sex in relation to death or discharge from hospital. Test for model assumptions and diagnostics were conducted. Findings showed that the SDH explained best the effect of the covariates to the probability of a patient being discharged or dying. Further the main factors affecting length of hospital stay among TB in-patients were age and HIV Status. HIV positive patients were 17.6% less likely to be discharged from hospital compared to HIV negative patients ($p=0.048$) and an increase in age, resulted in 2% decrease of chances of discharge. It is important to use the cumulative incidence function for calculating probability of an event. The SDH model was a better model when studying data that involves competing risks. To meet the objective of identifying prognostic factors of discharge in the presence of competing risks, the sub-distribution hazard model explained better the covariate effects on event discharge than the CSH model. The findings emphasize the importance to use competing methods which best meet the study objectives.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDECIES

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike's Information Criteria
AIDS	Acquired Immune Deficiency Syndrome
ART	Antiretroviral Therapy
BIC	Bayesian Information Criteria
CI	Cumulative Incidence
CIF	Cumulative Incidence Function
COMREC	College of Medicine Research Ethics Committee
CR	Competing Risks
CSH	Cause-Specific Hazards
CV	Cardiovascular
DOTS	Directly Observed Therapy-Short course
EPTB	Extra- Pulmonary Tuberculosis
ESRD	End-Stage Renal Disease
GDP	Gross Domestic Product
HD	Hemo-dialysis
HIV	Human Immuno-deficiency Virus
KM	Kaplan Meier
KMC	Kaplan Meier Complement
LOS	Length of Hospital Stay
PD	Peritoneal Dialysis
PH	Proportional Hazards

PTB	Pulmonary Tuberculosis
QECH	Queen Elizabeth Central Hospital
SDH	Sub-distribution Hazard
SPINE	Surveillance Programme of In-patients and Epidemiology
TB	Tuberculosis
UN	United Nations
WHO	World Health Organisation

CHAPTER ONE

INTRODUCTION

1.1 Background Information

Tuberculosis remains a major public health problem worldwide. It is estimated that one-third of the world population is infected with Mycobacterium tuberculosis (WHO., 2010).The severity of tuberculosis in the world has worsened with social inequality, the advent of acquired immunodeficiency syndrome (AIDS) and migratory movements between countries. Thus, it's still a public health challenge in most countries of the world (WHO., 2012). In 2013, WHO reported that 9 million people around the world were sick of TB and there were around 1.5 million TB-related deaths worldwide.

Globally interventions and measures such as the Directly Observed Therapy-Short course (DOTS), and International Stop TB strategy aimed at eliminating TB have been implemented. These efforts have led to a successful reduction in TB cases world-wide. Globally, a total of 56 million people were successfully treated and as a result, TB incidence has fallen by 2% each year. Although this is the case, the global burden of TB remains high especially in most developing countries (WHO., 2013).

1.1.1 Case of TB in Malawi

In Malawi TB is a major public health problem with the incidence of all forms of TB being estimated to be 164 per 100,000, as reported by in 2012. The report also estimated that in Malawi there were an estimated 29,000 new cases of TB (all forms) in the year 2011, and approximately 18,000 of these were HIV positive. A long term study by WHO shows that TB funding in low and middle income countries grew from 2002 to 2011. Despite the increment in funding, it is still inadequate in comparison to the magnitude of the problem. The majority of countries that have a heavy TB burden are classified as low income countries (GDP below 760 US dollars) (WHO., 2013). Malawi is among the 10 poorest countries in the world (UN Development Report) and has currently been ranked as the first poorest country in the world by data from the World Bank in 2015.

In Malawi TB has had a great impact on the socio-economic well-being of the country. It is reported that on average, patients spend 29 US dollars to access facilities offering diagnostic and treatment services for TB (Kemp. et al., 2007). Although this is the case, (WHO., 2013) reports that the cost per person successfully treated for TB with first line drugs is in the range of 100 USD to 500 USD in all countries with high burden of TB. In view of these high costs, there is a need to understand different aspects surrounding care for TB patients and this includes studying factors related to hospitalization.

Since TB-infected patients who are admitted to the hospitals tend to have more serious clinical conditions, the determination of their treatment outcomes carries a great clinical and public health importance (Zetola. et al., 2014). Equally important, is an analysis of

the patients' length of stay (LOS) in hospital as this can guide future resource allocation for the treatment of such patients.

1.1.2 Length of Hospital Stay

Determination of factors that increase LOS may provide information that can help to reduce costs and improve delivery of care (Collins. et al., 1999).Most studies on length of hospital stay have shown that LOS is an important measure of resource utilization (Frietas. et al., 2012) and it can partly explain hospital costs as some studies have shown that there is a strong correlation between LOS and hospital costs. Thus, understanding length of stay is vital for planning and funding services (Hinchliffe. et al., 2013).

Understanding LOS for Malawi, a developing country which provides free primary, secondary and tertiary health care to its citizens, would be very helpful since funding health services is costly and thus there is a need to understand ways in which the cost for hospital services could be managed better. Also understanding LOS would help in planning for the hospital services that are provided to patients, for example in terms of bed occupancy. It would provide the hospital an overview of how the TB wards are operating in terms of space and quantity of medical items used.

In Malawi, patients are diagnosed for Tuberculosis for free and are mostly treated as out-patients. This developed because before 2001, TB wards were congested with admitted TB patients on treatment. In urban Malawi, the bed occupancy rates were between 140 to 160%. These rate have gone down since 2002 when the national policy changed to giving

patients options of receiving initial phase of treatment from hospital wards or health centers or to have it provided by guardians at their homes (Nyirenda. et al., 2003).TB patients are only admitted to hospital care when either their clinical condition warrants it and / or access to community-based care is not available. It is equally important that TB patients be discharged for outpatient care at clinics as soon as they can be managed effectively in the community (Tamiru & Haidar., 2010).

1.2 Problem Statement

There is large body of literature on competing risk models for analysis of time-to-event data in medical research (Dignam. et al., 2012; Hinchliffe. et al., 2013; Kim., 2007; Lim. et al., 2010). Studies in the past have employed different statistical techniques such as Cox regression model, Logistic regression and Chi-square test to study Length of hospital stay. A few papers have appeared in the application of advanced statistical models on LOS such as the generalised linear mixed model (GLMM). For example a hierarchical Poisson regression model for maternity LOS (Lee. et al., 2001) was developed to capture the inherent correlations of patients clustered within hospitals. A finite mixture regression model with random effects and its application to neonatal hospital LOS has been proposed by(Yau. et al., 2003), leading to the development of the class of finite mixture GLMM where heterogeneity in LOS has been modeled. Despite these studies and papers on LOS, a few studies have looked at modeling length of hospital stay for TB patients and have often not accounted for competing events.

In survival analysis an individual who experiences an event of interest within a specified observation period is said to have an event, otherwise the individual is said to be censored if no such event is experienced in that period by the end of the study. When more than one event is considered (e.g., death from any of several causes), those events are known as competing risks or competing events(Coviello. et al., 2004; Kleinbaum & Klein, 2005). As Gooley (1999) stated, ignoring competing risks and applying standard survival models to a dataset that includes competing events leads to biased estimates thus leading to biased conclusion. Therefore there is need to account for competing risks where they exist.

In the study of length of stay for TB patients, death while in hospital is one of the well-known competing events since those who die do not have a chance to be discharged even if the observation time was extended. Failure to account for this would lead to invalid estimates of time to discharge. Therefore, this study aimed to estimate time to discharge while accounting for competing risk death.

Hospitalizing TB patients can be challenging especially in countries like Malawi, with limited health care resources or appropriate in-patient facilities (Dehghani. et al., 2011). An analysis of length of in-patient hospital stay and factors affecting hospitalization with an account of competing events is important in assessing and predicting the consumption of hospital resources which is an important tool in hospital planning for resource allocation. As stated by Hinchliffe et al, 2013, understanding length of hospital is important for planning and funding of hospital services. Therefore the study aims at

modelling the LOS for TB treated patients to observe the length of time TB patients remain in hospital and factors that influence the average length of stay of TB patients in Malawi using models that take competing risk into account; and choose the best model that explains the associated factors of length of hospital stay.

1.3 Objectives of Study

The following are the study objectives:

1.3.1 Broad Objective

To apply competing risk models on time to discharge for adult TB in-patients at QECH with death as a competing event.

1.3.2 Specific Objectives

1. To estimate and compare the Cumulative Incidence Function with the Kaplan Meier Estimator
2. To compare the Cumulative Incidence curves in the presence of the competing risk for the categorical variables; HIV Status, ART Status, Gender.
3. To fit and compare the Cause-Specific Hazard and Sub-distribution hazard models.
4. To identify prognostic factors affecting time to discharge in the presence of event death

1.4 Research Questions

The following were the research questions for the study:

1. Do the cumulative incidence curve and the Kaplan Meier curve give different probabilities to discharge?
2. Which model (Sub-distribution hazard or Cause-Specific Hazard models) best explains time to discharge for TB patients?
3. What are the factors that affect time to discharge for TB in-patients at QECH?

1.5 Significance of the Study

Examining length of hospital stay for TB patients will provide an insight into this public health problem and will contribute to the country's base knowledge of factors affecting length of hospitalization of TB patients. In addition, the study contributes to available work done in the field of survival analysis when modeling survival time while accounting for competing events. Furthermore results of the research will help researchers understand appropriate competing risk methods to use when studying length of Hospital stay or epidemiological diseases.

The subsequent chapters present the Literature review of the study area, the methodology used in the study, the results and discussion of the study and conclusion and recommendation(s).

CHAPTER TWO

LITERATURE REVIEW

This section provides literature review on survival analysis, survival function, hazard function, hazard ratio, Kaplan Meier (KM) methods, tests for survival analysis, Models in survival analysis, handling of time-varying covariates and competing risks approach.

2.1 Determinants of Length of Hospital Stay

Different studies have shown that age, HIV Status, ART therapy are some of the risk factors associated with hospital stay. A study done by Ferreira et al (2014) on factors associated with hospitalization of tuberculosis patients showed that increased length of hospital stay was proportional to increasing age, especially > 40 years; male; single; low education; tuberculosis/human immunodeficiency virus (TB/HIV) co-infection; previous TB episode; pulmonary and extra-pulmonary TB; previous opportunistic infection. They used an integrative literature review, using the MEDLINE, LILACS, and ISI databases, besides the SciELO collection, whose descriptors were: “tuberculosis”, “hospital”, “hospitalization”, “risk factors”, and “associated factors. Their study did a comprehensive literature review on factors associated with hospital stay. Despite reviewing different studies that have looked at this area of study, most of the studies focused on whether a patient was discharged or not. As a result methods such as the logistic regression, Chi-Square test of association, were used to determine associated factors of length of stay.

A retrospective study on Factors Associated with Length of Hospital Stay among HIV Positive and HIV Negative Patients with Tuberculosis in Brazil done by Ferreira et al (2013) used a Chi-square test or a T-test at a 5% significance level to obtain the associated factors. The study showed that there were no significant differences in the length of hospital stay in HIV positive patients but found that minimum wages, pulmonary tuberculosis form, negative smear test or no information in this regard, initial 6-month treatment scheme, were associated to prolonged hospital stay in HIV positive patients. Another study in Brazil also concluded that a high number of patients with TB/HIV are expected in hospitals as admission patients (Oliveira. et al., 2009).

Tuberculosis is highly associated with HIV status of a patient. Many retrospective studies have shown that tuberculosis is associated to HIV. A retrospective study done in Lilongwe, Malawi showed that HIV co-infection was associated with a slightly poorer TB treatment outcome. Only 38% of the TB/HIV new smear positive co-infected patients were on ART. Those on ART had successful TB treatment outcomes compared to those not on ART (Tweya. et al., 2013). Information on ART is very important when modeling survival of TB patients after admission, since there is a relationship between ART and TB treatment outcome, thus it needs to be considered when studying factors that affect length of hospitalization for TB patients.

The study by Tweya et al (2013), further found that both HIV and ART status influenced TB treatment outcomes. This explains that length of hospitalization of a TB treated patient is dependent on the HIV or ART status as based on the study. Those with HIV

who are not on ART are likely to have a poor TB outcome (are likely to have a high LOS). Therefore when studying TB patients who are HIV reactive it is very important to consider whether the patients are on ART or not and determine if ART contributes to the length of hospital stay of that patient.

2.2 Common Terms Used in Survival Analysis

2.2.1Censoring

Time to event analyses test hypotheses about the occurrence of an event of interest in two or more groups with data that are often subject to censored observations. Censoring occurs when information on time to outcome event is not available for all study participants. Three reasons of censoring are: when a person is lost to follow-up during the study period, and when a person withdraws from the study because of death (if death is not the event of interest) or some other reason like issues concerning ethics for example having adverse drug reaction. Censoring is of two types, right and left (Leung. et al., 1997). Right censored data is mostly encountered which involves lost to follow up. Left censored data can occur when a person's survival time becomes incomplete on the left side of the follow up period. Censored observations may not only be due to losses to follow-up or administrative cessation of the time period of consideration but can also be due to events not of interest. This situation is problematic if these "other events" preclude observation of the primary event under consideration. Experiencing a competing event acts as a right censor on the primary event. Because of this extra censoring, it is often useful to estimate and compare cumulative event probabilities of a specific event, rather than of all events as a whole.

Censoring in survival analysis should be “non-informative,” i.e. participants who drop out of the study should do so due to reasons unrelated to the study. Informative censoring occurs when participants are lost to follow-up due to reasons related to the study, e.g. in a study comparing disease-free survival after two treatments for cancer, the control arm may be ineffective, leading to more recurrences and patients becoming too sick to follow-up

2.2.2Survival Function

Let T be a non-negative random variable denoting the time to a failure event. The survivor function $S(t)$ gives the probability that a person survives longer than some specified time ***t:*** that is, $S(t)$ gives the probability that the random variable T exceeds the specified time t (Kleinbaum & Klein, 2005). In other words the survivor function also known as survivorship function is simply the reverse of the cumulative probability function of T. Where the cumulative distribution is given by

$$F(t) = \Pr(T < t) = \int_0^t f(u)du \qquad (1)$$

and the survivor function is given by

$$S(t) = 1 - F(t) = \Pr(T \geq t) \qquad (2)$$

It is simply the probability that there is no failure event prior to time t. The function is equal to 1 at t=0 and decreases toward zero as t goes to infinity. Its probability density function is expressed as;

$$f(t) = \frac{dF(t)}{dt} = \frac{d\{1-S(t)\}}{dt} = -S'(t) \qquad (3)$$

2.2.3 Hazard Function

The hazard function also known as the conditional failure rate is the instantaneous rate of failure. It is the limiting probability that the failure event occurs in a given interval, conditional upon the subject having survived to the beginning of that interval, divided by the width of the interval (Cleves. et al., 2010). In simple terms it is the probability that an individual encounters an event of interest at time t, conditional on having survived to that time. If **t** is a continuous function with density function ***f,*** then the hazard function is defined by:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\,(t+\Delta t > T > t | T > t)}{\Delta t} = \frac{f(t)}{S(t)} \qquad (4)$$

It can vary from zero (no risk at all) to infinity (certainty of a failure at that instant). It is different from survival function because it specifies the failure event while the survivor function talks of the survival rate past a time t (Kleinbaum & Klein, 2005). The importance of the hazard function is that it provides insight into conditional failure rates. It may also be used to identify a specific model form.

2.2.4 Hazard Ratio (HR)

In survival analysis the hazard ratio is the ratio of the hazard rates corresponding to the conditions described by two levels of an explanatory variable. The hazard ratios represent instantaneous risk over the study time period. A hazard ratio of 1 corresponds to equals hazards between the two groups (i.e. treatment arm and control arm). While a hazard ratio of 2 implies that at any time twice as many in the treatment group are having an event proportionately compared with the control group (Deurden., 2009).

2.2.5 Kaplan-Meier (KM) Estimate

The Kaplan Meier estimator is a non-parametric estimate of the survivor function S(t), which is the probability of survival past time t, or the probability of failing after t. It is a popular method because it requires very weak assumptions (assumes no form of distribution) but utilizes information content of both fully observed and right censored data. Suppose that k individuals have experienced an event of interest, such as death in a group of individuals. If we let $0 \leq t_1 < \cdots < t_k < \infty$ be the observed ordered death times. Let k_j be the number of individuals who are at risk at $t_{(k)}$. Let d_j be the number of observed deaths at t_j, j= 1…k. Then the Kaplan Meier estimate at any time t is given by

$$\hat{S}(t) = \prod_{j|t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) \qquad (5)$$

where n_j is the number of individuals at risk at time t_j, and the product is overall observed failure times less than or equal to t (Kaplan. & Meier., 1958). The estimator is a step function that changes values only at the time of each.

2.2.6 Cumulative Incidence Function

A competing risk must be accounted for in estimating failure rates. The best approach of assessing failure rates is by using the cumulative incidence curve to estimate the probability of failures actually observed in patients who are subject to censoring by competing risk (Dignam et al., 2012).

The cumulative incidence, which is closely related to the survivor function encountered in standard survival analysis, denotes the expected proportion of patients with a certain event over the course of time (Latouch. et al., 2007). The CIF at time t for cause *i* is the

probability of failing from cause *i* before (or up to) time t, it represents the probability that an event of type *i* has occurred by time t. It is represented as

$$CIF_i(t) = P(T \leq t \text{ and } failure\ from\ cause\ i) = \int_0^t f_i(u)du. \quad (6)$$

The cumulative incidence function helps to determine patterns of failure and to assess the extent to which each component contributes to overall failure.

2.2.7 Cause-Specific Hazard Function

Survival function and Hazard function are important quantities in the analysis of time to event data. The survival function quantifies the probability of a person being event free at a given point in time. While the hazard function quantifies the risk that a person who is event free at a given point in time will experience the event in the next instant. In competing risks, each event has an associated hazard function known as the cause-specific hazards (CSH). A cause specific hazard quantifies the risk of experiencing an event from a particular cause (Aban, 2014).

The cause-specific hazard refers to the instantaneous risk of failure from a specified cause given that no failure from any cause has yet occurred. Formally if failure can occur for any *i* =1, … , k causes. The CSH for cause *i* at time t is given as

$$h_i(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t+\Delta t, D=k | T \geq t)}{\Delta t} \quad (7)$$

T is equal to time to first failure from any cause *i*. A subject will still be at risk at time t given that the subject has not died of cause *i* or any of the *i*-1 other causes. For D ϵ {1,2, … , k}, It represents the hazard of failing from cause *i* in the presence of the competing events.

Emerging evidence now suggests that in the presence of competing risks, which will be further discussed, the cumulative incidence function, a method which takes into account competing risks occurrence, is the appropriate method use to estimate the probability of occurrence of the event of interest in the presence of other events. However, researchers often use the Kaplan Meier approach (1-KM) to evaluate the survival probability of occurrence of a cause-specific endpoint, even if the appropriate data contain competing-risk events (Gooley, 1999).

2.3 Competing Risk Approach

In medical research, each person studied can experience one of several different types of events over the follow-up period and survival times are subject to competing risks if the occurrence of one event type prevents other event types from occurring (Kleinbaum & Klein, 2005). For example, in order to determine the incidence of discharge among Tuberculosis patients, every patient will be followed from a baseline date (such as date of admission) until the date of discharge from hospital. A patient who is discharged during the study period would be considered to have an ‘event’ at their date of discharge. A patient, who is alive at the end of the study but still in hospital, would be considered to be ‘censored’. However, a patient can experience an event different from the event of interest. For example, a TB patient may die due to TB or unrelated causes. Such events are termed competing risk events.

When competing risks are present it is assumed that the subjects contribute independent and identically distributed observations to the data; the component fails when the first of

all the competing failure mechanisms reaches a failure state; each of the *k* failure modes has a known life distribution model. (Pepe, 1991; Crowder, 1994). One can assume that each failure mechanism leading to a particular type of failure proceeds independently of each other, including the risk of the event of interest, at least until a failure occurs. However, this is often not likely to be true, particularly when there is causal-effect between events. To assume independence one must be sure that a failure of one type of event has no effect on the likelihood of any other events (Crowder, 1994).

Competing risks modeling is important in time to studying length of stay because a large proportion of patients may either die or be discharged, where if one dies, the event of interest: discharged would not be observed. Competing risks models offer significant advantages over standard survival analysis when competing events exist (Putter. et al., 2007). Various studies (Gooley et al, 1999; Fine and Gray 1999, Dignam et al, 2012) have proposed the use of the cumulative Incidence Function other than the Kaplan Meier to estimate quantities pertaining to the probability of failure caused by an event of interest when other failure types may preclude it.

2.3.1 Comparison of the Cumulative Incidence Estimate and Kaplan Meier Estimate

A Comparison of the Kaplan Meier estimate to the cumulative incidence curves, shows that the KM estimates probabilities of one failure in the absence of any others while the cumulative incidence curves of each of these causes of failure will sum to the cumulative incidence of any failure (Chappell., 2012). Beuscart et al, (2012) in their study found that the Kaplan-Meier method overestimated the probability of each event, while the

cumulative incidence provided accurate estimations of event probabilities. The study looked at the efficacy of peritoneal dialysis (PD) in survival of patients explained that a patient on PD could experience a transfer to Hemodialysis, Renal transplantation or death which was considered as competing events. They found that the Kaplan-Meier method overestimated the probability of each event, i.e. death, transfer to HD, or renal transplantation during PD. When the event investigated was death, patients censored because of transfer to HD or renal transplantation was considered to be withdrawn alive on PD, which led to an overestimation of the probability of death during PD. When the event studied was transfer to HD or renal transplantation, patients who died were censored and considered to be withdrawn alive on PD (Beuscart. et al., 2012).

Most studies use the complement of KM (1-KM) for comparability sake against the cumulative incidence. The complement of KM is an estimator interpretable only if events due to all other causes are removed. The Kaplan Meier complement (1-KM) of event *i* at time t is defined as the cumulative probability of experiencing event *i* before time t in the absence of competing events. It is defined as;

$$KMC_i(t) = \int_0^t h_i(u)S_i(u)\Delta u \qquad (8)$$

$$= 1 - S_i(t) \qquad (9)$$

$$= 1 - \exp\big(-H_i(t)\big)\, i = 1, \dots, K \qquad (10)$$

Where the event-specific survival function for event *i*, $S_i(t)$, is defined as the probability that $T \geq t$ when $I=i$. $S_i(t)$ can be estimated by the Kaplan Meier estimator.

The probability of experiencing a competing event prior to time *t*is assumed to be zero when this does not actually reflect the true situation under competing events. Thus, the complement of KM cannot be considered the true probability of an event occurring before a certain time *t* because competing events are treated as censored observations (Dignam et al., 2012).

Gooley, et al, (1999) states that the CI gives a more accurate representation of the cumulative event probability than the complement of KM in the presence of competing events, because competing events are included in the risk set. 1-KM is equivalent to the CI in the absence of competing events. 1-KM always overestimates the CI in the presence of competing events because reducing the number of individuals in the risk set inflates the proportion of individuals at risk.

Verduijn et al, (2011) showed that when cumulative survival probabilities for competing events such as Cardiovascular (CV) and non- Cardiovascular (CV) mortality are estimated by the Kaplan–Meier method, these probabilities are profoundly overestimated for each of the two separate causes. This is in particular the case in populations with high mortality, such as in elderly dialysis patients, and/or long duration of follow-up. As a consequence, the sum of the estimated CV and non-CV mortality probabilities is (much) larger than the all-cause mortality probability and may even exceed 100%. For this reason, Kaplan–Meier should not be used to calculate and present cumulative probabilities curves for cause-specific mortality. The study looked at all-cause mortality and cause-specific mortality (CV and non-CV mortality) were analyzed by Kaplan–Meier

analysis and Cumulative Incidence Competing Risk analysis in two cohorts of patients with end-stage renal disease (ESRD) on dialysis.

Contrary to the different findings on CI and I-KM, Borrebach (2013) argues that choosing between the complement of the Kaplan Meier and the cumulative Incidence is ambiguous, since CI has a disadvantage of not removing failures due to competing events from the risk set. In his study, where data were simulated and analysis was done for four scenarios; When there's i) Primary event Hazard, ii) High competing event Hazard, iii) High random censoring, and iv) High sample sizes. The results showed that all except high competing event hazards had a difference in the estimates for the CI and 1-KM.

Borrebach (2013) explains that 1-KM's potential clinical advantages with an example that suppose a woman diagnosed with stage II breast cancer is due to receive a more aggressive treatment if, based on her characteristics her cumulative event probability is predicted to be above a certain threshold. If her cumulative event probability is predicted to be below that value, she will receive a less aggressive treatment. Suppose also that her 1-KM estimate lies above this threshold, whereas her CI estimate lies below this threshold. In this case, her clinician may decide to exercise caution and use 1-KM, giving the more aggressive treatment and presumably having a greater chance of treating her cancer. He further explains that there may be instances where clinicians would want to avoid overtreatment when the treatments (e.g., certain chemotherapies, radiation therapies) have potentially harmful side-effects of their own. In those cases, using the CI may be desired instead. The problem for clinicians becomes whether they want greater

predictive accuracy or to exercise caution in cases where the benefits of over-treatment are perceived to be greater than the risks (Borrebach, 2013).

The studies have clearly shown that using cumulative incidence estimates when dealing with competing events lead to unbiased estimates of the cumulative probabilities unlike using the Kaplan Meier estimate, although the complement of KM might still be clinically advantageous when making decisions.

2.4 Test of Hypothesis

In addition to estimating the survival functions, Kaplan-Meier Estimator in Origin provides three other methods to compare the survival function between two samples. These include; Log Rank, Wilcoxon and Tarone-Ware etc. These tests are very useful in assessing whether a covariate affects survival however they do not account for competing events available in dataset. Therefore two alternative methods: Gray's test and Pepe and Mori test, for comparing cumulative incidence curves for a particular failure type among different groups are presented in this section. This study used the Pepe and Mori test to test for equality of CIF between two groups.

2.4.1 Pepe and Mori's Test

Pepe and Mori's test is a 2-sample test that was introduced by Pepe and Mori (1993). This test compares the cumulative incidence functions (CIF's) directly for the event of interest. The null hypothesis is that there is no difference between the 2 groups.

2.4.2 Gray's Test

Gray's test is a K-sample test that was introduced by Gray (1988). It compares the weighted averages of the sub-distribution hazards across groups for the event of interest. The null hypothesis is that there is no difference among the K groups. The test is based on the K – 1 score statistics

2.5 Models in Survival Data Analysis

This section presents the survival models that are used to estimate the effect of the covariates on the hazard rate of an event. These models suggested in the literature include the Cox semi-parametric proportional hazard model and some parametric models like the exponential model, and Weibull Model and Log-Normal model. The Cox PH and competing risk models were discussed in this section, since they were used in the analysis of the TB in-patient data.

2.5.1 Cox Proportional Hazard Model

It is the most common approach to model covariate effects on survival. It takes into account the effect of censored observations (Cox., 1972). The model is based on the assumption of proportional hazards and no probability distribution assumption is made on the survival times. The only assumption made is on the proportionality of the baseline hazard. The model is therefore referred to as a semi-parametric model. The proportional hazard assumption means that the hazard ratio is constant over time or that the hazard for an individual is proportional to the hazard for any other individual (Therneau. &

Grambsch., 2000). Let $x_{1, \ldots, x_p}$ be the values of p covariates $X_{1, \ldots, X_p}$, according to the Cox regression model, the hazard function is given as follows;

$$h(t) = h_0(t)\exp\left(\sum_{i=1}^{p} \beta_i X_i\right) \qquad (11)$$

Where $\beta_i = (\beta_{1,} \beta_{2,\ldots,}\beta_p)$ is a 1 × p vector of regression coefficients and $h_0(t)$ is the baseline hazard function at time t.

In many applications, competing risks have been ignored (such as, patients experiencing competing events were censored at the time of these events) and standard Cox regression was applied. This approach is adequate when competing risks are rare because it assumes independence between the event of interest and censored observations. However, in the presence of strong competing risks, standard survival models may overestimate the hazard of the event of interest because subjects with a competing (and thus censored) event are treated as if they could experience the event of interest in future (Putter et al, 2007; Wolbers et al, 2009).

2.5.2 Cause Specific Hazard Model

The regression model on cause-specific hazards is as follows:

$$h_i(t|x) = h_{0i}(\mathrm{t})\exp(\boldsymbol{\beta x}) \qquad (12)$$

Where X is a vector of explanatory variables and β is a vector of coefficients. The total risk of any event happening, the overall hazard rate is

$$h(t|x) = \sum_i h_i(t). \qquad (13)$$

The typical “cause-specific” approach for analyzing competing risks data is to perform a survival analysis (standard Cox regression) for each event type separately, where the

other (competing) event types are treated as censored categories. There are two primary drawbacks of the above method. One problem is that the above method requires the assumption that competing risks are independent (Kleinbaum & Klein, 2005) which is not the case when dealing with competing risk data. As previously discussed, to estimate the survival probabilities, the CIF is much appropriate when dealing with competing risk. The Cox-Proportional hazards may be used to model the cause-specific hazards in regression modeling (Aban, 2014). However testing for equality of CSH is not equivalent to testing the equality of CIF(Gray, 1988).

Cause-specific hazard and corresponding hazard ratio's, are estimated using Cox proportional hazards model for each failure event. The comparison of the cause-specific hazards is made as if the other types of events did not exist. Kim (2007) regarded this approach as unrealistic.

Several modeling approaches are available for evaluating effects of covariates on the cause-specific outcome in competing risk data (Fine. & Gray., 1999). Two popular approaches are (1) modeling the cause-specific hazard of each event separately by applying the standard Cox regression for the event of interest and censoring all other observations. The second approach is Fine and Gray's (1999) extension of the Cox regression that models (the hazards) the CIF.

As already mentioned, the cause-specific hazard can be modeled using the Cox model, which is broadly used in medical research. The relationship between the $CIF_i(t)$ and the cause-specific hazard is mathematically represented as;

$$CIF_i(t) = \int_0^t h_i(x)S(x)dx$$

$$= \int_0^t h_i(x)exp\{-\sum_{j=1}^{k} H_j(x)\}dx$$

$$= \int_0^t h_i(x)exp\{\sum_{j=1}^{k} \int_0^x h_j(u)du\}dx \qquad (14)$$

Where S(x) is the overall survivor function, $H_j(x)$ is the cause-specific for cause j, which is integrated from 0 to x of the CSH for cause j.

A study done by Andersen et al (2012) on Competing Risk in Epidemiology Possibilities and Pitfalls deduced that a one to one correspondence between a single rate (cause-specific hazards) and the corresponding risk (cumulative incidence [CI]) no longer exists. This means that any given CI depends on all cause-specific hazards and vice versa. Also another consequence of lack of correspondence is that covariates may affect a cause i specific hazard and cause i CI differently. He suggested that cause-specific hazards may be more relevant when the disease etiology is of interest, since it quantifies the event rate among the ones at risk of developing the event of interest. Though this was their deduction, they concluded that CI's are easier to interpret and are more relevant for the purpose of prediction.

Cause-specific hazards can inform us about the impact of risk factors on rates of disease or mortality, while the cumulative incidence functions provide an absolute measure with which to base prognosis and clinical decisions on (Koller. et al., 2011). Although the CSH's and the CIF are reported separately, Hinchliffe, (2011) did a study that would model competing risks scenarios using an approach that estimates both the cause-specific hazards and the cumulative incidence functions as they believed both to be useful measures. Such an approach was defined by Fine and Gray (1999) and will be explained in the later section.

2.5.3 Sub-Distribution Hazard Model

In recent years, research methods centered on directly assessing covariate effects on a CIF have been developed (Jeong & Fine, 2007). One important work is the proportional sub-distribution hazards model proposed by Fine and Gray (Jeong. & Fine., 2007). This approach directly measures the covariate effects on the cumulative failure probability due to one risk, in the presence of other risks. Fine and Gray (1999) specify a model for the sub distribution hazard formally defined for failure cause *i* as

$$\bar{h}_i(t) = \lim_{\Delta t \to 0} \left\{ \frac{P(t<T \leq t+\Delta t and failure cause i | T>t \text{ or } (T \leq t and not failure cause i))}{\Delta t} \right\} \qquad (15)$$

This hazard generates failure events of interest while keeping subjects who experience competing events "at risk" so that they are counted as not having any chance of failing

As in any other regression analysis, modeling CIF for competing risks can be used to identify potential prognostic factors for a particular event in the presence of competing

risks, or to assess a prognostic factor of interest after adjusting for other potential risk factors in the model.

The cause-specific hazard model may be more clinically understandable when assessing the prognostic effect of the covariates on a specific cause because we see that the covariate effect would be to reduce or increase the instantaneous probability of the event of interest irrespective of other covariate effect. However, when the study objective is to compare the probability of the event of interest, then the sub-distribution hazards model is appropriate (Lim et al, 2010). The sub-distribution model is more desired because it assesses covariate effect on CIF directly unlike cause-specific model which is an indirect measurement. Although this is the case the sub-distribution hazards model might be limited to populations with similar characteristics and similar competing risk rate, the cause-specific hazard model is applicable for any population with similar characteristics regardless of the rates of competing risk events (Pintilie, 2007). The sub distribution hazard model can be used to calculate the CIF from it by the equation;

$$CIF_i(t) = 1 - \exp\{-\bar{H}_i(t)\} \qquad (16)$$

Where $\bar{H}_i(t) = \int_0^t \bar{h}_i(t)dt$ is the *cumulative sub-hazard*. The sub-distribution hazard model is semi-parametric in that the baseline subhazard $\bar{h}_{1,0}(t)$ (covariates set at zero) is left unspecified, while the effects of the covariates **x** are assumed to be proportional;

$$\bar{h}_i(t|\boldsymbol{x}) = \bar{h}_{1,0}(t)\exp(\boldsymbol{x}\boldsymbol{\beta}) \qquad (17)$$

No direct relationship exists between the cause-specific hazard and the cumulative incidence function in estimating effects of covariates. Therefore, in such situations, the

emphasis must shift from the conventional modelling of cause-specific hazard function to modelling of quantities directly tractable to the cumulative incidence function (Fine and Gray 1999; Klein 2003).

In their study, Methods of competing risks analysis of end-stage renal disease and mortality among people with diabetes, Lim et al, (2010) showed that the estimates of the covariates coefficients on the cause-specific hazards and on the sub-distribution hazards models were different. Their study applied a cause-specific and sub-distribution hazards model to a diabetes dataset with two competing risks (end-stage renal disease (ESRD) or death without ESRD) to measure the relative effects of covariates and cumulative incidence functions.

Latouche et al, (2007), also showed that the effects of covariate on the cause specific hazard and on the sub-distribution hazard were normally different. This clearly shows that to test for effect of covariates on the CIF, a suitable regression model for the competing risks must be used. Lim et al (2010) concluded that either the cause-specific hazards model or the sub-distribution hazards model can be used for a dominant risk. However, for a minor risk we do not recommend the sub-distribution hazards model and a cause-specific hazards model is more appropriate in competing risk data analysis

The Sub-distribution and Cause-specific hazard model were applied to assess the effects of covariates on the cumulative probability of being discharged taking into account that a patient can die within the hospital period. The study then compared the effects of the covariates on the cumulative incidence and cause specific hazard to choose the best

model that best explained the relationship between the covariates and the cumulative incidence function or the cause-specific hazard function for the event of interest.

2.6 Model Diagnostics

This section presents different approaches to assess the assumptions under different models. These include the use of time varying covariates, Cox Snell for goodness of fit test and graphical approach using schoenfeld residuals.

2.6.1 Cox Snell Residuals

The basic issue involving the use of the Cox-Snell residuals is goodness of fit of the Cox PH model. As defined by Collet (2003), Cox-Snell residuals are given as

$$rc_i = \exp\left(\hat{\beta}' x_i\right) \hat{H}_0(t_i). \qquad (18)$$

When assessing the model, the plot of the integrated hazard based on the residuals against the hazard rate estimates backed out of the Cox model should have a 45-degree slope. Therefore if Cox model fits, then the residuals should be distributed as unit exponential i.e. should behave as if they are from a unit exponential distribution. The Cox-Snell residual was applied to determine if the model fit well to the data.

2.6.2 Time- Varying Covariates

Kleinbaum and Klein (2005) define time-varying covariates as any covariate whose value for a given subject may differ over *t*, whereas a time-independent variable is a variable whose value for a subject remains the same over *t*. For our study, age and HIV status can be regarded as time-varying covariates. Collet (2003) defined an Internal and External

time-dependent variable. Internal variables are related to a patient within the study and can be measured if and only if the patient is alive e.g. Blood pressure, CD4 count etc. While external variables are variables that do not necessarily need the patient to be alive for example Age of a patient. Time varying covariates can be used in the different survival models and they produce time-varying coefficients. If for example the Cox PH model includes a time-dependent variable X(t)then the model becomes:

$$h_i(t) = h_0(\mathrm{t})\exp\{\sum_{j=1}^{p}\beta_i x_{ij}(t)\} \qquad (19)$$

Where $h_0(t)$ is the baseline hazard function for an individual for whom all the variables equal to zero and is constant. The values of the explanatory variables $x_{ij}(t)$ depends on time and in such a situation the proportional hazard assumption is violated. This study used fitted models with time-varying covariates to assess if the PH assumption was met in the Fine and Gray model.

2.6.3 Schoenfeld Residuals

In this study three types of models were considered. These are the non-parametric cumulative incidence function, Cox-cause specific hazards and sub-distribution hazards model. The Cox-CSH and SDH model assumes that the hazard ratio comparing any two specifications of a covariate is constant over time. This means that the hazard for one individual is proportional to the hazard for any other individual (Cleves, et al, 2010).

To check whether the PH assumption is met in respect to a particular covariate, the Scoenfeld residuals proposed by schoenfeld (1982) is used. Collet (2003) denotes the *ith* scoenfeld residual for X_j, *jth* explanatory variable in the model as given by;

$$r_{pji} = \sigma_i\{x_{ji} - \hat{\alpha}_{ji}\}; \qquad (20)$$

Where x_{ji} is the value of the *jth* explanatory variable, *j*=1,2,3, … ,p, for *ith* individual in the study. The schoenfeld residuals are particularly useful in evaluating the PH assumption after fitting a Cox regression model.

2.6.4 Martingale's Residuals

These residuals are used to check the functional form of continuous covariates. Hosmer and Lemeshow (1999) define the martingale residuals as;

$$\widehat{M}_i = C_i - \widehat{H}_i \qquad (21)$$

Where the components of the residual for the *ith* subject are the values of the censoring variable C_i and the estimated cumulative hazard $\widehat{H}_i=\widehat{H}(t_i, x_i, \hat{\beta})$. Therneau, Grambsch and Fleming (1990) proposed fitting the Cox model without the covariate. The results are then used to generate smoothed values such as lowess smooth. These are then plotted against the values of the excluded covariate.

CHAPTER THREE

METHODOLOGY

This chapter describes the methodology used in this study. In particular, study design; data collection and data analysis; analysis approach and lastly ethical consideration.

3.1 Study Design

The study used secondary data from Surveillance Programme of In-patients and Epidemiology (SPINE) project, collected at Queen Elizabeth Central Hospital, in Malawi. This study was a retrospective cohort analysis of data from people with all forms of TB in the year 2014, (from 1st January 2014 to 28th November, 2014).

3.2 The SPINE Data

SPINE (Surveillance Programme of In-patients and Epidemiology) project is a computerized real time data collection system. The information system recorded tracked and managed in-patient care and appointment data. The patient registration system allowed all patients to be recorded with relevant details. Using a unique barcode for each, it was able to identify patients so that their records could be retrieved from the system in future visits by simply scanning their assigned barcodes.

The SPINE data was availed for this thesis in Microsoft Excel spreadsheet format. It covered patient's diagnosis and admission information from January 2010 to November 2014. The dataset used in this study contained information on adult in-patients only. An adult here was defined as any individual 15 years of age and above.

3.3Data Collection and Management

The study extracted the TB cases into Microsoft Excel 2007 from the SPINE database admitted from January and followed up for 6 months. The study utilized information on adult male and female patients who had been admitted with TB and were on treatment. Time to discharge or death whichever occurs first was captured. Date of admission and date of discharge from hospital after a treatment outcome observed was collected. Socio-demographic characteristics and clinical information was collected from all subjects. The socio-demographic characteristics included were age and gender. The clinical data included HIV status; ARV status; date of HIV test and a patients TB class (whether pulmonary or extra pulmonary). Once patients were admitted, they were tested for their HIV status, if found reactive, they were put on ART treatment. These records were entered into SPINE under the medical records for the patient for future reference.

3.4 Sample size and Sampling procedure

The data of this study came from Queen Elizabeth Central Hospital through Malawi Liverpool Wellcome Trust. The data had information on patients with all forms of TB and admitted due to unstable clinical conditions. A representative sample of 4500 TB admissions were available in the dataset of adults in the required age group of 15years

and above which was the target population of this study. The study analysed information from 2220 TB patients who were admitted with TB during the interested study period.

3.4.1 Inclusion and Exclusion Criteria

- The study looked at TB patients (15 years old) who were admitted within January and June 2014, the entry point was admission in hospital due to TB or bad clinical conditions other than TB.
- Patients on TB treatment below 15 years old and who were treated as out-patient were not included in the study

3.5 Study Outcome

The main outcome variable of this study was

- Time to discharge from hospital

Time to death was also considered as an outcome variable but was used as purpose of explaining its effect on modeling time to discharge with death as a competing event.

3.6 Data Handling and Description

The data was collected from the QECH spine database and patient case records once authorization was sort and approval was given by the College of Medicine Ethical Committee. The data was explored to obtain important variables that would be used for analysis. The data was first cleaned, and then sorted for easy navigation when doing analysis.In this study the individual patient (with TB) was the unit of analysis and the outcome variable consists of situations which were times to: discharge (main event),

death (competing event) or censored (Transferred, Referred and absconded). The variables under censored were grouped into one variable “Censored” due to small sample data within each variable. Length of stay (time to discharge) was calculated from date of hospital admission to the date of discharge including any hospital transfers that occurred. Categorical variables were coded using numbers e.g. male =0, female=1; HIV positive = 1, HIV negative=0; died=0, discharged=1, censored=2. Age and time to discharge were continuous variables. Survival time was measured in days.

3.7 Data Analysis

The analysis first looked at some descriptive statistics (frequencies, Inter-quartile range, and median) for the baseline characteristics. The cumulative incidence curve and Kaplan Meier curves were compared. CI curves for categorical variables; gender, HIV status and ART status were obtained and comparison between the different groups for the patients in terms of survival was performed. Secondly the Pepe and Mori test was done to compare cumulative incidence to discharge between groups for gender, HIV status and ART status. The null hypothesis was defined as the cumulative incidence to discharge is the same for both groups. Inferential statistics involved obtaining hazard ratios for the calculated probability. Finally at a multivariate level factors that affect the time to discharge for a patient on TB treated were obtained.

Statistical analysis were done using STATA version 12, a statistical software package created in 1985 by StataCorp used in data management, Statistical analysis, graphics and Simulations. Two extra programs from Statistical Software Components archive was

needed to conduct analysis on non-parametric cumulative incidence function. To estimate nonparametric cumulative incidence function, the command stcompet (refer to Appendix 1)by Coviello and Boggess (2004) was installed. To test equality of cumulative incidence functions among groups, the command stpepemori written by Coviello (2008) was used. The sub-distribution hazards were performed using Stata 12 command stcrreg. The Schoenfeld residuals and plots were used to test the PH assumption. The Martingale residuals were used to check the Linearity of variable age. Time-varying covariates were used to test for PH assumption for the Sub-distribution hazard model.

3.8 Model Specification

Competing risks are represented by the failure time T, the failure cause D and a vector of covariates Z. T is assumed to be a continuous and positive random variable, D takes values in the finite set $\{1, \dots, i\}$. The failure cause D can be either the event of interest, in our case D=1 representing “Discharge” D=2 representing “Death” and D=3 representing “Censored”. This study used semi-parametric proportional hazard models because of their flexibility (no distributional assumption on time and availability of software for fitting these models).

3.8.1 Plotting Cumulative Incidence Function

The estimation of the probability of occurrence by time t, for a particular failure can be handled by fitting 1-KM, the complement of the Kaplan-Meier estimator or the cumulative incidence function. This study did consider 1-KM for the estimation because

it leads to bias when dealing with competing events, but a comparative analysis was done between the two.

CIF is the probability of experiencing an event by a given time. Denoted as I_k it describes the risk of failing from cause *k* until time *t*: $I_k(t) = P(T \leq t \text{ and } D = k)$.

3.8.2 Modelling Cause-Specific Hazards

As stated in the previous chapter, the cause-specific hazard function for failure cause *k* is the instantaneous failure rate of failing at time *t* of cause *k*.

The cause-specific hazard function for the k-th cause is defined by;

$$h_k = \lim_{\Delta t \to 0} \left\{ \frac{P(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t} \right\}$$

For D ϵ {1,2, … , k}. It represents the hazard of failing from cause j in the presence of the competing events. The regression model on cause-specific hazards is as follows:

$$h(t|z) = h_{0k} e^{\beta' z}$$

The total hazard $h(t; z)$defined in terms of the cause specific hazards equals the corresponding hazards function summed up to time t as follows;

$$h(t|z) = \sum_{k=1}^{2} h_k(t)$$

This implies that the all cause hazard rate is the sum of K hazards (Grey 1988).Several studies have pointed out that the Cox-Proportional hazards can be used to model the cause-specific hazards in regression modeling (Aban, 2014). Although this is the case, cause-specific hazards have some shortfalls, one of the problems being that the above method requires the assumption that competing risks are independent (Kleinbaum &

Klein, 2005) which is not the case when dealing with competing risk data. In this study the cause specific hazard was modeled using the Cox model, which is broadly used in medical research.

3.8.3 Modelling Sub-distribution Hazards

Fine and Gray (1999) developed a semi-parametric model that considers all important factors in a competing risk setting. These factors are the baseline hazard effect for the outcome events, the covariate effect for the outcome events and the effect of time itself. It directly links the covariates to the cumulative incidence function. The Fine and Gray is a proportional hazards model for the sub-distribution hazard of the event of interest defined as

$$\bar{\lambda}_1(t) = -\frac{d\log(I - I_1(t))}{dt}$$

Given covariate X, the model is of the form$\bar{\lambda}_1(t|X) = \bar{\lambda}_{1,0}(t)\exp(\beta^t X)$, where $\bar{\lambda}_{1,0}(t)$ is the baseline sub-distribution hazard for the event of interest. This study used this method as explained in literature that it is a model that directly links covariates to the cumulative incidence of discharge, therefore this method was appropriate to identify prognostic factors of length of stay.

3.9 Ethical Consideration

Full ethical approval was granted by College of Medicine Research Ethics Committee (COMREC) to collect data from Queen Elizabeth Central Hospital. Patients' names were not used during analysis so as to uphold confidentiality. Refer to certificate of approval in Appendix 4.

CHAPTER FOUR

RESULTS AND DISCUSSION

This chapter presents and discusses the results that have been obtained from the study analysis. Section 4.1 presents the exploratory data analysis, section 4.2 presents the fitted models, and section 4.3 presents model assumption assessment.

4.1 Exploratory Data Analysis

The SPINE dataset had 1325 patients who were admitted at QECH between January 2014 and November 2014 for TB related diseases. Out of 1325, the study analysis considered a total of 1220 TB-infected patients.

Table 1: Baseline Characteristics

Frequency		**Interquatile Range**		
	Median	**Std. Dev.**	**25th**	**75th**
Age (Years)	35	11.9	29	43
Time to discharge (Days)	11	20.4	6	20
Categorical Variables	**n(%)**			
Gender				
Male	679 (55.6)			
Female	541 (44.4)			
HIV Status				
Positive	996 (86)			
Negative	162 (14)			
ART Status				
No	252 (25.3)			
Yes	731 (73.3)			
Defaulter	14 (1.4)			
Health Outcome				
Discharged Alive	891(73.03)			
Dead	322(26.39)			
Censored (Transferred, referred & absconded	7(0.57)			

Table 1 gives a summary of the baseline characteristics of the patients included in the study. The median time to discharge for TB patients in the year 2014 was 11 days. Out of 1220 TB patients, 891(73.03%) were discharged alive while 322 (26.39%) died while in hospital and 7 (0.57%) where either transferred, referred or absconded the admission, 678 were males representing 55.6% and 996 (86%) patients were registered as HIV positive. Out of the 996 HIV positive patients, 731 (73.3%) were on ART therapy, while 252 (25.3%) were not on ART therapy. Table 1 show that, the percentage of TB patients who were HIV positive was high (86%) as compared to TB patients without HIV. This observation agrees with the WHO 2003 report, which stated that most common cause of

immuno-suppression in Malawi is HIV infection that leads to AIDS and that HIV infection leads to rapid progression from TB infection to disease and increases the risk of re-activation of old infection into active disease. The lifetime risk of developing TB of HIV non-infected individuals is between 5 to 10% while that of infected individuals is between 30 to 50% or 5 to 15% per year(WHO, 2010).

The primary diagnosis variable which explained the patients diagnosis at admission constituted of different type of Tuberculosis, which included; Tuberculosis Miliary, TB sepsis, TB spinal, TB meningitis, TB pulmonary etc. Appendix 2summarizes the various TB categories that the patient's in this study were primarily diagnosed of. The table shows that 469 patients had Pulmonary TB and that there were some categories stated as TB, Pleural PTB (Pulmonary Tuberculosis), PTB relapse, some of these fell under Pulmonary TB. It also shows that 80 patients had Tuberculosis EPTB (Extra-pulmonary Tuberculosis) but they were also others who had EPTB relapse, Tuberculosis Iris, Tuberculosis Anemia etc. Length of hospital stay was based on the primary diagnosis of all forms of TB.

Table 2: Characteristics of the TB patients by Outcome category

Variable	Categories	Outcome Categories			
		Alive	**Dead**	**Censored**	**Total**
Sex	**Male: n (%)**	478 (70.3)	199 (29.3)	3 (0.4)	680
	Female: n (%)	415 (76.6)	123 (22.7)	4 (0.7)	542
Art Status	**No: n (%)**	194 (76)	58 (23)	1 (0.4)	253
	Yes: n (%)	526 (72)	200 (27)	5 (0.7)	731
	Defaulter: n(%)	12 (86)	2 (14)	0 (0.0)	14
HIV Status	**Non-Reactive: n(%)**	126 (78)	36 (22)	0 (0.0)	162
	Reactive: n(%)	731 (73)	259 (26)	6 (0.6)	996
Age (Years)	**Mean (IQR)**	34 (13)	38 (14)	42 (20)	35 (14)
Failure time(Days)	**Median (SD)**	12 (21.7)	10(15.8)	5.5(5.4)	11(20.4)

Table 2 shows the baseline characteristics against the dependent variable type of failure. The results from the table showed that there was a high in-hospital mortality rate in the various categories. Out of 680 male TB patients, 478 (70.3%) were discharged alive while 199 (26%) died while in hospital. Although this is the case, it can be observed that the percentage of death in males is higher than in females. One of the reasons suggested of this difference is that males in general have higher risk of acquiring *Mycobacterium tuberculosis* infection because of a wider network that leads to a greater exposure to the organism (Johansson. et al., 2000). Out of 996 HIV positive TB patient's 731(73%) patients were discharged alive. Out of 162 HIV negative patients 126 (78%) were discharged alive while 36 (22%) died while in hospital. There were no patients who were transferred or referred from this group. Among the HIV positive patients, out of 731 who started ART therapy, 526 (71.96%) were discharged from hospital and 58 (76.68%) died

while in hospital. A higher percentage of HIV negative patients were discharged than HIV positive patients but the results were not statistically proven if significant.

Table 2 shows that, the median time to discharge alive was 12 days and 10days for patients who died in hospital. The median hospital stay for TB patients at QECH was 11 days. Holmquist et al. (2008) found that in 2006, the average hospital stay in the US for a primary Tb diagnosis was 15.0 days more than twice the average stay for a patient with a secondary TB diagnosis (6.6 days). A study on Tuberculosis on African refugees from Eastern sub-Saharan Africa found that the average length of hospitalization for the TB patients they studied was 8.7 days. The patients were admitted due to TB related diseases or due to clinical conditions other than TB (Nesher. et al., 2012). A study done in Botswana showed that Mean duration of stay in the hospital for TB patients was 12 days (Stolp. et al., 2013). The duration of stay in this study was determined by TB illnesses and not diseases un-related to TB. These non-TB diseases could be contributing factors to length of stay. Appendix 3 summarizes patient's secondary diagnosis. A study done in Israel studied length of stay of patients with TB who were admitted due to various reasons on top of TB illnesses. Their study showed that the mean LOS was 8.7 days. One reason explaining this difference with results from this study, could be because the hospital under study in Israel did various tests to identify TB and the patients were admitted mainly due to TB which is not the case with Malawi where simple tests are done but advanced tests that require high technology are not available such as extensive radiological investigations: a chest computed tomography scan, abdominal CT and spine

CT and magnetic resonance imaging, trans-bronchial biopsy and pleural biopsy (Nesher. et al., 2012)

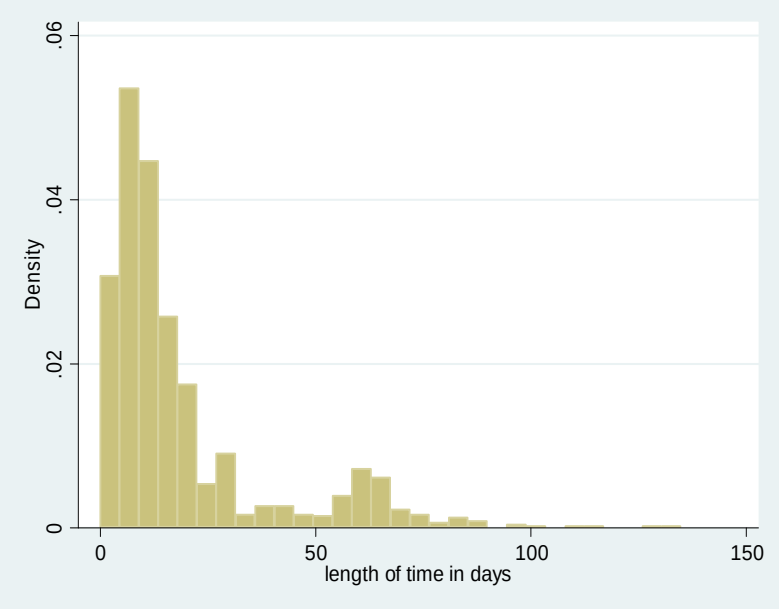


Figure 1: Distribution of Patient's Length of time in hospital

Figure 1 shows the distribution of patient's time in hospital. The figure shows that time to discharge for a patient was skewed to the right with most patients being discharged around day 11.

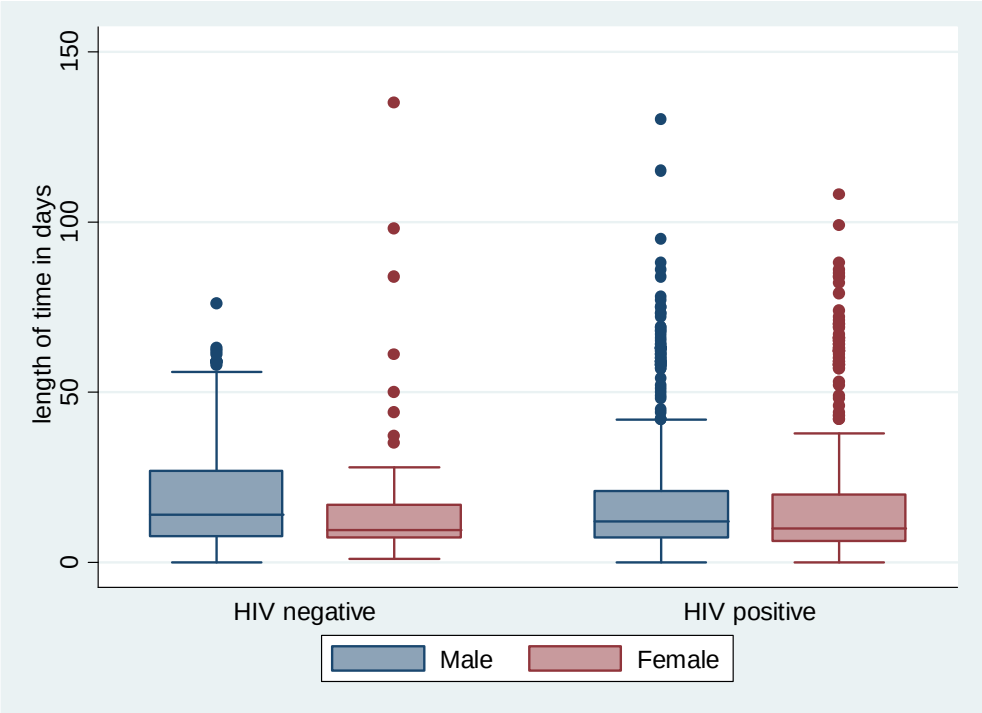


Figure 2: Distribution of time to discharge by Gender and HIV Status

Figure 2 shows that for females HIV positive or negative the median length of hospital stay is similar. The same applies to HIV positive and negative males, the median length of hospital stay is similar and the median length of stay is higher for males than for females. The median length of hospital stay for males is 12 days and 10 days for females with varying time outliers. Figure 2 shows that the distribution of time to discharge was right skewed for males and females who are HIV positive and negative.

4.2 Comparison of Cumulative Incidence and Complement of Kaplan Meier (1-KM)

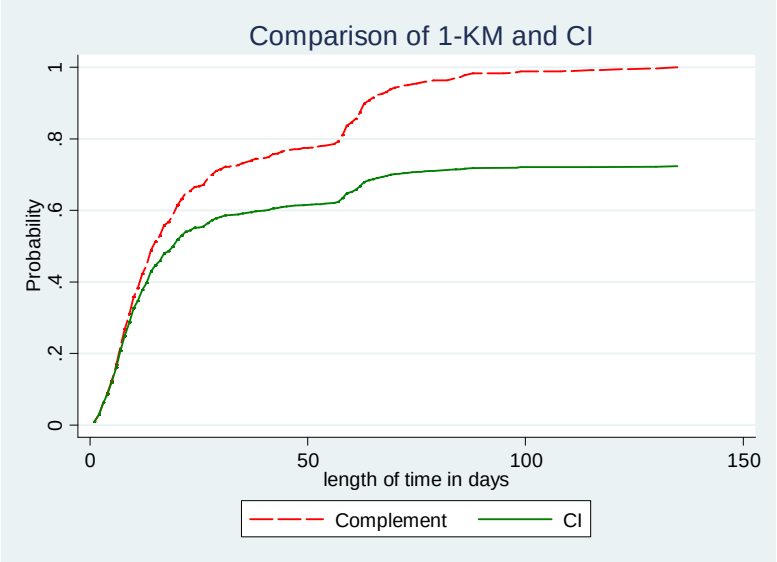


Figure 3: Comparison of 1-KM and Cumulative Incidence (CI) curves

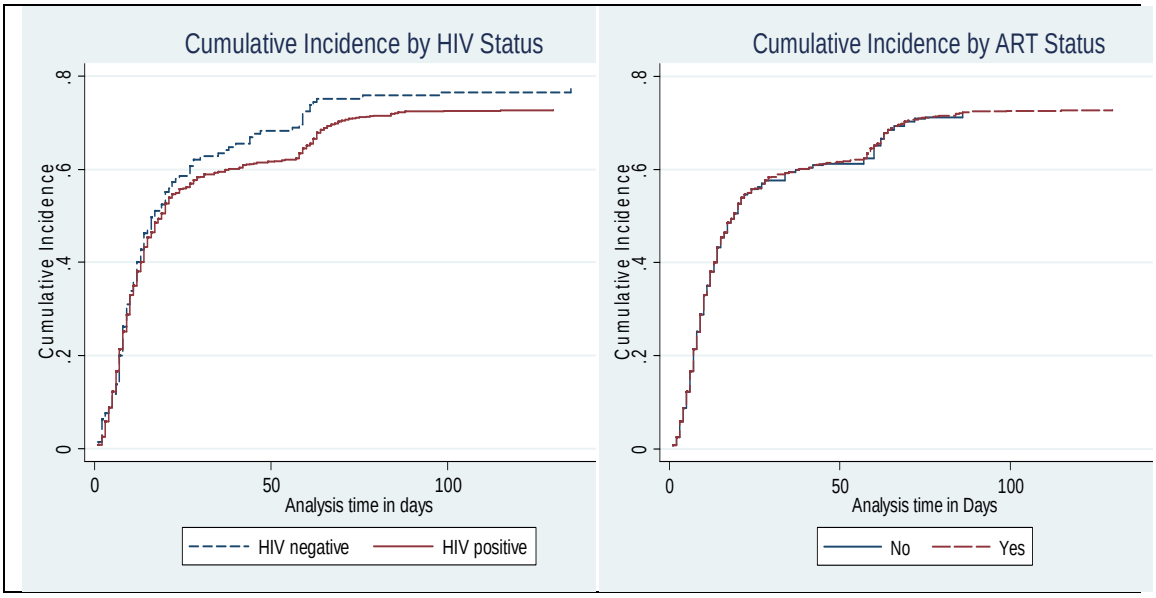
The 1 minus Kaplan-Meier (1-KM) estimates and cumulative incidence estimates were generated, plotted and compared. From day 1 the 1-KM estimates and the CI estimates were similar. As shown in Figure 3 the estimates for CI and 1-KM are similar but as the number of day's increases, the estimates greatly differ from each other. The 1-KM estimator provides inflated probabilities of discharge among the TB patients as compared to the cumulative Incidence. The difference is very noticeable after 10 days and increases with more competing events i.e. death as evidenced from Figure 3. Whereas the cumulative Incidence estimates the probability of discharge before time t and its cause specific hazard which is the conditional probability of being discharged before a time interval given that an individual has survived and did not die up to time t. Thus the CI estimates the probability of discharge taking into account that one might die before a discharge resulting in true and realistic estimates of probability of event discharge. The

cumulative Incidence estimates and compares cumulative event probabilities of a specific event.

This finding is similar to several studies and authors (Borrebach, 2013; Gooley. et al., 1999; Sherif, 2007)that have pointed out that the CI is an appropriate tool to use for estimation in the presence of competing risks. Sherif (2007) stated that the use of 1-KM to estimate cause-specific cumulative probabilities leads to inflated estimates of proportion of patients at risk of failure at time t. Since the 1-KM makes an assumption that the probability of failing prior to time t from cause k is equal to 0.

4.3 Model Estimation Results

4.3.1 Non-Parametric Cumulative Incidence functions



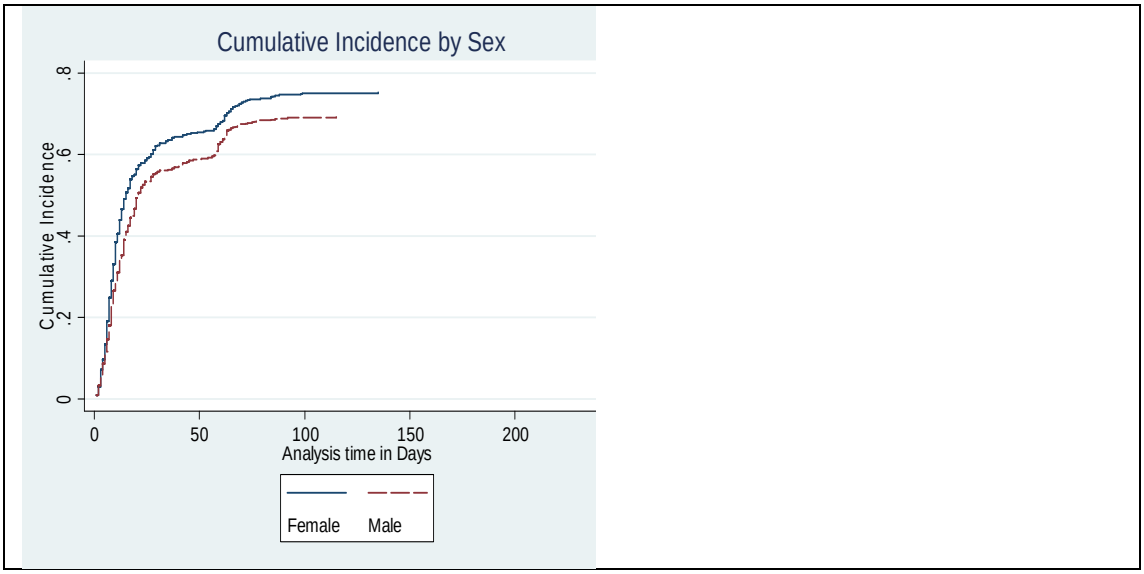


Figure 4:Non-parametric cumulative Incidence functions for HIV Status, ART Status and Gender

Figure 4 presents a comparison of the CIF's for categories within groups. The figure shows that HIV negative patients had a higher likelihood of being discharged than HIV positive patients. HIV negative patients had a 0.65 probability of being discharged by day 50 while HIV positive patients had a 0.6 probability of being discharged. The CI curve for ART shows no difference, implying that there is no difference in the probability to discharge for patients on ART and for those not on ART.

Table 3: Pepe and Mori cumulative Incidence tests

Parameter	Outcome Event		χ_1^2	P-Value
HIV Status	**Main Event**	Discharged	0.158	0.691
	Competing Event	Death	1.993	0.158
ART Status	**Main Event**	Discharged	0.650	0.420
	Competing Event	Death	0.834	0.361
Patient's Sex	**Main Event**	Discharged	0.476	0.490
	Competing Event	Death	2.307	0.129

The p-values for the Pepe and Mori tests, for both events (discharge and death) from Table 3 above lead to failure to reject the null hypothesis which states that the cumulative incidence for the categories are similar. This shows that there is no significant difference in cumulative incidence of discharge and death for the three categorical variables (p-value > 0.05). Based on the test at a 5% probability of making an error, HIV positive and HIV negative TB patients have the same likelihood of being discharged or dying from hospital. Although this is the case, Figure 4 shows that HIV negative patients with TB at QECH had a higher likelihood of being discharged from hospital than HIV positive patients. Figure 4 also shows that females had a higher likelihood of being discharged than the males though the Pepe and Mori test showed that there was no significant difference in the cumulative incidence between males and females. In terms of length of hospital stay, Figure 4 showed that HIV positive TB patients and male TB patients seemed to have a longer hospital stay before discharge than HIV negative patients and female patients respectively.

4.4 Competing Risk Regression Models

4.4.1 Semi-Parametric Analysis

The study was interested in the taking into account the competing event (death) when estimating the effects of Age, Sex, Primary Diagnosis and HIV Status on the hazard of discharge for admitted TB treated patients. This effect was determined by observing the estimates obtained from the CSH model and estimates from the SDH model, if the estimates are similar between the models then the assumption that is used when modeling CSH of independence between the main event and competing event holds (Dignam et al,

2012). This would imply that death does not affect estimation of the covariates on the main event “Discharge”.

The probability of discharge within the study period was not different for various groups within the variables; HIV status, ART status and gender under the Pepe and Mori test. Despite such results, the non-parametric curves for HIV status and gender in Figure 4, explain existence of some chance of differences in time to discharge between the levels within these categorical variables. This section presents Cause-specific and Sub-distribution models for time to discharge and time to death against the covariates age, HIV status and gender. The variable ART status was not included in the model due to its insignificance.

4.4.2 Analysis of Covariate effects on events Discharge and Death

Estimates obtained from fitting the sub-distribution hazard based on Fine & Gray (1999) and the Cox cause-specific hazard models are presented in Table 4 and 5. These models provide a good check for independence of events assumption made when implementing cause-specific models. The Univariate and multivariate sub-distribution hazard and cause-specific models for main event discharge and competing event death are presented.

Table 4: Comparison of Univariate CSH and SDH for event "Discharge" and competing event "Death"

		Model Effect Estimates					
		CSH			**Fine and Gray- SDH**		
Event Type **Discharge**		**HR**	**p-value**	**95% Estimate CI**	**SHR**	**p-value**	**95% estimate CI**
Age(in Years)		0.99	0.023	(0.99, 1.00)	0.985	<0.001	(0.98,0.99)
HIV Status	**Negative**	Reference			Reference		
	Positive	0.963	0.716	(0.79, 1.18)	0.894	0.257	(0.74,1.09)
Sex of Patient	**Male**	Reference			Reference		
	Female	1.11	0.134	(0.97, 1.28)	1.23	0.004	(1.07,1.41)
Death							
Age(in Years)		1.02	<0.001	(1.012,1.032)	1.02	<0.001	(1.015,1.033)
HIV Status	**Negative**						
	Positive	1.41	0.069	(0.97 2.05)	1.499	0.029	(1.04-2.15)
Sex of Patient	**Male**						
	Female	0.92	0.499	(.72 1.17)	0.85	0.189	(0.67-1.08)

Table 4 presents Univariate cause-specific hazard and sub-distribution estimates for event discharge and competing event death. HIV status was a significant effect on the sub-hazard of discharge but was not significant for the CSH model. Variable Age significantly affects time to discharge among the TB patients for both models (*$p<0.05$)*. Based on the p-value for variable age, it is regarded as a significant predictor, but the confidence interval does contain a 1. In this instance age is still a significant predictor but its effect is based after a large increase in age. Therefore with a 20 years increase, older patients were less likely to be discharged by 2% compared to younger TB patients. Patients Sex did not show a significant effect but it is observed that female patients have a 9% likelihood of being discharged than male patients when death is treated as censored and 12.7% with death as a competing event.

Univariate CSH models and SDH models were again fitted for competing event death. Table 4 presents the hazard and sub-hazard estimates obtained. Similar to the CSH model for event discharge, Age significantly affects the cause-specific hazard for death. As age increase, the cause-specific hazard of dying for a TB patient in hospital increases by 2%. In other terms, older TB patients are more likely to die while in hospital than younger patients. For the CSH model HIV status and Sex were not significant predictors in explaining hazard of death.

The Univariate SDH models in Table 4, showed that age significantly affected the cumulative Incidence of discharge (via the sub-distribution hazard) with a sub-hazard of 0.985 (95% CI: 0.98-0.99) and p-value<0.001. Gender is also statistically significant with

a sub-hazard of 1.23 (95% CI: 1.07, 1.41) and p-value=0.004. Older TB patients are 1.5% less likely of being discharged with time. Female patients are 23% more likely to be discharged within the 6 months than male patients. The results show that for the univarate CSH and SDH model, HIV status had no significant effect on time to discharge from hospital.

Table 5: Comparison of Multivariate CSH and SDH for event "Discharge" and competing event "Death"

		Model Effect Estimates					
		Cox CSH			Fine and Gray- SDH		
Event Type	**Category**	**HR**	**p-value**	**95% Estimate CI**	**SHR**	**p-value**	**95% estimate CI**
Discharge							
Age(Years)		0.993	0.040	(0.99,1.00)	0.986	0.001	(0.98,0.99)
HIV Status	**Negative**	Reference					
	Positive	0.949	0.611	(0.78, 1.16)	0.824	0.048	(0.68,1.00)
Gender	**Male**	Reference					
	Female	1.09	0.244	(0.94, 1.26)	1.127	0.097	(0.98, 1.30)
Death							
Age(Years)		1.02	<0.001	(1.01, 1.03)	1.02	<0.001	(1.013, 1.29)
HIV Status	**Negative**	**Reference**					
	Positive	1.25	0.227	(0.87, 1.80)	1.24	0.238	(0.87, 1.78)
Gender	**Male**	Reference					
	Female	0.811	0.078	(0.64, 1.02)	0.75	0.017	(0.597, 0.95)

A multivariate SDH model was fitted for covariates age, Sex and HIV status as presented in Table 5. The results showed HIV status and age were statistically significant in predicting the sub-hazard of discharge. HIV status had a sub-hazard ratio of 0.824 with a p-value of 0.048. HIV positive patients had a 17.6% less sub hazard of being discharged from hospital than HIV negative patients. Older patients again were 1.4% less likely to being discharged with an increase in age by 20 years. Variable sex turned out to be statistically insignificant. A univariate CSH model for effect of ART on time to discharge with death as a competing event was fit. ART status of a patient was found not to be a significant predictor in modeling time to discharge ($\boldsymbol{\chi^2_{(1)} = 8.39}$, p=0.079), A SDH model of event discharge was fitted for ART Status and it was found to be statistically insignificant. Overall ART was not significant in explaining time to discharge in the presence of a competing event death.

The multivariate model of CSH was fitted for event discharge and competing event death, the estimates are shown in Table 5. The results show that HIV positive TB patients were 41% more likely to die in hospital than HIV negative patients. Age was the only significant factor affecting time to death in the CSH model. The CSH model for event discharge showed no significant factor. Females are 8% less likely to die in hospital than males, though this effect is not significant. ART still remains insignificant with $p>0.05$.
The Multivariate SDH model for competing event death showed that age and HIV status are statistically significant in explaining the cumulative incidence of death. HIV positive patients are 50% (SDH=0.499) more likely to die in hospital than HIV negative patients. Once again older patients are more likely to die in hospital within the admission time.

ART status was again not statistically significant in explaining cumulative incidence of death with $p>0.05$.

Two approaches of modeling can be used when competing risks are present: modeling the cause-specific hazard and modeling the sub-distribution hazard which takes into account the competing risk. The results show that, the SDH estimates and the CSH estimates were slightly different. This shows that the contribution of death in reducing association between covariates and discharge was minimal. This is in-line with various studies that have shown that the covariate effects using the CSH model or the SDH model differ (Teixeira. et al., 2013) as shown in this study. A covariate not significant on hazard of main event can be significantly associated with cumulative probability of that main event if the covariate influences the hazard of the competing event (Dignam. et al., 2013). Fine & Gray., (1999) also showed that the parameter estimates for the CSH and SDH model differ for the main event.

Several authors have differed on the type of model to use to estimate effects of covariates on the probability of the event of interest. Andersen. et al., (2012) pointed out that the cause-specific hazards may be more relevant when the biological mechanism of the disease is of interest, since it quantifies the event rate among the ones at risk of developing the event of interest. One of the drawbacks for the CSH model is that it fails to directly link the effect of the covariates on the CIF of the event of interest. Many have proposed use of the Sub-distribution hazard model since it directly links the covariates to the cumulative incidence function. The results from this study are in agreement with

various authors (Fine. &Gray., 1999; Kim., 2007; Lim. et al., 2010), who stated that the SDH model is a better model when studying data that involves competing risks. Therefore, to meet the objective of identifying prognostic factors of discharge in the presence of competing risks, the sub-distribution hazard model was a better model than the CSH model.

4.4.3 Comparison of Cumulative Incidence curves for Predictors based on Fitted Models

In addition to the CSH and SDH models fitted, Cumulative Incidence curves for the SDH models were plotted to compare the CI's for categories within the variable Sex and HIV status to clearly observe the differences. Figure 5 and 6 show the results after plotting the CI for sex and HIV status after fitting the SDH models.

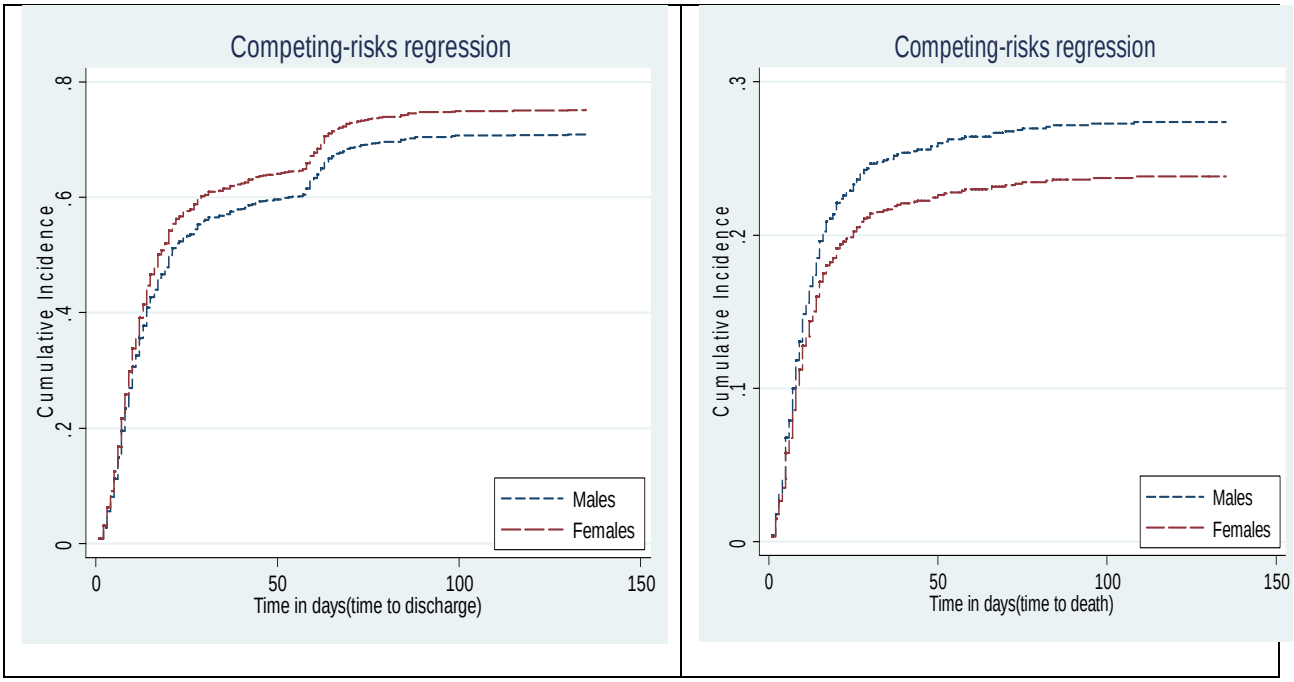


Figure 5: Cumulative Incidence by Sex for events "Discharge" and "Death"

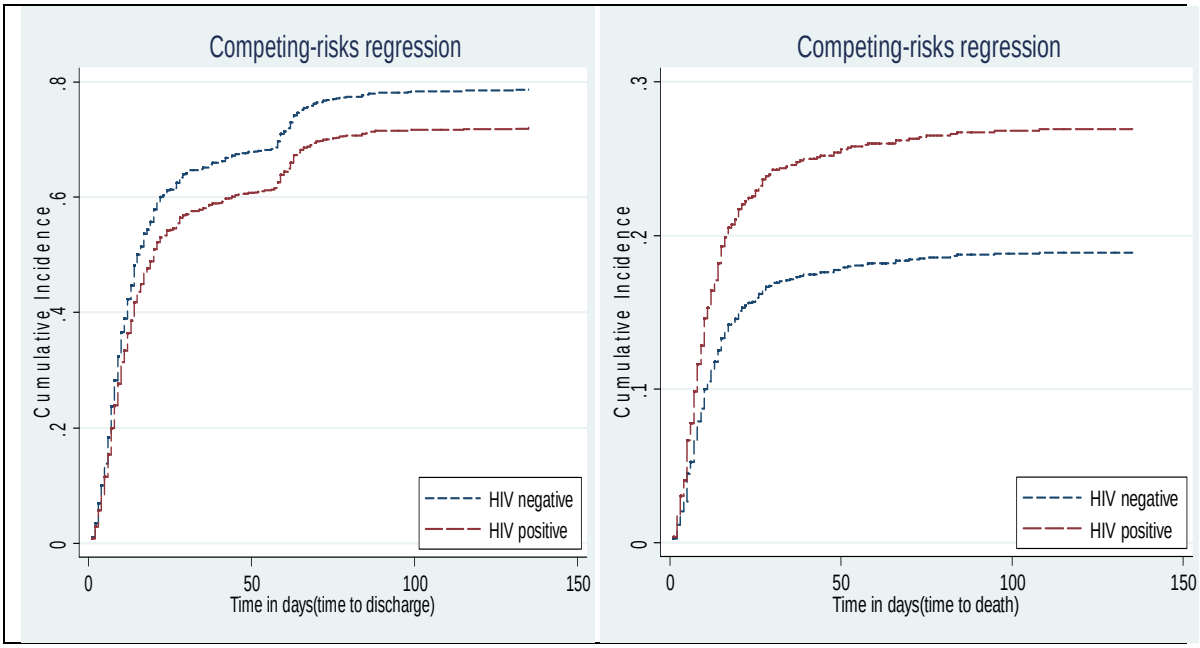


Figure 6: Cumulative Incidence by HIV Status for events "Discharge" and "Death"

The results from the plots in Figure 5 and 6 represent what has been obtained in the SDH models. For the main event discharge, there's a slight difference between the CI's for males and females. Females had a higher likelihood of being discharged than males. For competing event death, at day 25, females had a 0.2 cumulative incidence of death than males who had a 0.25 probability of dying within the study period. The CI curves for the variable HIV status for event discharge showed a slight difference in the CI's between HIV positive and negative patients. For event death, HIV positive patients had a higher probability of dying than HIV negative patients as evidenced from the CI curves. With a 0.15 probability of dying for HIV negative patients by day 25 and 0.25 probability of dying for HIV negative patients.

4.4.4 Factors Affecting LOS

Age and HIV status were identified as factors associated with a lower probability of discharge occurring, and a higher probability of death occurring in the sub-distribution multivariable regression modeling (Table 6). This is similar to one study which showed that Increasing age was associated with increasing risk of death for TB patients (Roberts & Daley, 2003). The results showed that older TB patients were less likely to be discharged from hospital than younger patients. This is likely since older people are more frail to diseases than younger patients, there immunities are much weaker than for the young people, therefore once they get infected or as soon as develop a disease it takes time for them to recover. This finding is in-line with findings from a study done by Holmquist et al, 2007 which showed that elderly patients were more likely to remain in hospital than younger ones. The 2007 US Vital and Health Statistics also reported that older patients have a longer average length of stay. This can be explained since it is commonly known that with advancing age, patients tend to have more comorbid chronic illnesses making them more vulnerable during hospitalization (Marengoni. et al., 2008) to this findings is a study done by (Çelik. et al., 2001), who mentioned that age, sex, residence, institution at which the patient admitted and insurance status determine unnecessary stay but statistically do not affect the average length of stay.

Gender was significant for Univariate sub-distribution hazard model. Females were 23% more likely of being discharge than males, which implied that as the days of admission progressed, the probability of discharge for females was higher than for males. This was clearly observed from the Cumulative incidence curves, which showed females having a

higher cumulative incidence of discharge than their male counterparts. This is likely, since most males might visit the hospital when very sick unlike females who might visit the hospital once they observe a discomfort. Ferreira et al 2014 showed that increased length of hospital stay was proportional to increasing age, especially > 40 years; and that males were more likely to stay longer in hospital than females.

Another interesting result on covariates affecting the probability of discharge taking into account that a death can happen was comparison between HIV positive TB patients and HIV negative TB patients. This variable was significant only in the sub-distribution model. This is possible where more HIV patients experienced the competing event death before a discharge and thus the effect of the competing event on the probability of the main event was noticeable. In this case, if the data involves a lot of the competing events it is best to use the SDH model which takes into account effect of the competing event on the probability of the main event(Teixeira. et al., 2013). HIV positive patients were 17.6% less likely of being discharged from hospital and 50% more likely of dying in hospital as compared to HIV negative patients. This is quite a high difference and would need attention of medical researchers to find out why there is such a gap between these two groups. Oliveira et al. (2009) in Brazil concluded that a high number of patients with TB/HIV are expected in hospitals as admission patients. In addition to what Oliveira et al, found, according to the Government-funded research conducted in South Africa, HIV positive patients stay in hospital four times longer than other patients on average. Malawi is among one of the countries severely affected by the dual epidemic of HIV and TB

(WHO report, 2012). Therefore it is expected to have more HIV positive patients in hospital than HIV negative patients.

4.5 Model Assumptions and Goodness-Of-Fit

This section presents results for assessment of model adequacy. The proportional hazard assumption for the Cause-specific model was performed. Cox Snell residual test was performed to test goodness of fit and Martingale residual plot were used to linearity for covariate age. Time-varying covariates were used when modeling the sub-distribution hazard model to test for proportionality assumption for SDH model.

4.5.1 Proportional Hazards Assumption of the Cause-specific Hazards for Event Discharge and competing event Death

Table 6 and 7 present results obtained after carrying out a proportional hazards assumption test on the full model for cause-specific hazard when the failure event for the patient's was discharge and death respectively.

Table 6: The Schoenfold's test for event Discharge

Covariate	Rho	Chi-Square	P-value
Age	0.019	0.28	0.5978
HIV Status: Positive	0.029	0.61	0.4331
Sex: Female	0.092	6.43	0.0112
Global test		7.59	0.0552

The Schoenfeld's global test assesses the assumption that the hazards in the time-to discharge and time to death (the Cox-proportional hazard models) are proportional over time, i.e. testing whether effects of covariates on the risk remain constant over time.

Specifically, the test computes a test for each covariate i.e. testing the null hypothesis that the model fits the data. The alternative states that the data does not fit the data. A p-value ($p<0.05$) means that the null hypothesis that the data fits well cannot be rejected.

Table 7: The Schoenfold's global test for Death

Covariate	Rho	Chi-Square	P-value
Age	0.004	0.00	0.955
HIV Status: Positive	-0.072	1.32	0.251
Sex: Female	-0.001	0.00	0.986
Global test		1.41	0.733

We observe that, at a 95% confidence level, the global test for the CSH model of discharge or death are not statistically significant (p-values > 0.05). This is evident from Table 6 and 7 where the global test is $p=0.055$ and $p=0.733$ respectively Therefore we accept the hypothesis of zero slopes, that means the assumption of constant proportional hazard for the CSH model of discharge or death holds.

In regression analysis, it is recommended to look at the graphs of the regression in addition to performing the tests of non-zero slopes. Therefore, Fig. 7 presents the graphs for the scaled Schoenfeld residuals for each explanatory variable versus survival time. The solid line is a smoothing-spline fit to the plot. The graph clearly shows that the fitted lines (slopes) for the scaled Schoenfeld residuals for each covariate are not significantly different from zero (i.e. no systematic departures from a horizontal line), that is confirming the test results obtained in the Schoenfeld global test.

4.5.2 Test for PH assumption ofCSH model for event “Discharge”

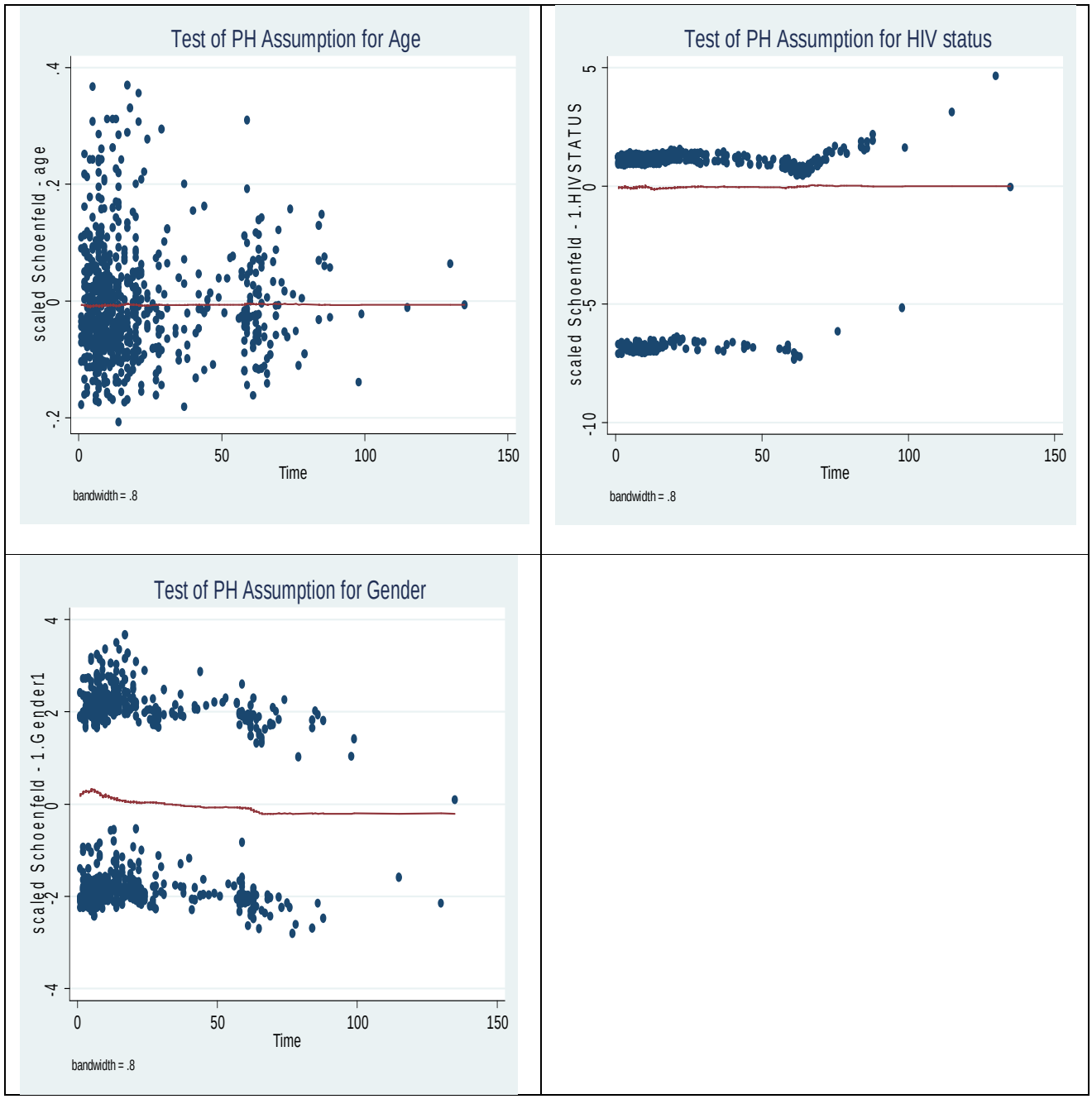


Figure 7: Schoenfold residual plots for each predictor for event discharge

4.5.3 Test for PH assumption of CSH model for competing event "Death"

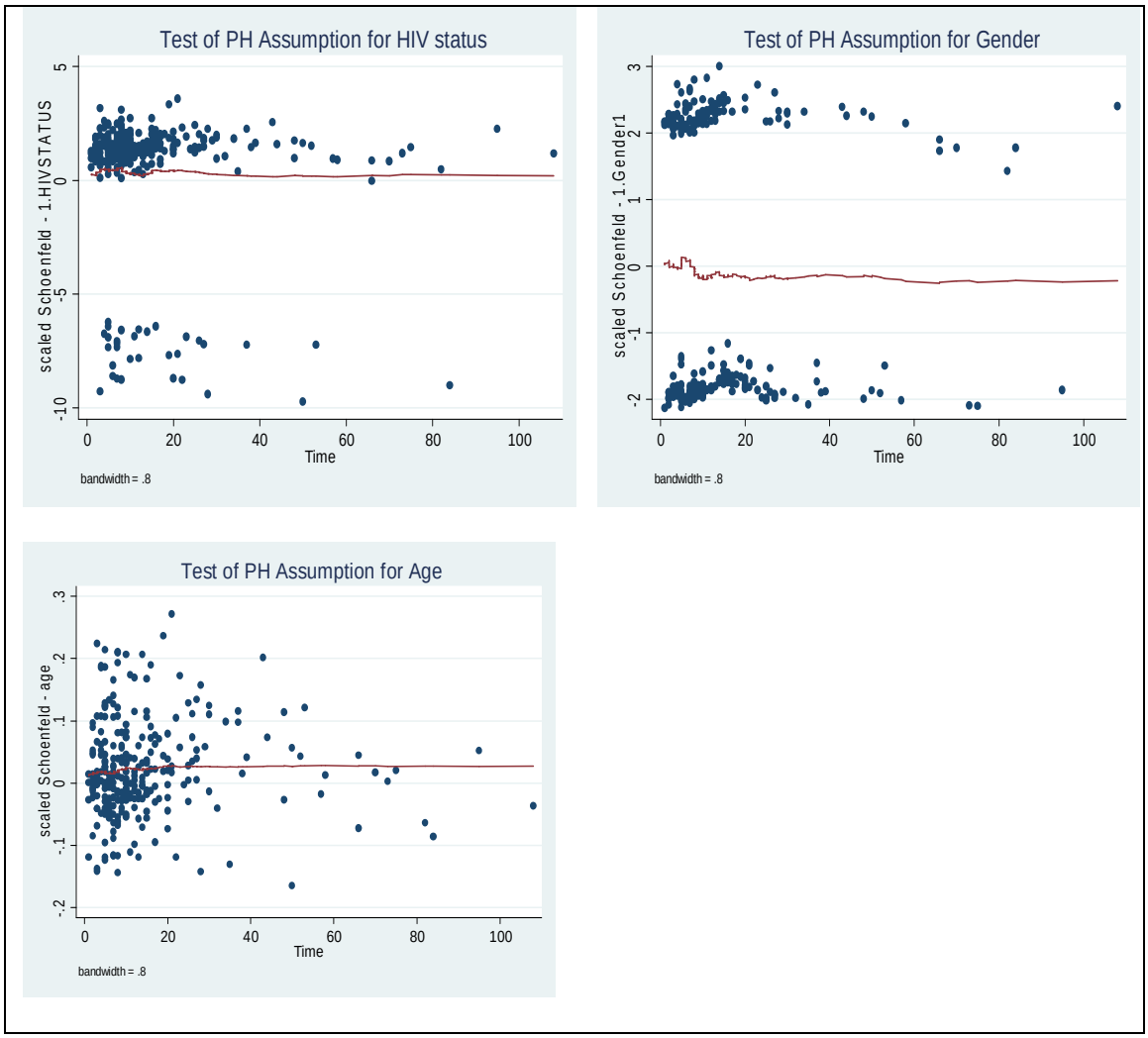


Figure 8: Schoenfold residual plots for each predictor for event death

4.5.4 Time-Varying covariates

In order to test if the Sub-distribution hazard model satisfied the proportional hazard assumption, a SDH model was performed with age, HIV Status as time-varying covariates interacting with the analysis time. Table 8 and 9 present's results that were obtained after fitting the SDH models for failure event "Discharge" or "Death"

Table 8: Time varying covariates for failure event "Discharge"

Model	Categories	SHR	95% CI	P-Value
Main: **HIV Status**	Negative(reference)			
	Positive	0.896	(0.68 1.18)	0.433
Age		0.99	(0.98 .999)	0.033
Time-Varying: **HIV Status**	Negative(reference)			
	Positive	0.996	(0.99 1.01)	0.376
Age		0.99	(0.999 1.00)	0.067

The estimated hazard ratios are split into two categories in Stata, hazard ratios for variables with constant time and HR for time-varying covariates. From table 9, it is observed that HIV status and Age did not significantly interact with time (p>0.05), therefore a conclusion can be made that the PH assumption for the Fine and Gray regression is not violated.

Table 9: Time varying covariates for failure event "Death"

Model	**Categories**	**SHR**	**95% CI**	**P-Value**
Main: **HIV Status**	Negative(reference)			
	Positive	1.726	(1.08 , 2.75)	0.022
Age		1.025	(1.01 , 1.04)	<0.001
Time-Varying: **HIV Status**	Negative(reference)			
	Positive	0.991	(0.97 1.01)	0.367
Age		0.99	(0.999 1.00)	0.931

The same conclusion on the PH assumption can be made for the SDH model with death as the failure event, HIV Status and Age are not significant, thus failing to reject the null hypothesis of PH assumed. The PH assumption is not violated for this model (p>0.05).

4.5.5 Checking Linearity for Age

To check if the variable age is appropriate in a continuous form, the Martingale's residuals were plotted against age. Figure 9 presents the results for the analysis.

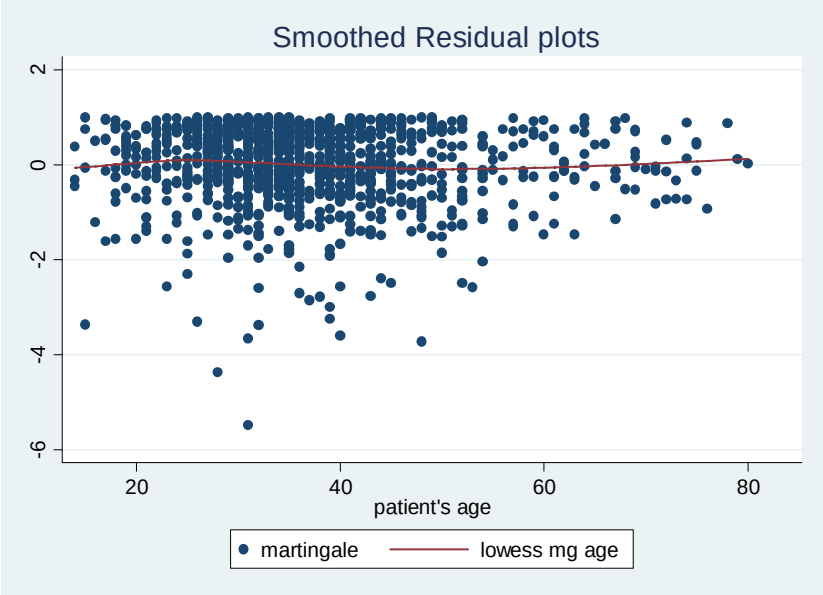


Figure 9: Testing Linearity on variable age.

There was an approximate linearity in the functional form of the covariate age. This indicates the need to transform the covariate Age was minimal. This shows that the log-hazard is slightly linear in age. Therefore, in addition to the un-violated PH assumption, results of age on the cause-specific models were acceptable too.

4.5.6 Goodness of Fit Test

To evaluate the fit of the model the Cox-Snell residuals were used. If the model fits the data well then the true cumulative hazard function conditional on the covariate vector has an exponential distribution with a hazard rate of one. First the Cox CSH models were fitted for failure event discharge and competing event death. The Nelson-Aalen cumulative hazard functions were plotted to compare the hazard functions to the diagonal line. Goodness of fit was determined if the hazard function follows the 45 degrees line, implying that the cumulative hazard was approximately exponential with a hazard rate of one.

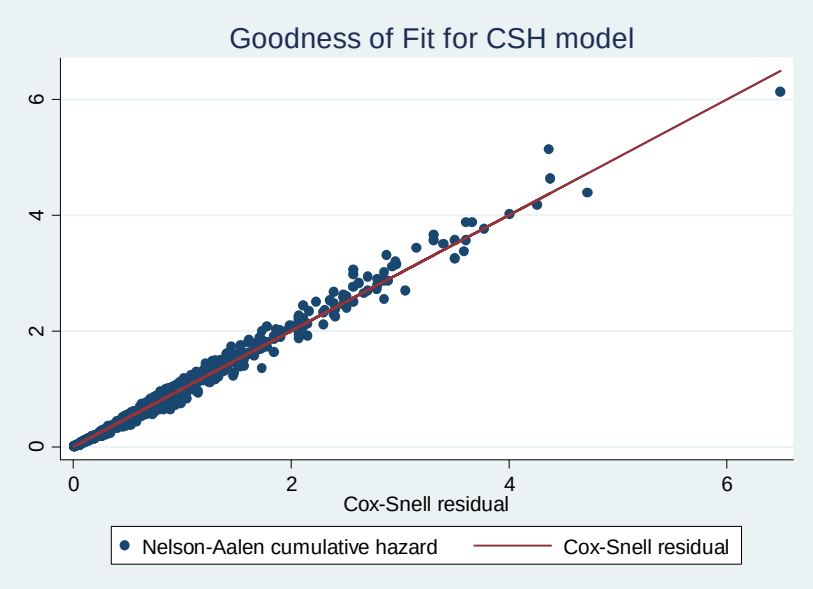


Figure 10: Cox Snell residual plot for event "Discharge"

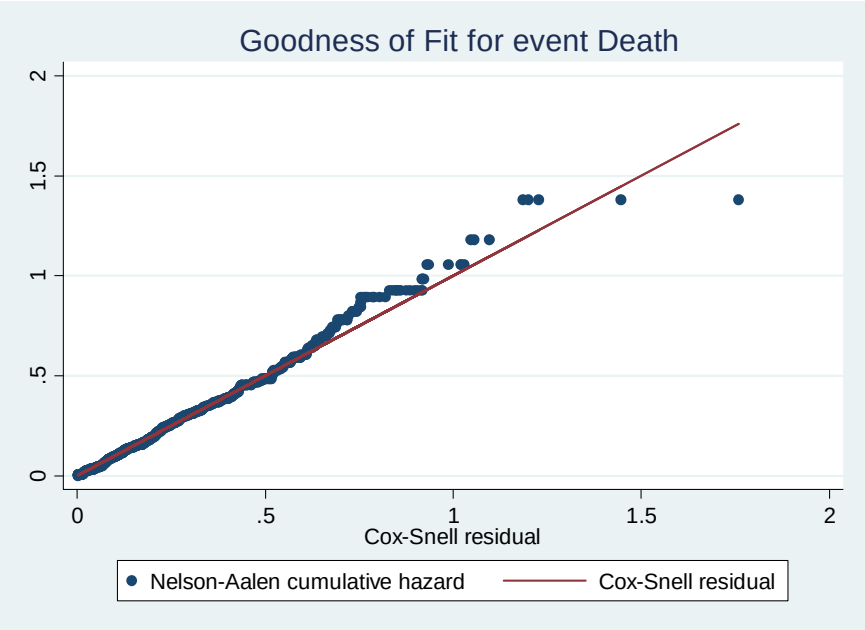


Figure 11: Cox Snell residual plot for competing event "Death"

Figure 10 and 11 shows that the hazard functions follow the 45 degree line very closely except for very large values of time especially for hazard event death. We can conclude that the models fit the data well, that is to say, the model adequately fits or describes the data.

In the following chapter, we the conclusion, recommendations and study limitations are presented.

CHAPTER FIVE

CONCLUSION, RECOMMENDATIONS, LIMITATIONS AND AREA FOR FURTHER RESEARCH

5.1 Conclusions

A comparison of the Cumulative Incidence (CI) and complement of Kaplan-Meier (1-KM) showed that 1-KM produced higher probability of the failure event discharge unlike the Cumulative Incidence function. Since 1-KM considers the competing event death as censored and calculates the probability of discharge without taking into account the effect of the competing event death. It is important to use the cumulative incidence function to obtain survivorship of an event of interest in the presence of competing events.

The study showed that the influence of the covariates on the cause-specific hazard and on the sub-distribution hazard (cumulative incidence) of the event of interest gave different results. Age was the only significant covariate in the CSH model. While the SDH model showed that age and HIV status had a significant effect on the cumulative incidence of discharge. The difference arises since the CSH model looks at the effect of covariates on the event of interest only without regards to how the covariates act on the competing event. Individuals who fail from a death are treated as censored. While this is the case for CSH models the SDH models measure the effect of the covariate on the specific event cumulative probability (Dignam et al., 2012). The estimates, between these two models

were slightly similar showing a moderate dependence between the failure event discharge and the competing event death. The study showed that to clearly observe prognostic factors on time to discharge in the presence of competing events, it is best to use the SDH model because covariate effect would take into account the effect of the competing event on observing the main event.

The study revealed that factors Age and HIV status of a patient were significant predictors in determining hospital stay for TB patients. Older patients and HIV negative patients were more likely to be discharged from hospital within a short stay of hospital admission. The study showed a non-significant effect of ART on length of stay for TB patients. The Univariate SDH model showed that female patients were more likely to be discharged than males. Thus gender was considered an important factor affecting time to discharge.

5.2 Recommendations

- It is important to use competing risk methods when handling data that involves competing events. It is best to use the CIF other than the 1-KM to estimate the survivorship function.

- If independence is observed between the event of interest and the competing event, the Cause-specific hazard model can be used to estimate effect of covariate on the hazard of interest but the SDH model is best to estimate effects of covariates on length of hospital stay. The use of the CSH model and the SDH

model depends on the objectives of the study. Thus it is important to report estimates from both models, since at times both might be informative.

5.3 Area for Further Research

- The study proposes further research on competing risk modelling and handling missing information.
- The study used non-parametric and semi-parametric CR models on the data. It proposes if parametric models are used especially flexible parametric models which enable one to make vast assumptions on the baseline hazard.
- The study proposes an extension of competing risks to repeated time to event data.
- The study proposes a similar area of research but using a prospective study with all necessary variables captured such as treatment status and type, weights; previous TB episode; pulmonary and extra-pulmonary TB; previous opportunistic infection.

5.4 Limitations

The study used a retrospective design, thus had to use any information that was there, thereby some important information might not be available. The data obtained from SPINE did not capture variables such as Treatment, patient height, weight thus limiting explanation of the factors affecting the health outcome. The study incorporated missing values of failure time, HIV status and ART status which might have affected the results of the analysis thereby distorting the real picture of time to discharge. The study analyzed

time to discharge for the TB patients based on their primary diagnosis of TB but did not consider if prolonged stay was affected by a secondary diagnosis unrelated to TB. The data may have had some patients who may have been in the database twice (as recurrent or relapse patients). But the numbers were few considering the fact that the study assessed a short time period. Such patients could not be easily identified as identifiers were removed from the database in order to maintain confidentiality.

REFERENCE

Aban, I. (2014).Time to event analysis in the presence of competing risks.*Journal of Nuclear Cardiology, 22;* 466-7.

Andersen, P. K., Geskus, R. B., de Whitte, T., & Putter, H. (2012). Competing risks in Epidemiology: Possibilities and Pitfalls. *International Journal of Epidemiology, 47*(3), 861-70.

Beuscart, J., Pagniez, D., Boulanger, E., Foy, C. L., &Salleron, J. (2012). Overestimation of the probability of death on peritoneal dialysis by the Kaplan-Meier method: advantages of a competing risks approach. *BMC Neohrology, 13*, 31.

Borrebach, D. J. (2013).*Comparisons between The Kaplan Meier Complement and the cumulative Incidence for survival Prediction in the Presence of Competing events.*Master's Thesis, University of Pittsburgh.

Çelik., Y., Çelik., S. Ş., Bulut., H. D., &Kısa., A. (2001). Inappropriate use of hospital beds: a case study of university hospitals in Turkey. . *World Hospitals and Health Services, 37*(1), 6-13.

Chappell., R. (2012). Competing Risk Analyses: How Are They Different and Why Should You Care. *Clinical Cancer Research, 18*(8), 2127-2129.

Cleves., M., Gutierrez., R. G., Gould., W., &Marchenko., Y. V. (2010). *An Introduction to Survival Analysis Using Stata* (3rded.). Texas.

Collins, T. C., Henderson, W. H., &Khuri, F. S. (1999). Risk Factors for Prolonged Length of Stay After Major Elective Surgery. *Pub Med, 230*(2), 251.

Coviello., V., &Boggess., M. (2004). Cumulative Incidence Estimation In the Presence of Competing Risks.*The Stata Journal, 4*, 103-112.

Cox., D. R. (1972). Regression Models and life-tables.*Journal of Royal Statistical Society Series, B34*(187), 220.

Dehghani., K., Allard., R., Gratton., J., Marcotte., L., &Rivest., P. (2011). Trends in Duration of Hospitalization for Patients with Tuberculosis in Montreal, Canada from 1993 to 2007.*Can J Public Health, 102*(2), 108-111.

Deurden., M. (2009). What are Hazard Ratios? Accessed from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/what_are_haz_ratios.pdf.

Dignam, J. J., Zhang, Q., &Kocherginsky, M. N. (2012).The Use and Interpretation of Competing Risks Regression Models.*Clinical Cancer Research, 18*(8), 2301-2308.

Ferreira., G. M. J., &Alves, A. A. (2013). Factors Associated with Length of Hospital Stay among HIV Positive and HIV Negative Patients with Tuberculosis in Brazil. *PLOS ONE, 8*(4).

Fine., J. P., & Gray., R. J. (1999). A Proportional Hazards Model for the Sub-distribution of Competing Risks.*Journal of American Statistical Association, 94*, 496-509.

Frietas., A., Silva-Costa., T., Lopes., F., Garcia-Lema., L., Teixeira-Pinto., A., Brazdil., P., et al. (2012). Factors Influencing hospital high length of stay outliers.*BMC Health Services Research, 12*, 265.

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*, 695-706.

Gray, R. J. (1988). A Class K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk.*Annals of Stat, 116*(1141), 1154.

Hinchliffe., S. R., Seaton., S. E., Lambert., P. C., & Draper, E. S. (2013). Modeling Time to Death or Discharge in Neonatal Care: An application of Competing Risks.*Pediatric and Perinatal Epidemiology Methodology 27*(4), 426-33.

Johansson E, Long NH, &Diwan V.K. (2000). Gender and Tuberculosis control, Perspectives on health seeking behaviour among men and women in Vietnam. *Health Policy*, 52133-52151.

Jeong., J. H., & Fine., J. P. (2007). Parametric Regression on Cumulative Incidence Function.*Biostatistics 8*(2), 184-196.

Kaplan., E. L., & Meier., P. (1958). Non-Parametric Estimation from Incomplete Observations.*Journal of American Statistical Association, 53,* 457-481.

Kemp., R. J., Mann., G., Simwaka., B. N., Salaniponi., F. M., & Squire., S. B. (2007). Can Malawi's Poor Afford Free Tuberculosis Services? Patient and Household Costs Associated with Tuberculosis in LL. *Bulletin of WHO, 85*(8), 580-585.

Kim., H. T. (2007). Cumulative Incidence in Competing risk data and competing risks Regression analysis. *Clinical Cancer Research 13,* 559-565.

Kleinbaum, D. G., & Klein, M. (2005).*Survival Analysis, A Self-Learning Text* (2nd ed.). New York: Springer Science and Business media

Koller., M. T., Raatz., H., Steyerberg., E. W., &Wolbers., M. (2011). Competing risks and the Clinical Community: Irrelevance or Ignorance? *Stat Med, 31*, 1089-1097.

Latouche., A., Boisson., V., Chevret., S., &Porcher., R. (2007). Mis-specified Regression Model for the Sub-distribution hazard of a competing risk *Statistics in Medicine*(26), 965-974.

Lee, A. H., Ng, A. S. K., &Yau, K. K. W. (2001). Determinants of maternity length of stay: A gamma mixture risk-adjusted model. *Health Care Management Science 4*, 249-255.

Leung, K.-M., Elashoff, R. M., &Afifi, A. A. (1997).Censoring Issues in Survival Analysis.*Annual Review of Public Health, 18*(1), 83-104.

Lim., H. J., Zhang., X., Dyck., R., & Osgood., N. (2010). Methods of Competing Risks Analysis of end-stage Renal disease and Mortality among people with diabetes. *BMC Medical Research Methodology 10*, 97

Marengoni, A., Winblad, B., A, K., & L, F. (2008).Prevalence of Chronic diseases and multimorbidity among the elderly people in Sweden. . *Am J Public Health, 98*(1198).

Nesher., L., Riesenberg., K., Saidel-Odes., L., Schlaeffer., F., &Smolyakov., R. (2012). Tuberculosis in African Refugees from the Eastern Sub-Sahara Region.*IMAJ, 14*, 111-114.

Nyirenda., T. E., Harries., A. D., &Gausi., F. (2003). Decentralisation of TB Services in an Urban Setting, Lilongwe, Malawi.*International Journal Tubere Lung Disease, 7*, S21-28.

Oliveira., H. M. M. G., Brito., R. C., Kritski., A. L., &Ruffino-Netto., A. (2009). Epidemiological profile of Hospitalized Patients with TB at a referral hospital in the city of Rio de Janeiro.*Brazil. J Bras Pneumol*(35), 780-787.

Putter., H., Fiocco., M., &Geskus., R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Stat Med, 26*, 2389-2430.

Roberts, M. S., & Daley, K. (2003).A national study of clinical and laboratory factors affecting the survival of patients with multiple drug resistant tuberculosis in the UK.*Thorax*, *9*(57), 810-816.

Sherif, B. N. (2007).*A Comparison of Kaplan-Meier and Cumulative Incidence Estimate in the Presence or Absence of Competing Risks in Breast Cancer Data*.Pittsburgh,University of Pittsburgh.

Stolp., S. M., Huson., M. A. M., Janssen., S., Beyeme., J. O., &Grobusch., M. P. (2013). Tuberculosis patients hospitalized in the Albert Schweitzer Hospital, Lambarene, Gabon—a retrospective observational study. *ClinMicrobiol Infect 19*, E499–E501.

Tamiru, M., &Haidar, J. (2010).Hospital Bed Occupancy and HIV/AIDS in three Major Public Hospitals of Addis Ababa, Ethiopia.*International Journal of Biomedical Science, 6*(3), 195-201.

Teixeira., L., Rodrigue., A., Carvalho., M. J., Cabriata., A., &Mendonca., D. (2013). Modelling Competing Risks in Nephrology Research: An Example in Peritoneal Dialysis.*BMC Neohrology, 14*(110).

Therneau., T. M., &Grambsch., P. M. (2000). *Modeling Survival Data: Extending the Cox Model.*New York; Springer.

Tweya H, Feldacker C, Phiri S, Ben-Smith A, & L, F. (2013).Comparison of Treatment Outcomes of New Smear-Positive Pulmonary Tuberculosis Patients by HIV and Antiretroviral Status in a TB/HIV Clinic, Malawi.*PLoS ONE, 8*(2).

WHO.(2010). *Global Tuberculosis Report 2010*. Geneva. WHO

WHO.(2010). *Global Tuberculosis Report 2010*. Geneva. WHO

WHO.(2012). *Global tuberculosis report 2012*. Geneva. WHO

WHO.(2013). *Global Tuberculosis Report 2013*. Geneva. WHO

Yau, K. K. W., Lee, A. H., & Ng, A. S. K. (2003). Finite mixture regression model with random effects: Application to neonatal hospital length of stay. *Computational Statistics & Data Analysis, 41*, 359-366.

Zetola., N. M., Macesic., N., Modongo., C., Shin., S., Ncube., R., & Ronald., G. (2014). Longer Hospital Stay is Associated with Higher rates of Tuberculosis-related Morbidity and Moratlity within 12 months after discharge in a referral hospital in sub-saharan Africa. *BMC Infectious Diseases, 14*, 409.

APPENDICES

Appendix 1: Analysis Stata Commands

```
set more off

cd "E:\School\thesis\TB data"

capture log using New TB data.log

use "E:\School\thesis\TB data\New TB data.dta", clear

***Descriptive Statistics***

graph box ftime, over(Gender1) over(HIVSTATUS) asyvars

tabprimary_diagnosis

tabsecondary_diagnosis

tabstat age, by(Gender1)

tabstatdftime, by(HIVSTATUS)


***Cumulative Incidence Curves***

Stsetftime, fail(failtype=1)

stcompetcif = ci stderr=se upper=hi lower=lo, compet1(0) by
(HIVSTATUS)

gencif_HIV_Negative = cif if failtype==1 & HIVSTATUS ==0

gencif_HIV_positive = cif if failtype==1 & HIVSTATUS ==1

twoway line cif_HIV_* _t, connect(J J) sort ytitle(Cumulative
Incidence) xtitle(Analysis time in days) lpattern(shortdash)

stpepemori HIVSTATUS, compet(0)


stcompet CI = ci Stderr=se Upper=hi Lower=lo, compet1(0) by
(ARTSTATUS)

gencif_ART_No = cif if failtype==1 & ARTSTATUS ==0

gencif_ART_Yes = cif if failtype==1 & ARTSTATUS ==1
```

```
twoway line cif_ART_* _t, connect(J J) sort ytitle(Cumulative
Incidence)xtitle(Analysis time in Days) lpattern(solid shortdash
dot)

stpepemori ARTSTATUS, compet(0)


stcompet ci = ci STderr=se UPper=hi LOwer=lo, compet1(0) by
(Gender1)

gencif_Gender_male = cif if failtype==1 & Gender1==0

gencif_Gender_female = cif if failtype==1 & Gender1==1

twoway line cif_Gender_* _t, connect(J J) sort title(Cumulative
Incidence by Sex) ytitle(Cumulative Incidence) xtitle(Analysis
time in Days) lpattern(solid longdash) legend(lab(1 "Male") lab(2
"Female") stack)

stpepemori Gender1, compet(0)

stset, clear


*********Comparison of 1-KM vs CI*************
stsetftime, fail(failtype=1)

sts gen KM=s

gen Complement = 1-KM

stcompetCumInc=ci, compet1(0)

gen CI=CumInc if failtype==1

twoway line Complement CI ftime, ytitle("Probability")
lpattern(dash longdash) lcolor(green orange)

stset, clear

*****Standard Cox PH model******

stsetftime, fail(failtype)

xi:stcox age

xi:stcoxi.Gender

xi:stcoxi.HIVSTATUS

xi:stcox age i.Gender1 i.HIVSTATUS

estatphtest, detail

stset, clear
```

```
******Cause-Specific Hazards Model for event Discharge******
Stsetftime, fail(failtype==1)
xi:stcox age
xi:stcoxi.Gender
xi:stcoxi.HIVSTATUS
xi:stcox age i.Gender1 i.HIVSTATUS
********Test for PH assumption*********
estatphtest, detail
stphtest, plot(HIVSTATUS)
stphtest, plot(Gender1)
********model selection criterion*********
estatic
stset, clear
******* Cause-Specific Hazards Model for event Death********
Stsetftime, fail(failtype==0)
xi:stcox age
xi:stcoxi.Gender
xi:stcoxi.HIVSTATUS
xi:stcoxi.ARTSTATUS
xi:stcox age i.Gender1 i.HIVSTATUS
********Test for PH assumption*********
estatphtest, detail
stphtest, plot(HIVSTATUS)
stphtest, plot(Gender1)
**********model selection Criteria*************
estatic
stset, clear
****Fine and Gray model for event discharge and competing event
death****
```

```
stsetftime, fail(failtype==1)

xi:stcrreg  age, compete(failtype==0)

xi:stcrreg  i.Gender1, compete(failtype==0)

xi:stcrregi.HIVSTATUS, compete(failtype==0)

xi:stcrreg  age i.Gender1i.HIVSTATUS, compete(failtype==0)

********Test for PH assumption*********

xi:stcrreg  age i.Gender1 i.HIVSTATUS,
compete(failtype==0)tvc(age HIVSTATUS)

estatic //model selection criteria

stset, clear

**** Fine and Gray model for event death and competing event
discharge****

stsetftime, fail(failtype==0)

xi:stcrreg  age, compete(failtype==1)

xi:stcrreg  i.Gender1, compete(failtype==1)

xi:stcrregi.HIVSTATUS, compete(failtype==1)

xi:stcrreg  age i.Gender1 i.HIVSTATUS, compete(failtype==1)

xi:stcrreg  age i.Gender1 i.HIVSTATUS,
compete(failtype==1)tvc(age HIVSTATUS) //Test for PH assumption

estatic // model selection criteria

stset, clear

************* Model Diagnostics***************

 ********************Goodness of Fit test*****************

******for event discharge*******

stsetftime, fail(failtype==1)

xi: stcox age i.Gender1 i.HIVSTATUS, mgale(mg)

predictcoxsn, csnell

stsetcoxsn, fail(failtype==0)

sts generate H=na

twoway (scatter coxsn H) (line coxsncoxsn)
```

```
stset, clear

******for event death***********
drop mg
dropcoxsn
stsetftime, fail(failtype==0)
xi: stcox age i.Gender1 i.HIVSTATUS, mgale(mg)
predictcoxsn, csnell
stsetcoxsn, fail(failtype==1)
sts generate H=na
twoway (scatter coxsn H) (line coxsncoxsn)
stset, clear
********Checking Linearity for Age*********
stsetftime, fail(failtype==1)
xi: stcoxi.HIVSTATUS i.Gender1, mgale(mg)
twoway (scatter mg age) (lowess mg age)
stset, clear
```

Appendix 2: Summaries of Primary Diagnosis

Primary diagnosis	Frequency	Percent
?ART FAILURE ?ABDO TB	3	0.25
?TB	2	0.16
ANAEMIA;UNKNOWN TB	1	0.08
ASPIRATION TUBERCULOSIS	1	0.08
DRUG RESISTANT TB	2	0.16
EPTB	4	0.33
EPTB RELAPSE	2	0.16
MENINGITIS SUB-ACUTE PRESUMED TUBERCULO	13	1.07
PERIPHERAL NEUROPATHY DRUG RELATED – TB	1	0.08
PLEURAL EFFUSION FPTB	1	0.08
PLEURAL EFFUSION TB	3	0.25
PNEUMONIA PTB	1	0.08
PNEUMONIA TB	1	0.08
PTB RELAPSE	1	0.08
RENAL TUBERCULOSIS	2	0.16
SEPSIS TB	1	0.08
TB	2	0.16
TB SPINE	1	0.08
TUBERCULOSIS	38	3.11
TUBERCULOSIS ?DISSEMINATED TB	1	0.08
TUBERCULOSIS ?TBM	1	0.08
TUBERCULOSIS ABDOMINAL	15	1.23
TUBERCULOSIS ADENTIS	1	0.08
TUBERCULOSIS ASCITES	3	0.25
TUBERCULOSIS ASCITIS	1	0.08
TUBERCULOSIS BONE TUMOUR	1	0.08
TUBERCULOSIS DISSEMINATED	317	25.98
TUBERCULOSIS E P T B	13	1.07
TUBERCULOSIS EPTB	67	5.49
TUBERCULOSIS EPTB RELAPSE	2	0.16
TUBERCULOSIS EPTB PIEURAL EFFUSION	1	0.08

TUBERCULOSIS EXTRAL PULMONALY	1	0.08
TUBERCULOSIS INFECTIVE	1	0.08
TUBERCULOSIS IRIS	1	0.08
TUBERCULOSIS LYMPHADENTIS	1	0.08
TUBERCULOSIS M D R	1	0.08
TUBERCULOSIS MACROCYTIA ANAEMIA	1	0.08
TUBERCULOSIS MALIGNANCY	1	0.08
TUBERCULOSIS MARIGNANCY	1	0.08
TUBERCULOSIS MDR	3	0.25
TUBERCULOSIS MELIGNANCY	1	0.08
TUBERCULOSIS MENINGITIS	4	0.33
TUBERCULOSIS MILIARY	63	5.16
TUBERCULOSIS OTHER	9	0.74
TUBERCULOSIS PENCORDINAAL EFFUSION	1	0.08
TUBERCULOSIS PERICARDIAL	1	0.08
TUBERCULOSIS PERICARDIAL EFFUSSION	1	0.08
TUBERCULOSIS PERICARDIAL EFFISSION	1	0.08
TUBERCULOSIS PERICARDIAL EFFUSSION	2	0.16
TUBERCULOSIS PERICARDIAL EFFUSION	2	0.16
TUBERCULOSIS PIEURAL EFFUSION	1	0.08
TUBERCULOSIS PLEURAL EFFUSION	8	0.66
TUBERCULOSIS PLEURAL EFFUSSION	5	0.41
TUBERCULOSIS PLEURAL EFFUSION	1	0.08
TUBERCULOSIS PLEURAL EFFUSIONS	2	0.16
TUBERCULOSIS PLEURAL EFFUSSSION	1	0.08
TUBERCULOSIS PTB	1	0.08
TUBERCULOSIS PTB RELAPSE	14	1.15
TUBERCULOSIS PTB RELAPSE	2	0.16
TUBERCULOSIS PULMONARY	469	38.44
TUBERCULOSIS PULMONARY RELAPSE	1	0.08
TUBERCULOSIS RECURENT	1	0.08
TUBERCULOSIS RELAPSE	6	0.49
TUBERCULOSIS RELAPSE PTB	2	0.16
TUBERCULOSIS RELAPSE PTB SMEAR P	1	0.08
TUBERCULOSIS SMEAR NEG	1	0.08
TUBERCULOSIS SMEAR NEGATIVE	1	0.08
TUBERCULOSIS SPINAL	17	1.39

TUBERCULOSIS TB CXR	1	0.08
TUBERCULOSIS TUBERCULOUS MENINGITIS	83	6.8
TUBERCULOSIS TUBERCULOUS PERICARDITIS	3	0.25
Total	**1,220**	**100**

Appendix 3: Summaries of Patient's Secondary Diagnosis

Secondary diagnosis	Frequency	Percent
TUBERCULOSIS	1,186	89.58
ANAEMIA	1	0.08
ANAEMIA MACROCYTIC IRON DEFICIENCY	1	0.08
ANAEMIA MICROCYTIC CHRONIC DISEASE	2	0.15
ANAEMIA NORMOCYTIC CHRONIC DISEASE	2	0.15
ANAEMIA UNKNOWN CHRONIC DISEASE	3	0.23
ANAEMIA;NORMOCYTIC OTHER	1	0.08
ANAEMIA;UNKNOWN	1	0.08
ANAEMIA;UNKNOWN OTHER	1	0.08
ANAEMIA;UNKNOWN PANCYTOPEMIA	3	0.23
ANAEMIA;UNKNOWN SEVERE	2	0.15
ANGINA	3	0.23
ART FAILURE	3	0.23
CANCER UNKNOWN PRIMARY	3	0.23
CANDIDIASIS OESOPHAGEAL	3	0.23
CANDIDIASIS ORAL	3	0.23
CAP	2	0.15
CHEST INFECTION	1	0.08
CIRRHOSIS	4	0.30
DIABETES MELLITUS HYPOGLYCAEMIA	1	0.08
GASTROENTERITIS ACUTE	4	0.30
GASTROENTERITIS CHRONIC	2	0.15
HEART FAILURE CONGESTIVE CARDIAC FAILUR	1	0.08
HTN	1	0.08
HYDROPNEUMOTHORAX	1	0.08
HYPERTENSION OTHER	2	0.15
KAPOSI'S SARCOMA CUTANEOUS	3	0.23
LUNG CANCER OTHER	1	0.08
LYMPHADENOPATHY	1	0.08
MALARIA CEREBRAL	2	0.15
MALARIA UNCOMPLICATED	2	0.15

MALIGNANCY	1	0.08
MALNUTRITION	1	0.08
MENINGITIS BACTERIAL CLINICAL	3	0.23
MENINGITIS CRYPTOCOCCAL	1	0.08
MENINGITIS;BACTERIAL OTHER	1	0.08
PANCREATITIS OTHER	3	0.23
PCP	1	0.08
PHARYNGITIS	1	0.08
PID	1	0.08
PLEURAL EFFUSION PRESUMED DUE TO KAPOSI	1	0.08
PNEUMONIA	3	0.23
PNEUMONIA ASPIRATION	2	0.15
PNEUMONIA BRONCHOPNEUMONIA	2	0.15
PNEUMONIA CAP	3	0.23
PNEUMONIA COMMUNITY ACQUIRED	2	0.15
PNEUMONIA LOBAR	2	0.15
PNEUMONIA OTHER	2	0.15
PULMONARY EFFUSION	2	0.15
SCHISTOSOMIASIS OTHER	1	0.08
SCHIZOAFFECTIVE DISORDER	3	0.23
SEPSIS	2	0.15
SEPSIS NTS ISOLATED	12	0.91
SEPSIS OTHER	2	0.15
SEPSIS S PNEUMONIAE ISOLATED	1	0.08
SEPSIS TYPHOID	1	0.08
SEVERE IMMUNOSUPRESION	3	0.23
TB ADENITIS	1	0.08
TB BACTERAEMIA	1	0.08
TB IRIS	3	0.23
TUBERCULOSIS DISSEMINATED	2	0.15
TUBERCULOSIS EPTB	1	0.08
TUBERCULOSIS PULMONARY	5	0.38
TUBERCULOSIS RELAPSE	2	0.15
TUBERCULOSIS RELEPSE	1	0.08
TYPHOID	4	0.30
ULCER GASTRIC	1	0.08
Total	1,324	100.00

Appendix 4: Certificate of Ethical Approval

(See next page)