FLEXIBLE PARAMETRIC PROGNOSTIC RISK SCORE MODELS FOR SHORT-TERM RISK OF AIDS ON PATIENTS NEWLY INITIATED ON HIGHLY ACTIVE ANTIRETROVIRAL THERAPY (HAART) IN NYANZA, KENYA

BY

KEVIN OWUOR

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN QUANTITATIVE RESEARCH METHODS

SCHOOL OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE

Maseno University

Declaration

This thesis is my own work and has not been presented for a degree award in any other institution.

KEVIN OWUOR
EL/PG/MSC/0012/2011

Signature: ..... Date: .....

This thesis has been submitted for examination with our approval as the university supervisors.

Dr. Edgar Ouko Otumba, PhD
Department of Statistics and Actuarial Sciences,
Maseno University

Signature: ..... Date: .....

Dr. Starley B. Shade, PhD
AIDS Research Institute,
Center for AIDS Prevention Studies,
University of California, San Francisco

Signature: ..... Date: .....

Acknowledgments

Special thanks to Jayne Kulzer, Charles Kibaara, Cinthia Blat, Dr. Patrick Oyaro , Dr. Elizabeth Bukusi, Dr. Craig Cohen and the entire Kenya Medical Research Institute, Family AIDS Care and Education Services and University of California, San Francisco for availing and allowing the use of their invaluable data for the thesis. I would to thank my supervisors Dr. Edgar Ouko Otumba (Department of Statistics and Actuarial Sciences – Maseno University) and Dr. Starley B. Shade (AIDS Research Institute, Center for AIDS Prevention Studies, University of California, San Francisco) for their insight, support and dedication throughout this research work.

Dedication

To my late father, Lawrence Ojiambo and mother Lucy Ojiambo for the belief and support towards my education

Abstract

About 22.9 million people living with HIV/AIDS reside in sub-Saharan Africa, many of whom have progressed to AIDS over time. Kenya has high numbers of new infections; a total of 104,000 in the general population with paediatrics at 13,000 and adults at 91,000. Risk scores constructed using prognostic factors may be valuable in the early identification and intervention to patients at risk of progression to AIDS. There was therefore a necessity to come up with robust risk models that use a limited number of easily available factors. The main objective was to come up with a risk score utilizing routine care data which can be easily applicable in a clinical setting to asses for risk of AIDS among HIV infected patients. It was a prospective cohort study done using 2 year follow-up (initiated on Highly Active Antiretroviral Therapy (HAART)) between 1^{st} of June, 2010 and 30^{th} of May, 2011) data from 1454 HIV/AIDS on ART care and treatment. Age, sex, marital status, CD4 cell count, haemoglobin level, BMI, prior TB medication and whether or not patients were currently receiving any ART was modelled to describe the short term risk of new AIDS event. Flexible parametric survival regression analysis (Royston Parmar) was used instead of Cox-PH regression. Strong predictors of progression were Body Mass Index, haemoglobin, World Health Organization staging and Tuberculosis treatment prior to HAART initiation. The study was able to develop a two group risk categorization based on the risk model developed. The discriminative ability of the model was moderately strong (Harrell's c-index of 0.69). The rate of progression to AIDS between the high and low risk groups was well defined. The rate of progression was 0.38 and 0.93 per thousand person-years of followup for the low risk and high risk groups respectively which was more than twofold risk of progression to AIDS among high risk group, (HR= 2.47 95% CI: 1.66 - 3.69; p <0.001). This study developed a prognostic risk score model to be utilised in the early identification and intervention to patients at risk of progression to AIDS.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background to the study

The human immunodeficiency virus (HIV) destroys or impairs the immune system leading to a deterioration of the well being of the immune system while Acquired immunodeficiency syndrome (AIDS) refers to the most advanced stages of HIV infection. Majority of the newly infected people, about 22.9 million people living with HIV/AIDS (PLWHA) reside in sub-Saharan Africa accounting for about 68 percent of the total number of people living with HIV/AIDS (PLWHA) worldwide. There were about 1.9 million new infections in 2010 alone signifying a drop from 2.6 million new infections in the previous year. About 1.2 million deaths of adults and children also occurred in 2010 worldwide [21]. Kenya has high numbers of new infections, a total of 104,000 in the general population with paediatrics at 13,000 and adults at 91,000. The HIV and AIDS-related deaths stand at about 57,000 in 2011 [20]. The high risk of progression to AIDS among HIV infected patients on HAART is an issue many care and treatment programs in resource limited settings especially in Sub Saharan Africa are grappling with currently. This in turn has a great negative impact on important national aspects of many African nations which includes economy and business, education and knowledge, health and social welfare. Globally, 56 percent of children orphaned by AIDS live in six countries in sub-Saharan Africa namely Nigeria (2.5 million), South Africa (1.9 million), Kenya (1.2 million), Uganda (1.2 million), Tanzania (1.3 million), and Zimbabwe (1 million). The HIV/AIDS epidemic is setting back decades of progress in increasing the life expectancy of the people of sub-Saharan Africa. The vast majority of people in Africa who have HIV/AIDS are between

the ages of 15 and 49, and millions of adults are dying young or in early middle age [45]. Sub Saharan Africa countries are the most affected by this epidemic, on all components of human development, as shown by the national human development reports from Zambia, Mozambique, Kenya, Zimbabwe, South Africa and Botswana [46, 36, 35, 34, 33, 37, 31].

A number of studies carried out in sub-Saharan Africa have indicated high early mortality among patients starting antiretroviral treatment in sub-Saharan Africa [42]. The major impediment is late diagnosis of HIV/AIDS hence late start in ART care due to poor prognosis systems; yet ART care and treatment is readily available [26].There is a great focus and commitment by many countries to the sixth Millennium Development Goal (MDG), which is combating HIV/AIDS, malaria and other diseases. These efforts need to be enhanced, however there are challenges such as shortage of health care providers in sub-Saharan Africa: provider ratios fall significantly below World Health Organization recommendations. In the year 2012, an estimated 2.2 physicians and 9.0 nursing personnel served every 10,000 people in the region. This figure is way below the worldwide average of 16 physicians and 36 nursing personnel per 10,000 [32]. This in turn impedes the quality of care by giving inadequate time to each HIV/AIDS infected patient. There is therefore needed an enhanced system for HIV/AIDS care and treatment delivery. Risk scores have increasingly been used to quantify the amount of risk that an individual is predisposed to given their values on a set of prognostic factors. They are useful because they are derived from models that take into account the contribution of each of the risk factors jointly [17, 1].

There are several studies previously done which have identified a number of predictors of HIV progression to AIDS. They include haemoglobin level, cluster of differentiation 4 (CD4) cell count, body mass index (BMI), previous AIDS-defining illness or condition and age [16, 9, 2, 12, 39, 40, 4, 3, 13]. These prognostic factors can be useful in the early identification and intervention to patients at risk of progression to AIDS. The risk models would be crucial in the short term prognosis of early signs and symptoms of progression to HIV/AIDS indicators which in turn could help in timely interventions to avert further disease burden. The simplicity of the models is of essence; they should be able to use minimal laboratory tests and patient demographic characteristics [16]. Taking into consideration the high volume of patients to service provider ratio in Sub-Saharan

Africa the simplicity of application would come in handy to the health care providers. The efforts to develop easy to use risk models have been stepped up in various parts of the world [16, 21, 22, 43]; but none in particular to Kenya, East Africa and the greater Sub-Saharan Africa. There was therefore a necessity to come up with robust risk models that use a limited number of easily available factors but have equally high predictive power to assist in timely prognosis of HIV patients at risk of AIDS. Most of the previous models have utilized other statistical methods among them Poisson and proportional hazards Cox regression, to develop the risk models. However, flexible parametric survival models are able to accurately estimate the baseline hazard function unlike in Cox model which does not [27, 30].

1.2 Basic Concepts

HIV

The human immunodeficiency virus (HIV) infects cells of the immune system, destroying or impairing their function. Infection with the virus results in the progressive deterioration of the immune system, leading to "immune deficiency". The immune system is considered deficient when it can no longer fulfil its role of fighting infection and disease. HIV can be transmitted through unprotected sexual intercourse (vaginal or anal), and oral sex with an infected person; transfusion of contaminated blood; and the sharing of contaminated needles, syringes or other sharp instruments. It may also be transmitted between a mother and her infant during pregnancy, childbirth and breastfeeding.

AIDS

Acquired Immunodeficiency Syndrome (AIDS) is a term which applies to the most advanced stages of HIV infection. It is defined by the occurrence of any of more than 20 opportunistic infections or HIV-related cancers. It may also be defined medically as having a cluster of differentiation 4 (CD4) count <200 cells/ml or World Health Organization (WHO) stage III/IV.

ART

It refers to any treatment that suppresses or stops a retrovirus such as HIV virus that causes AIDS.

HAART

Highly Active Antiretroviral Therapy (HAART) is the term given to aggressive treatment regimens used to suppress HIV viral replication and the progression of HIV disease. Usually comes in a combination of three or more drugs. They work together to lower the amount of active virus in the body for better outcomes in HIV/AIDS patients.

CD4

Known as cluster of differentiation 4, they are the cells that help initiate the body's fight against disease. It is assessed by a blood test called the CD4 count which measures the number of functioning CD4 cells in the body and therefore measures the health of the immune system. The fewer functioning CD4 cells, the weaker the immune system and therefore the more vulnerable a person is to infections and illnesses.

WHO Stage

This a 4 stage HIV disease progression classification system developed by the World Health Organization (WHO). The four distinct stages are: primary infection (stage I), clinically asymptomatic stage (stage II), symptomatic HIV infection (Stage III), and progression from HIV to AIDS (Stage IV). There are certain illnesses and opportunistic infections that characterize each stage.

BMI

Body mass index is the individual's body mass (kg) divided by the square of his or her height (m). The formula is used in medicine to produce a unit of measure expressed in

(kg/m^2).

1.3 Statement of the problem

The intention of study therefore was to come up with such a robust prognostic system that would be used by various stakeholders among them clinicians, doctors and health care service providers to immediately assess the risk of progression to AIDS among HIV patients attending routine treatment and care at clinics and hospitals. This study therefore aimed to model easily available demographic, clinical and laboratory variables using flexible parametric survival models which do not have the restrictive assumption of proportional hazards for covariate effects and has advantages for prediction, extrapolation, quantification (e.g., absolute and relative difference in risk) and modelling time-dependent effects to come up with the prognostic risk score model with good predictive power and which would be easily translated into a simple risk score system.

1.4 Main and Specific Objectives of the study

The main objective of this study was to come up with a risk score utilizing routine care data which can be easily applicable in a clinical setting to asses for risk of progression to AIDS among HIV infected patients.

The specific objectives included:

1. To identify baseline patient characteristics that are associated with progression to AIDS among patients newly initiated on HAART in Nyanza, Kenya
2. To build a multivariable short term risk model for progression to AIDS in Nyanza, Kenya
3. To refine and test the predictive power of the risk model

1.5 Significance of the study

Deaths or morbidity due to HIV/AIDS is a big problem that sub-Saharan Africa countries have to deal with. In the long run this negatively impacts on various aspects of national economic and social development. A number of prognostic models in different disease areas have been developed to enable the prognosis of patients to be more accurately assessed and appropriate treatment and risk management given. There is a great need to assess the prognostic importance in assessment of the outcomes for each covariate on its own as well as in a multivariable model. As ART drugs become more available in places such as sub-Saharan Africa, these risk models may also be used to evaluate the effectiveness of ART care and treatment programs in such resource limited settings [21]. The development of these models to be easy and readily used by clinicians and other health care providers would in turn positively impact the delivery of timely quality of care and treatment to HIV/AIDS patients. The development of these models is also relevant to the continuing discussion on better ways of HIV/AIDS care and management.

Chapter 2

Literature Review

2.1 Introduction

This chapter focuses on the critical aspects of the existing literature in relation to this study while covering studies on predictors of progression to AIDS and risk scores development.

2.2 Studies on predictors of progression to AIDS

There are a number of studies that have been done which highlight some laboratory and clinical predictors of progression to AIDS. From the TAHOD (TREAT Asia HIV Observational Database) a collaborative observational cohort study involving 17 sites in the Asia-Pacific region Srasuebkul et al. [29] found that in both their clinical and CD4 cell count Poisson risk models,those at high risk of AIDS or death were patients with severe anaemia, mild anaemia, male gender,age above 40 years, CD4 cell count <50 cells/mL, 51–200 cells/mL and BMI$\leq$ 18. Byakwaga et al.[13] also pointed out that a lower CD4 count at 6 months and hemoglobin level less than 80 mg/dL compared to hemoglobin level greater than 130 mg/dL from the start of cART was significantly associated with a greater risk of disease progression [13].

In a study comparing mortality rates observed in HIV-1 infected patients starting ART with non HIV-related background mortality in four countries in sub Saharan Africa

from the International epidemiological Databases to Evaluate AIDS (IeDEA) Brinkhof et al. [23] found that longer time period on ART indicated a risk reduction of 90% (eHR(estimated Hazard Ratio) 0.10). Females were at 16% (eHR 0.84) lower risk of excess mortality than males. There was strong evidence for a decline in excess risk with increasing baseline CD4 count: patients with a baseline CD4 count of 200 cells/ml or more experienced an 81% (eHR 0.19) reduction in risk over 2 years as compared to patients with a baseline CD4 count of less than 25 cells/ml (eHR 0.19). There was also a reduction in the excess risk by 72% (eHR 0.28) in patients with less advanced baseline disease (WHO stage I/II) compared to those with advanced baseline disease (WHO stage III/IV). This study however does not find sufficient evidence for an association between patient's age, treatment regimen or calendar period and risk of death. This study also did not look at progression to AIDS but only death as the outcome of interest.

Significant risk factors associated with death of HIV/AIDS patients according to Zachariah et al. [40] from a study in a rural district hospital in Malawi include WHO stage IV, a baseline CD4 cell count < 50 cells/ml and low BMI (values < 16.0 kg/m^2). They noted that these factors are important as screening criteria for individuals starting ART who are at a high risk of death due to AIDS complications. However, this study's focus was on death outcome only in the first 3 to 6 months of starting ART and not progression to AIDS.

Malvy et al. [9] lay emphasis on the importance of age and CD4 count as predictors of the short term risk of AIDS. They show there is a clear relationship between increasing risk with increasing age. However Babiker et al. [3] found that the age effect seen on HAART treatment may be seemingly closer to the natural effect of aging rather than the pre-treatment, HIV-related increase in mortality, suggesting that HAART modifies the effect of age at seroconversion on HIV disease progression.

There are no such related studies or literature with a specific focus on predictors of progression to AIDS that have been done in Nyanza, Kenya , East Africa or sub Saharan Africa as a whole and in African populations.

2.3 Studies on Risk scores development for progression to AIDS

This section has a review of studies that developed risk scores related to AIDS outcome. The study by Srasuebkul et al. [29] had risk of AIDS or death classified into low, high, and very high risk groups. The model was able to discriminate very well patients at high risk, however not quite for low and moderate risk patients. This model was constructed for and using data from patients in Asian populations to identify high short-term risk of AIDS or death. The risk model may be generalized, however it may not be suitably adapted for other populations e.g. patients in sub Saharan Africa. Also of note is that the model was neither fit on a training dataset nor validated on a separate independent dataset. It is well known that fitting and validating models on a single dataset can lead to over-optimistic estimates of predictive value [27, 30]. They also never used formal bootstrap approaches as an alternative. Poisson regression is fully parametric and assumes a particular form for the baseline hazard; that it is constant over time . Srasuebkul et al.[29], used a Poisson regression approach to determine factors associated with the short-term risk of clinical progression. Brinkhof et al.[23] used generalized linear models with a Poisson error structure to model the excess mortality and SMRs (Standardized Mortality Ratios). Majority of the studies have utilized Poisson regression and more so Cox proportional hazards regression. However parametric survival analysis specifically flexible parametric models seems to be a more robust method due the flexibility and also its proper handling of the baseline hazard while predicting the risk scores. The use of flexible parametric (Royston Parmar) models could impact very positively on the ability of researchers to accurately predict survival. It may even be used in the validation of earlier published models for which the original data is unavailable [38, 5].

There are no such related published studies or literature with a specific focus on risk scores development for progression to AIDS that have been done in Nyanza, Kenya, East Africa or sub Saharan Africa as a whole and in African populations.

Chapter 3

Research Methodology

3.1 Study Design, Population and Sample

This was a prospective cohort study done among HIV/AIDS patients (clients) on ART care and treatment at four KEMRI-FACES-UCSF (Kenya Medical Research Institute - Family AIDS Care and Education Services - University of California, San Francisco) clinics from over 50 sites spread across Kisumu, Suba, Migori, Nyatike and Rongo Districts of Nyanza province, Kenya. The study population consisted of patients who enrolled into care from 23^{rd} of July 2007 to 30^{th} of May 2011 and were initiated on HAART treatment between 1^{st} of June, 2010 and 30^{th} of May, 2011 and followed prospectively for two years up to 1^{st} June , 2013. The study cohort consisted of patients aged 15 years and above who were previously ART naive. There were 1454 clients who met the inclusion criteria from a total population of about 10,000 clients on ART care and treatment within the study period.

3.2 Study Outcomes

The primary outcome was time to progress to AIDS. We defined AIDS as a composite of clinical or immunologic progression to AIDS in the context of this study. Clinical progression was defined as a new WHO stage III or IV or, if not correctly re-staged, presence of any AIDS defining condition such as Tuberculosis after starting HAART

while Immunologic progression was defined as CD4 count level at <200 cells/mL within the 2 year follow-up period. Some of the patients may or may not have had a diagnosis of AIDS prior to start of therapy. Death occurred if the patient died within the 2 year follow-up period. If no progression to AIDS or death occurred by the date of the last clinic visit in the 2 year period it was considered the censoring date. This study was a 2 year AIDS-free survival from initiation of HAART. Clinical and laboratory measurement taken 3 months prior or within a week to ART initiation were considered as valid baseline measurements.

3.3 Covariates

Baseline variables included age, sex, marital status, history of prior ART use, HAART initiation date, Prior TB medication (before initiation), baseline WHO staging (at HAART initiation), follow-up WHO staging and dates, baseline CD4 (at HAART initiation), follow-up CD4 cell counts and dates, AIDS diagnosis before HAART and date (determined by WHO staging and follow-up CD4 cell count drop to <200 cells/ml) and death (date of death).

Age, sex, marital status, CD4 cell count, haemoglobin level, BMI, prior TB medication and whether or not patients were currently receiving any ART were modeled as time updated values used to describe the short term risk of new AIDS event. Continuous variables, such as age or CD4 cell count, were categorized *a priori* with use of commonly used cutoff values to ensure roughly equal numbers of events within each category. Furthermore effect of continuous variables with small changes in hazard per unit increase in their value were reported for every appropriate unit increase. Mild anaemia was defined as a haemoglobin level of 80–140 g/L for females and 80–120 g/L for males. Severe anaemia was defined as a haemoglobin level < 80 g/L for both sexes. The BMI (Body Mass Index) cut-off's were <18.5 kg/m^2 for underweight , 18.5 to 25 kg/m^2 normal (healthy weight) and >25 kg/m^{2}overweight.

3.4 Data Extraction and Management

The motivating data was obtained from KEMRI-FACES-UCSF (Kenya Medical Research Institute –Family AIDS Care and Education Services – University of California San Francisco) which provides HIV care and treatment technical support to government health facilities and other local partners who provide direct patient HIV care and treatment. They offer HIV patient care and management which includes prevention, screening, diagnosis, treatment, and management of opportunistic infections and sexually transmitted infections, reproductive health services, cotrimoxazole prophylaxis and multi-vitamins, Prevention with Positives (PwP), and close monitoring of HIV disease progression through physical exams, WHO staging, CD4 cell count taking, and other laboratory investigations. Thereafter those patients meeting antiretroviral therapy (ART) criteria are initiated on ART following adherence counselling sessions. All HIV/AIDS services are provided in compliance with Kenya National Guidelines and best practices.

HIV care and treatment requires efficient information management to monitor patient clinical care. The routine patient care data collected at KEMRI-FACES-UCSF supported sites was entered into Open Medical Records System (OpenMRS; http://openmrs.org) which is an electronic medical record (EMR) system. The data from remote sites was then synchronized to a central server located at Lumumba Health Center using internet connectivity. At the central server; the data management, verification and cleaning was done regularly by a team of data quality officers, records staff and data managers. The extraction of socio-demographic characteristics, baseline and follow-up clinical and laboratory measurements, and treatment outcomes was done from patient medical record system using Structured Query Language (MySQL) queries. The extracted data was imported into Stata software, version 12.1 (StataCorp, Texas, USA), for study specific data management and analyses. All data was entirely observational with tests or interventions performed according to Kenya national guidelines on HIV Care and treatment at each clinical site.This study utilized only secondary data which was duly covered under Institutional Review Board (IRB) and Ethical Review Committee (ERC) obtained through KEMRI and UCSF, therefore no separate approvals were required.

3.5 Background on Survival Analyses approaches

During the past 10 years or more survival analysis has been ruled by two approaches namely Kaplan-Meier (non-parametric approach) and Cox proportional hazards regression (semi-parametric approach) [10, 8]. The survival time to develop new AIDS event T may be evaluated as a random variable having a probability distribution $F(t)$ and probability density function $f(t)$. The outcome of interest, the survivor function $S(t)$, is the probability of not developing a new AIDS event up to time t or beyond denoted by equation (3.1).

$$S\left(t\right) = P\left(T \geq t\right) = 1 - F(t) \tag{3.1}$$

Another function of importance is the hazard function, which basically represents an instantaneous failure rate. The hazard ratio (HR) has been used as the measure of the relative survival experience. It is the probability that an individual develops a new AIDS event at a time point given that the event has not yet occurred as shown in equation (3.2).

$$h\left(t\right) = \frac{f(t)}{S(t)} \tag{3.2}$$

The following section details the generalized proportional hazards model for new AIDS free survival data with covariates $\mathbf{z}$ as age, sex, marital status, CD4 cell count, haemoglobin level, BMI and prior TB medication. It is the standard Cox proportional hazards (Cox PH) model defined through a hazard function $h(t;\mathbf{z})$ as shown in equation (3.3).

$$h(t;\mathbf{z}) = h_0(t)\exp(\beta'\mathbf{z}) \tag{3.3}$$

where the baseline hazard function is $h_0\left(t\right) = \ h_0\left(t;\mathbf{0}\right)$ and beta is the coefficient. It may also be written in the integral form as in equation (3.4).

$$H(t;\mathbf{z}) = \left(\int_0^t h_0(u)du\right)\exp(\beta'\mathbf{z}) = H_0(t)\exp(\beta'\mathbf{z}) \tag{3.4}$$

where $H(t;\mathbf{z})$ represents the cumulative hazard function. According to Bennet [41] the proportional (cumulative) odds model with covariate vector $\mathbf{z}$ is defined as in equation (3.5)

$$O(t;\mathbf{z}) = \frac{1 - S(t;\mathbf{z})}{S(t;\mathbf{z})} = O_0(t)\exp(\beta'\mathbf{z}) \tag{3.5}$$

where $O_0(t) = O(t;\mathbf{0})$ and $O(t;\mathbf{z})$ is the odds of a new AIDS event occurring in $(0,\mathbf{t})$ for an individual with covariate vector $\mathbf{z}$. Covariates in the model act multiplicatively on the odds of a new AIDS event, as with more familiar logistic regression model.

Taking T as having a Weibull distribution with scale parameter $\sigma = p^{-1}$ and with a characteristic life μ.

We have our cumulative hazard function as in equation (3.6).

$$H(t) = -\ln S(t) \tag{3.6}$$

Then equation (3.7) below

$$ln\ H(t) = ln\left\{\left(\frac{t}{\mu}\right)^{p}\right\} = px - p\ln\mu = \frac{x - \ln\mu}{\sigma} \tag{3.7}$$

is linearly related to . If T is a distribution similar to a log-logistic, the log cumulative odds function will have a curvilinear relation to x by a function $s = s(x)$.

The survival function is denoted by equation (3.8).

$$S(t) = (1 + \exp s)^{-1} \tag{3.8}$$

The density function is given by equation (3.9).

$$f(t) = \frac{ds}{dt}\exp(s)(1 + \exp s)^{-2} \tag{3.9}$$

Equation (3.10) is the hazard function

$$h(t) = \frac{ds}{dt}\exp(s)(1 + \exp s)^{-1} \tag{3.10}$$

The approach used by Royston and Parmar [27] to estimate the hazard, density, and survival functions is to smooth either the baseline cumulative odds function or the baseline cumulative hazard function. With the same notation previously used in the paragraphs before, but for the time being suppressing $\mathbf{z}$, suppose that T is a survival-time random variable to a new AIDS event, having a log-logistic distribution with location parameter μ and scale parameter σ. Let $x = ln\,(t)$

We have it in equation (3.11) as.

$$S\,(t) = \left\{1 + \exp\left(\frac{x - ln\mu}{\sigma}\right)\right\}^{-1} \tag{3.11}$$

So much in equation (3.12),

$$lnO\,(t) = \ln\frac{1 - S(f)}{S(t)} \;=\; \frac{x - \ln\mu}{\sigma} \tag{3.12}$$

3.6 Limitations of Cox proportional hazard regression

Cox proportional hazards regression is often widely chosen and used choice of analysis for modelling survival data in medical studies. However, it has some intrinsic features that may cause problems for the analyst or the interpreter of the data:

1. It treats the baseline distribution of the observations as a high-dimensional nuisance parameter and is usually highly erratic. For example, a typical estimate of the baseline hazard function following Cox is a "noisy" step function. However, the contribution of the baseline survival is important as it impacts on the absolute survival probabilities over time.

2. It assumes that covariate effects act proportionally on the baseline hazard function, independent of time. It is not often the case; however this strong assumption is often not checked by the analysts.

3. Extending it to allow for non-proportional hazards is by no means a trivial modelling exercise. It does not give a complete probability specification for the data. Validation of the model and simulation of data sets realistically similar to a given one are impeded.

It is for the reasons outlined above that the choice of flexible parametric models is more desirable . For this thesis we used flexible parametric (RP) models as implemented in Stata's *stpm2* routine as proposed by Royston and Parmar [27]. When creating a prognostic survival model using regression, emphasis has been placed on value of the prognostic index based on covariates only, while ignoring the role of the baseline survival function. However, that is corrected by the use of flexible parametric models. Moreover, the time-dependent differences between the hazard functions are clearly and easily displayed after estimating a Royston Parmar (RP) model[38].

3.7 Spline-based parametric survival models

Spline-based parametric survival models are a better alternative to Cox proportional hazards regression. Splines are flexible mathematical functions defined by piece-wise polynomials which are used in regression models for non-linear effects. The points at which the polynomials join each other are called knots. Constraints are introduced to ensure the function is smooth. Splines can be of any degree (n), however the most commonly used splines are cubic splines where the function is forced to have continuous 0^{th}, 1^{st} and 2^{nd} derivatives.

Since the distribution of survival times to new AIDS event may be neither log-logistic nor Weibull, hence the need for more flexible models. The approach taken by Royston and Parmar [27] is to model the logarithm of the baseline cumulative odds or hazard function as a natural cubic spline function of log time, so the general function $s(x)$ is approximated by a spline. The proportional hazards model (PH) spline model with fixed covariate vector $\mathbf{z}$ may be written as below in equation (3.13).

$$\ln\{-lnS(t;\mathbf{z})\} = \ln H(t;\mathbf{z}) = \ln H_0\}(t) + \beta'\mathbf{z} = s(x) + \beta'\mathbf{z} \quad (3.13)$$

While the proportional odds (PO) spline model is as in equation (3.14).

$$\ln\left\{S(t;\mathbf{z})^{-1}-1\right\} \ =\ln O\left(t;\mathbf{z}\right)=\ln O_0\}\ \ (t)+\ \ \beta'\mathbf{z}{=}s\left(x\right)+\beta'\mathbf{z} \tag{3.14}$$

Summarized in equation (3.15)

Proportional hazards (PH) spline model :

$$\ln H\left(t;\mathbf{z}\right){=}s\left(x\right)+\beta'\mathbf{z} \tag{3.15}$$

and in equation (3.16) as shown Proportional odds (PO) spline model :

$$\ln O\left(t;\mathbf{z}\right){=}s\left(x\right)+\beta'\mathbf{z} \tag{3.16}$$

Natural cubic splines are defined as splines constrained to be linear beyond boundary knots k_{min} , k_{max} . Such knots are usually, but not necessarily, placed at the extreme observed x-values. In addition, m internal knots $k_1 < \cdots < k_m$ with $k_1 > k_{min}$ and $k_m < k_{max}$ are specified. It can be shown that the natural cubic spline may be written as shown in equation (3.17).

$$s\left(x\right)=\ \gamma_0+\gamma_1 x\ +\gamma_2 v_1\left(x\right)+\cdots+\gamma_{m+1}v_m\left(x\right) \tag{3.17}$$

where the $j\ th$ basis function for $j=1,\ldots,\ m$ as in (3.18) , (3.19) and (3.20).

$$v_j\left(x\right)=\left(x-k_j\right)_+^3\ -\lambda_{\ j}(x-k_{min})_+^3\ -(1-\lambda_{\ j})\left(x-k_{max}\right)_+^3 \tag{3.18}$$

$$\lambda_{\ j}=\frac{k_{max}-k_j}{k_{max}-k_{min}} \tag{3.19}$$

$$\left(x-a\right)_+^3=\max\left\{0,\ \left(x-a\right)^3\right\} \tag{3.20}$$

The curve complexity is dictated by the number of degrees of freedom (df), which ignoring γ_0 equals $m{+}1$. By convention, $m=0$ is taken to mean that no internal and no boundary knots are specified. The straight line model $x=\gamma_0+\gamma_1 x$ with df $=1$ is then obtained [27].

3.8 Statistical Analysis

Baseline descriptive statistics were provided as appropriate for all the variables.

Univariate Flexible parametric regression models were done geared towards identification of important predictors of the primary outcome (progression to AIDS), for consideration in the subsequent multivariable models.

Before fitting each of the Flexible parametric models, the proportionality assumption was informally assessed for categorical variables by graphical techniques. Some of the continuous variables associated with progression to AIDS were each categorized into quantiles and subsequently graphed the hazard ratios and corresponding 95% confidence intervals while assessing their relationship.

Kaplan-Meier graphs were plotted for the categorical variables exhibiting significant association with outcome while corresponding log rank test was displayed. P-values from the likelihood ratio test was reported for each univariable analysis while assessing statistical significance at 5% level. The effects of continuous variables with small changes in hazard per unit increase in their value were reported for a number of units increase. All Hazard ratios were computed from complete cases at this stage. The Kaplan-Meier 2-year survival probability was presented for the risk categories.

Summary of the final prognostic risk models obtained after *a priori* inclusion of all variables given their prognostic importance from literature. The significance level for each predictor in the multivariable model was set at 5% level. Adjusted hazard ratios associated with the predictors in the final model were shown besides respective P values.

The Akaike information criterion (AIC), equation (3.21), was used in selecting the model with the lowest AIC, implying it had the best fit [12].

$$AIC = 2k\ - 2log_e L \tag{3.21}$$

where k is number of parameters fitted and L is the estimated model's maximum likelihood function.

Missing data was expected (if any significant proportion) so there was a plan to that effect. To deal with missing data the multivariable model was re-estimated after multiple imputations of missing values using chained equations [7]. Ten imputed datasets were used to calculate mean estimates of parameters of interest, and appropriate adjusted standard errors using the within and between imputation standard errors of the estimates using Rubin's rules. The new hazard ratios were compared to those estimated from complete case analysis. However, the proportion of missing data across the variables was consistently low (less than or equal to 10 percent). In theory, if material variances were not noted as a consequence of bias then the complete cases represented the ignored incomplete cases. The estimated coefficients from the complete case analysis were used to assign risk scores to subjects if there was no marked bias expected.The changes in the P values are however not directly important as the changes in the actual estimates in the distribution of risk scores to patients. The imputed multivariable risk score regression model was also presented.

For model validation, bootstrapping method of internal validation which involved taking 50 samples with replacement from the original sample was used. It provides nearly unbiased estimates of high predictive accuracy and are of relatively low variance. It also has the advantage of using fewer model fits than in cross-validation and using the entire dataset as available [19]. We then used the data to predict 2-year AIDS free survival probability for each subject. The rate of progression to AIDS was done per 1,000 person-years of observation or follow-up.

The risk of progression to AIDS was derived from predicted probabilities based on final multivariable model. We displayed the results of classification of HIV patients into risk groups or classes according to their progression to AIDS risk scores, which are linear predictors from the prognostic models obtained. The discriminative power of the model was measured by Harrell's C-index ,bias corrected for possible over-fitting using the bootstrap method [19]. It was aimed at assessing the probability of concordance between predicted and observed responses where an index of 0.5 depicts no predictive discrimination while 1.0 depicts perfect discrimination [19]. Similarly the Somers' D-statistic [28] as measures of discrimination was calculated for the models. The D-statistic was also used as a measure of discrimination of the survival models because of its ability to stratify the

risk of progression to AIDS among groups of patients. Larger D statistics in prognostic models depict greater degree of separation. Given the difficulty of having a very practical prognostic tool for use by health caregivers that uses a complicated regression model, the estimated risk scores were translated to simple 18 unit score grading using the Z values for each covariate from the multivariable Royston Parmar model, by rounding off to the nearest whole number. These individual scores were then added together to provide an overall risk score for each patient whereby a score of 10 and above was considered high risk, if otherwise low risk which would provide the health caregivers with a probability of progression to AIDS for each possible final risk score. Statistical analyses were performed using Stata software, version 12.1 (StataCorp, Texas, USA).

Chapter 4

Results

4.1 Baseline Descriptive Statistics

The mean follow-up time for the clients was 490.8 days with a median of 665 days. The minimum follow-up time for the clients was 6 days while a maximum of 731 days. The total sample consisted of 1454 adults. Over half of the clients were female 981 (67.5%) and with a mean age of 32.1 years (SD = 9.4). Slightly above half of the clients, 716 (53.4%), were married. The mean BMI was 20.9 (SD = 4.1). Over half of the clients were underweight 798 (56.7%). The mean haemoglobin level of the clients was 112.7 (SD = 22.4), translating to over half, 768 (52.8%), with mild anaemia. Majority of the clients were either in WHO stage I or II, 559 (38.5%) and 518 (35.6%) respectively. Just under half, 553 (45.9%), had CD4 counts less than 200. Only 235 (17%) had prior TB treatment(Table4.1).

Table 4.1: Baseline Descriptive Statistics

Variables	**Total n(%)**
Female	981 (67.5)
Age(Years)Mean(SD)	32.1 (9.4)
Marital Status	
Married	716 (53.4)
Never Married	44 (3.3)
Single/Divorced	373 (27.8)
Widowed	209 (15.6)
Clinical	
Weight(kg) Mean(SD)	57.4 (11.0)
Height(cm) Mean(SD)	165.9 (9.0)
BMI (kg/m^2) Mean(SD)	20.9 (4.1)
BMI Categories	
Normal	398 (28.3)
Underweight	798 (56.7)
Overweight	212 (15.1)
Hemoglobin(g/L)Mean(SD)	112.7 (22.4)
Anaemia	
Normal	605 (41.6)
Mild Anaemia	768 (52.8)
Severe Anaemia	81 (5.6)
WHO Stage	
Stage I	559 (38.5)
Stage II	518 (35.6)
Stage III	347 (23.9)
Stage IV	30 (2.1)
CD4 Median(IQR)	212 (100 - 299)
CD4 categories	
<200 cells/ml	553 (45.9)
200 – 349 cells/ml	535 (44.4)
>= 350 cells/ml	118 (9.8)
Medical History	
Prior TB treatment	235 (17.0)

4.2 Baseline Descriptive Statistics by progression to AIDS

A total of 221 out of a total sample of 682 (who had no AIDS at baseline) followed up progressed to AIDS (outcome) either clinically or immunologically while 461 did not progress to AIDS. There was no notable difference in the mean baseline age and gender proportions among those who progressed to AIDS or not. The percentage of married or widowed clients also didn't differ markedly among those who progressed to AIDS or not.

The baseline BMI (kg/m^2) for those who progressed to AIDS was lower by a single unit, 21.5 vs. 22.5. In terms of the BMI categories, those with AIDS were less likely to be overweight, 16.5% vs. 24.2%. Similarly the mean baseline haemoglobin was lower by a unit among those with AIDS and when categorized a higher proportion had either mild or severe anaemia compared to those who did not progress to AIDS. Majority of those who progressed to AIDS had a baseline WHO stage II compared to those who did not, 57.9% vs. 35.8%. Clients without AIDS had a slightly higher baseline CD4 compared to those without, 283 vs. 280 cells/ml. A larger percentage of those who progressed to AIDS had a history of TB treatment prior to HAART initiation, 15% vs. 4.6% (Table 4.2). There were 55 (3.8%) who died during the follow-up period. We had low proportions of missing data; 4% missing BMI values, 9% missing haemoglobin values, 5% missing WHO staging and 10% missing CD4 values.

Table 4.2: Patient Characteristics by progression to AIDS from Baseline

Variables	**AIDS Baseline n(%)**	**AIDS Follow-up n(%)**	**No AIDS n(%)**
Female	485 (62.8)	158 (71.5)	338 (73.3)
Age(Yrs)Mean(SD)	32.5 (9.8)	33 (9.2)	31 (8.7)
Marital Status			
Married	362 (50.1)	104 (51.7)	250 (59.7)
Never Married	19 (2.6)	10 (5.0)	15 (3.6)
Single/Divorced	227 (31.4)	49 (24.4)	97 (23.2)
Widowed	114 (15.8)	38 (18.9)	57 (13.6)
Clinical			
Weight(kg)Mean(SD)	54.6 (10.2)	58.6 (10.2)	61.6 (11.2)
Height(cm)Mean(SD)	166.2 (9.3)	165.7 (9.2)	165.5 (8.4)
BMI (kg/m2)Mean(SD)	19.9 (3.9)	21.5 (4.2)	22.5 (3.9)
BMI Categories			
Normal	297 (39.6)	50 (22.9)	51 (11.6)
Underweight	384 (51.1)	132 (60.6)	282 (64.2)
Overweight	70 (9.3)	36 (16.5)	106 (24.2)
Hgb (g/L) Mean(SD)	110 (23.0)	111.5 (20.5)	117.8 (21.5)
Anaemia			
Normal	329 (42.6)	78 (35.3)	198 (43.0)
Mild Anaemia	395 (51.2)	131 (59.3)	242 (52.5)
Severe Anaemia	48 (6.2)	12 (5.4)	21 (4.6)
WHO Stage			
Stage I	170 (22.0)	93 (42.1)	296 (64.2)
Stage II	225 (29.2)	128 (57.9)	165 (35.8)
Stage III	347 (45.0)	0 (0)	0 (0)
Stage IV	30 (3.9)	0 (0)	0 (0)
CD4 Median(IQR)	124 (49-188)	280 (237-328)	283 (239-329)
CD4 categories			
<200 cells/ml	553 (79.3)	0 (0)	0 (0)
200 – 349 cells/ml	95 (13.6)	131 (86.8)	309 (86.3)
>= 350 cells/ml	49 (7.0)	20 (13.3)	49 (13.7)
Medical History			
Prior TB treatment	183 (24.7)	32 (15.5)	20 (4.6)

4.3 Univariable Flexible Parametric (RP) Models for progression to AIDS

On univariable analysis there was a significant increase in risk of progression to AIDS per five year increase in the age of the client (HR=1.08 95% CI: 1.01 - 1.16; p = 0.024). The widowed clients compared to the married were associated with a 49% increase in risk of progression to AIDS (HR=1.49 95% CI: 1.03 - 2.16; p = 0.036). Clients with higher baseline BMI values were significantly less likely to progress to AIDS (HR=0.93 95% CI: 0.90 - 0.97; p = 0.001). Higher baseline haemoglobin counts were associated with a small but significant lowered risk of progression to AIDS. Clients initiating HAART at WHO stage 2 almost had a twofold risk of progressing to AIDS in comparison to those starting at WHO stage 2 (HR= 1.94 95% CI: 1.48 - 2.53; p <0.001). Similarly, clients with a history of TB treatment prior to HAART initiation had a twofold risk of progression to AIDS (HR= 2.40 95% CI: 1.65 - 3.50; p <0.001). The higher baseline CD4 values showed a trend of reduced risk of progression, however without any statistical significance. There was no clear association of risk to gender (Table 4.3).

Table 4.3: Univariable RP models for progression to AIDS

Variables	Crude HR	(95% CI)	P value
Demographics			
Gender			
Female	Ref		
Male	1.05	0.78 - 1.41	0.747
Age (per 5 years)	**1.08**	**1.01 - 1.16**	**0.024**
Marital Status			
Married	Ref		
Never Married	1.22	0.64 - 2.34	0.542
Single/Separated/Divorced	1.2	0.86 - 1.69	0.289
Widowed	**1.49**	**1.03 - 2.16**	**0.036**
Clinical characteristics			
Weight (kg)	**0.97**	**0.96 - 0.99**	**<0.001**
Height (cm)	1	0.99 - 1.02	0.722
BMI (kg/m^2)	**0.93**	**0.9 - 0.97**	**0.001**
BMI Categories			
Normal	Ref		
Underweight	**0.52**	**0.37 - 0.72**	**<0.001**
Overweight	**0.38**	**0.25 - 0.58**	**<0.001**
Hgb (g/L)	**0.99**	**0.98 - 0.99**	**<0.001**
Hgb Categories			
Normal	Ref		
Mild Anaemia	1.11	0.84 - 1.48	0.448
Severe Anaemia	1.25	0.68 - 2.3	0.468
Hgb 10 g/dL cutoff			
<=10 g/dL	Ref		
>10 g/dL	**0.62**	**0.46 - 0.83**	**0.002**
WHO Stage			
Stage I	Ref		
Stage II	**1.94**	**1.48 - 2.53**	**<0.001**
CD4(per 100 cells/ml)	0.97	0.84 - 1.11	0.643
Medical History			
Prior TB treatment	**2.4**	**1.65 - 3.5**	**<0.001**

4.4 Multivariable Flexible Parametric (RP) Models for progression to AIDS

Complete data results

On multivariable analysis of the covariate and factors associated with progression to AIDS. There was a difference in trend towards a reduced risk per five year increase in age, however, without any significance. There was no association of risk with respect to gender. The widowed still had an increased risk compared to the rest of the marital categories, however with borderline significance (HR=1.64 95% CI: 0.99 - 2.72; p = 0.056). Higher baseline BMI and haemoglobin levels were significantly associated with a reduced risk of progression to AIDS. Clients initiating HAART at WHO stage 2 had a 64% increased risk of progression to AIDS compared to those starting at WHO stage 1 (HR=1.64 95% CI: 1.09 - 2.47; p = 0.016). There was no association of baseline CD4 counts with an increased or lowered risk of progression. Clients with a history of TB treatment prior to HAART initiation still had a twofold risk of progression adjusting for the other covariates and factors (HR=2.39 95% CI: 1.38 - 4.13; p = 0.002) (Table 4.4). The Akaike Information Criterion (AIC) was used to select the optimum number of spline knots which was optimally determined as 2 internal spline knots (d.f. = 3 for the *stpm2* command).

Table 4.4: Complete data multivariable RP model

Variables	Adj. HR	(95% CI)	P value
Demographics			
Gender			
Female	Ref		
Male	1	0.56 - 1.78	0.999
Age (per 5 years)	0.94	0.83 - 1.07	0.335
Marital Status			
Not widowed	Ref		
Widowed	1.64	0.99 - 2.72	0.056
Clinical characteristics			
BMI (kg/m^2)	**0.94**	**0.89 - 1**	**0.045**
Haemoglobin (g/L)	**0.99**	**0.98 - 1**	**0.046**
WHO Stage			
Stage I	Ref		
Stage II	**1.64**	**1.09 - 2.47**	**0.016**
CD4 (per 100 cells/ml)	1.01	0.87 - 1.18	0.884
Medical History			
Prior TB treatment	**2.39**	**1.38 - 4.13**	**0.002**

Imputed data results

The hazard ratios of the imputed predictors in the multivariable model did not indicate any major variations in the hazard ratios with the exception of marital status variable 'widowed and previous TB treatment where there was a change of 0.5 and 0.4 in the

hazard ratio respectively. The confidence intervals are also not fundamentally different after imputation (Table 4.5).

Table 4.5: Imputed data multivariable RP model for progression to AIDS

Variables	**Adj. HR**	**(95% CI)**	**P value**
Demographics			
Gender			
Female	Ref		
Male	1.08	0.74 - 1.57	0.697
Age (per 5 years)	1.03	0.95 - 1.12	0.407
Marital Status			
Not widowed	Ref		
Widowed	1.14	0.78 - 1.65	0.496
Clinical characteristics			
BMI (kg/m^2)	**0.95**	**0.91 - 0.98**	**0.007**
Haemoglobin (g/dL)	0.99	0.98 - 1	0.073
WHO Stage			
Stage I	Ref		
Stage II	**1.68**	**1.27 - 2.24**	**<0.001**
CD4 (per 100 cells/ml)	0.98	0.85 - 1.08	0.748
Medical History			
Prior TB treatment	**2**	**1.36 - 2.93**	**<0.001**

The KM survival curves and the AIDS hazard rate show a clear separation in the curves and also indicates that clients in WHO stage 2 at HAART initiation have lower AIDS free survival and high risk of progression to AIDS compared to clients at WHO stage 1 throughout the entire follow-up period as shown in Figure 4.1 and Figure 4.2 respectively.

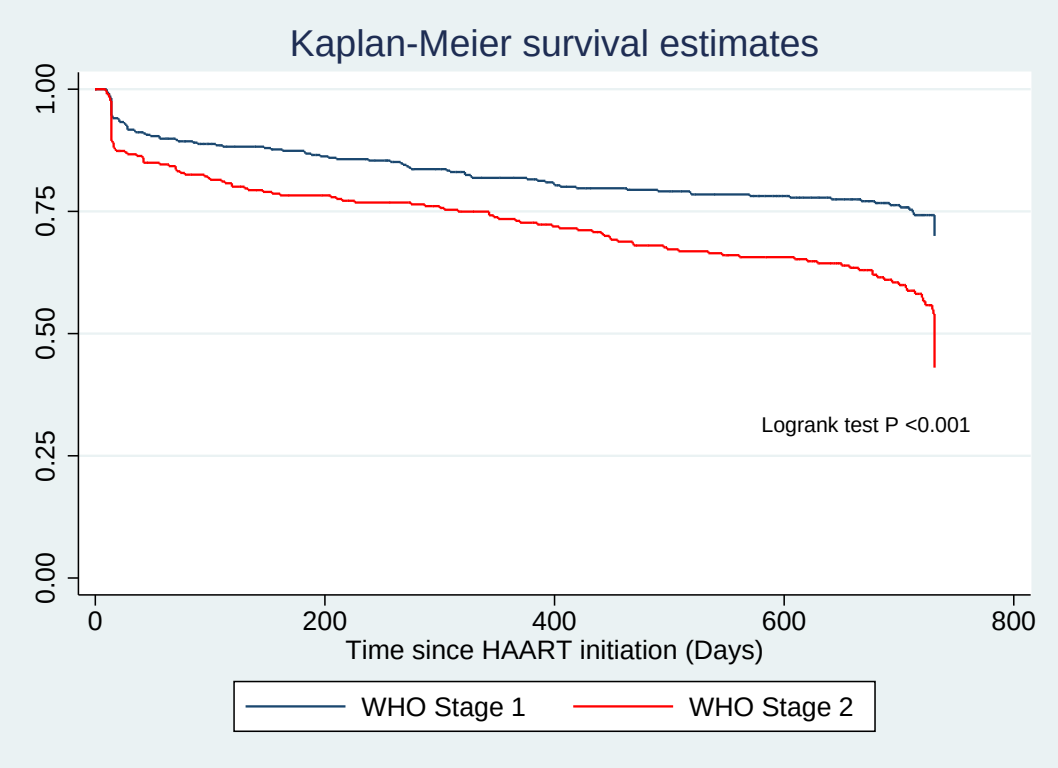

Figure 4.1: KM AIDS free survival estimates by WHO stage

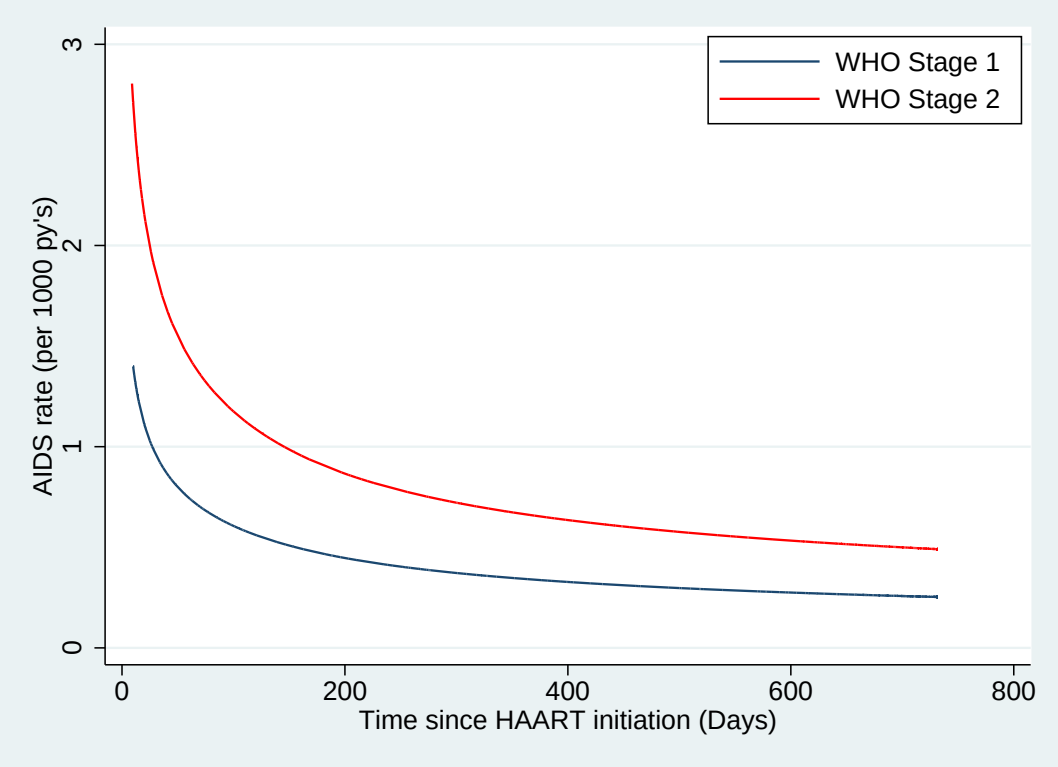

Figure 4.2: Rate of progression to AIDS by WHO stage

In terms of baseline BMI categories, the underweight clients had the poorest AIDS free survival and also had the highest AIDS progression rate as depicted in Figure 4.3 and Figure 4.4 respectively.

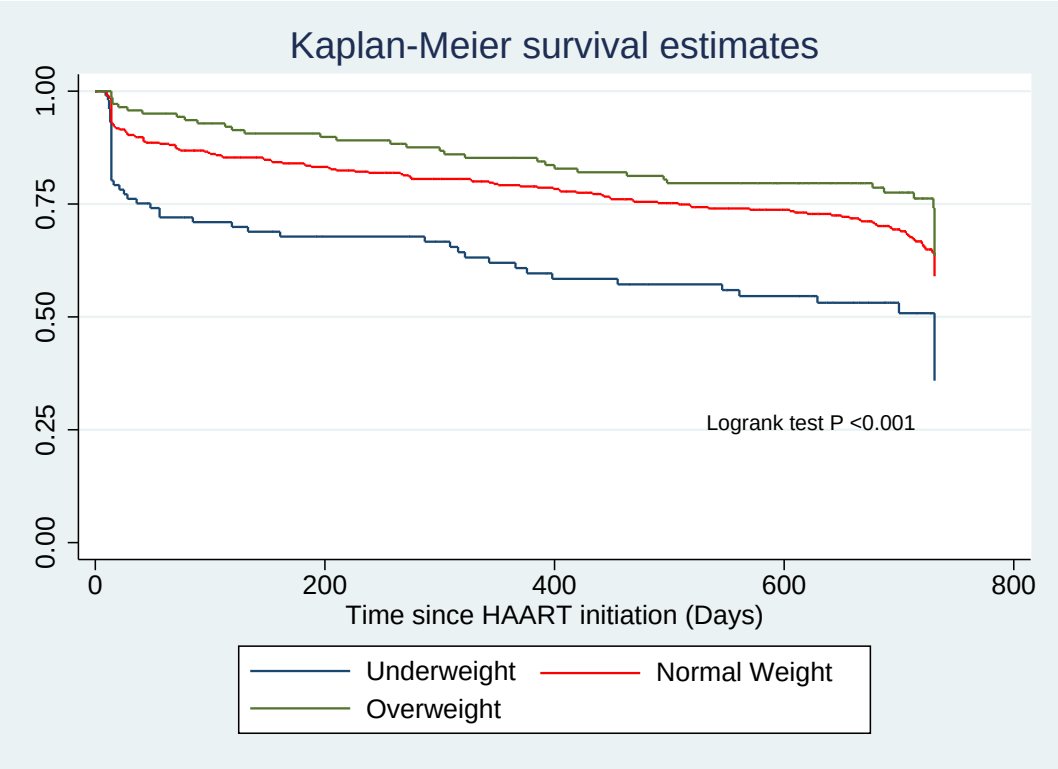

Figure 4.3: KM AIDS free survival estimates by BMI

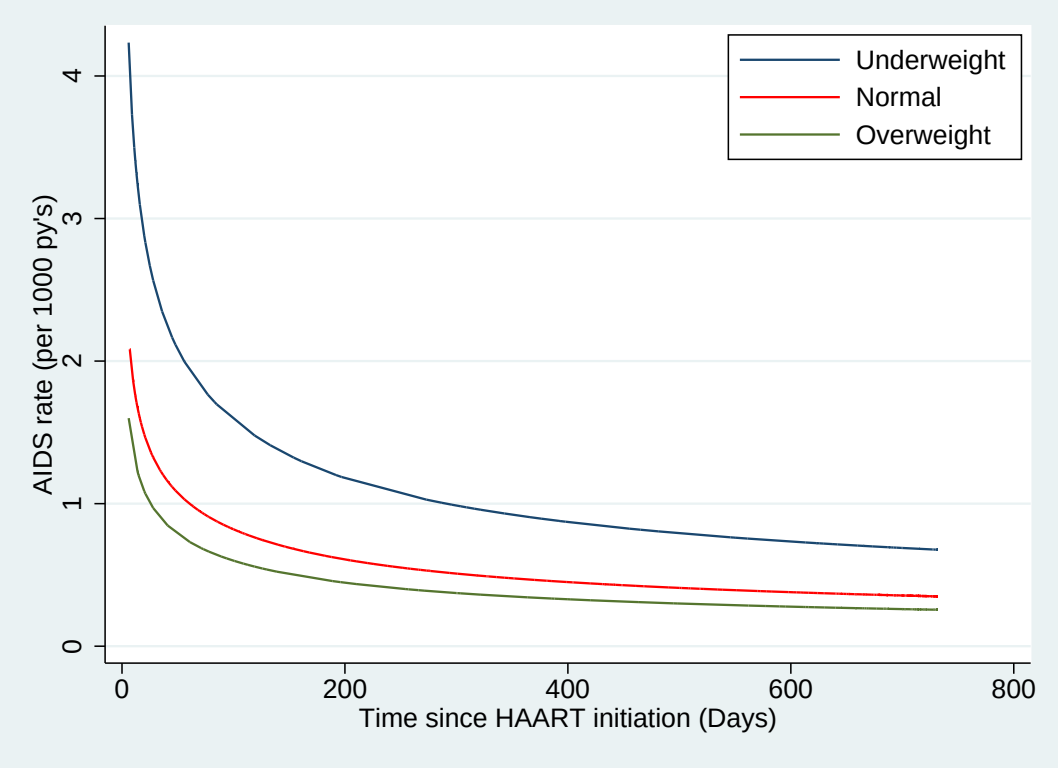

Figure 4.4: Rate of progression to AIDS by BMI

Clients with a history of TB treatment prior to HAART initiation had a poor AIDS free survival compared to those without and also had a higher AIDS progression rate as shown in Figure 4.5 and Figure 4.6 respectively.

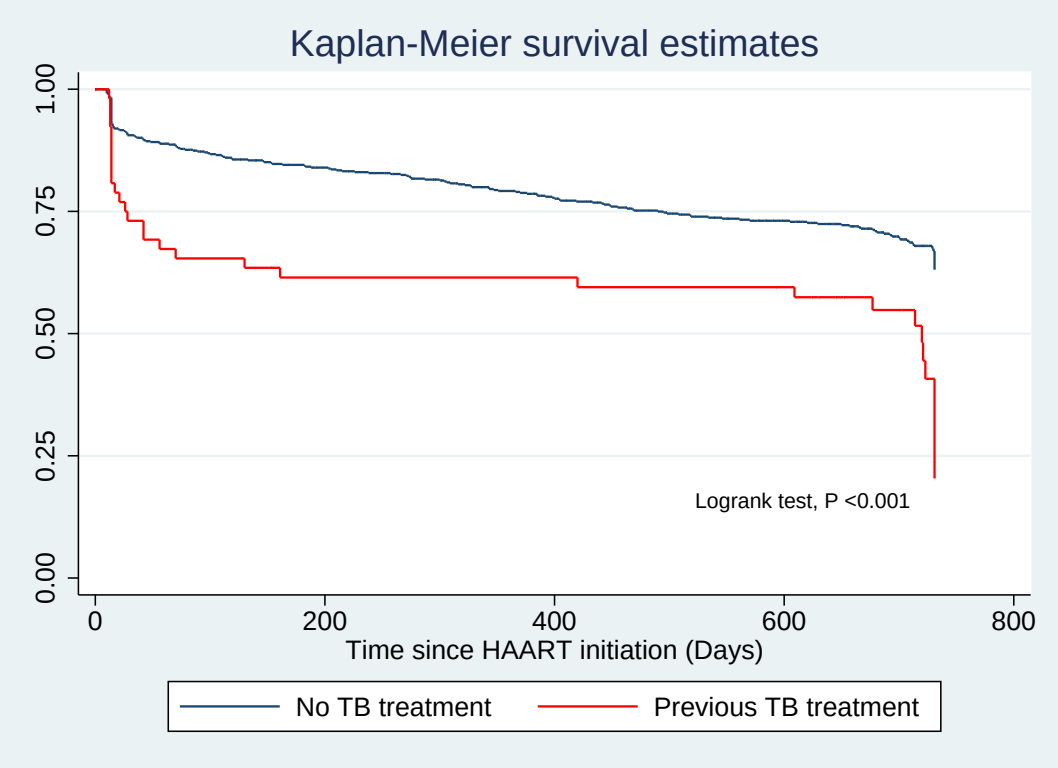

Figure 4.5: KM AIDS free survival estimates by TB trx

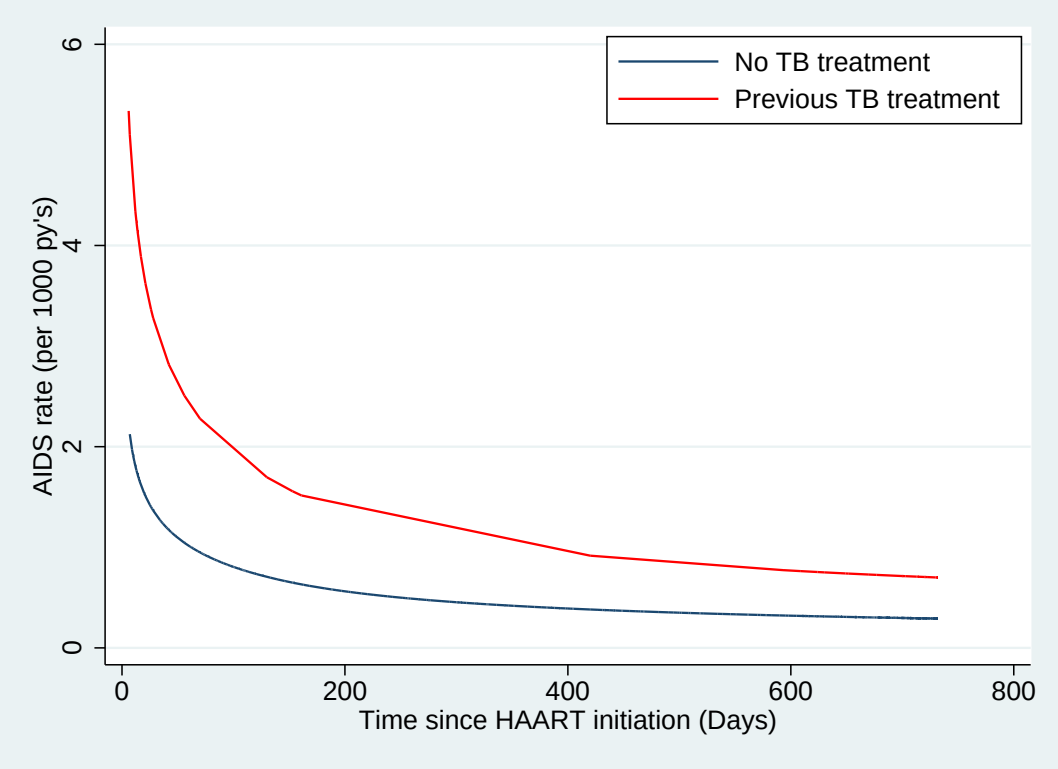

Figure 4.6: Rate of progression to AIDS by TB trx

4.5 Risk Score for progression to AIDS

Subjects were divided into two risk groups with individuals in the risk group 1 (Low Risk) having the lowest risk scores and those in risk group 2 (High Risk) having the highest risk scores. The risk score cutoff was developed based on the lowest score of the upper quintile (High risk group) which was 0.572.

The median risk score was 0.41 (IQR: 0.28 - 0.48) in the Low Risk group and 0.95 (IQR: 0.70 - 1.89) in the High Risk group. The rate of progression to AIDS in the Low Risk group was 0.38 per thousand person-years of observation while 0.93 per thousand person-years of observation in High Risk group (Table 4.6).

The hazard of progression to AIDS for the High Risk group was twice more than that of the Low Risk group (HR=2.47 95% CI: 1.66 - 3.69; p <0.001) (Table 4.6).

An assessment of the predictive power of the prognostic survival model estimated by internal validation using bootstrap samples showed a moderately strong concordance and discrimination ability, Harrell's C-index (C 0.69 95% CI 0.62 - 0.77) and Somers' D (D 0.41 95% CI 0.24 - 0.58).

Table 4.6: Risk group statistics for progression to AIDS

Risk	Median Risk (IQR)	Events	Rate per 1000 pyo (95%CI)	KM	HR(95%CI)
1 (Low)	0.41 (0.28 - 0.48)	35	0.38 (0.27 - 0.52)	0.81	Ref.
2 (High)	0.95 (0.70 - 1.89)	72	0.93 (0.74 - 1.17)	0.65	2.47 (1.66 - 3.69)

pyo - Person Years of Observation, KM- Kaplan-Meier estimate, HR- Hazard Ratio, CI- Confidence Interval

Figure 4.7 and Figure 4.8 respectively shows the AIDS free survival experiences of subjects in the two risk groups for the prognostic model. The individual Kaplan Meier (KM) survival curves (Log rank test p<0.001) are well separated from each other, suggesting a good discriminative potential, with wider gaps between high and low risk groups of

throughout the 2 year follow-up period. Similarly the line graphs of AIDS hazard rate show a good discrimination, with higher hazard rate in the first few days and stabilizing with a difference in the curves throughout the follow-up period. There is very high discrimination in the curves especially during the first 90 days of ART initiation.

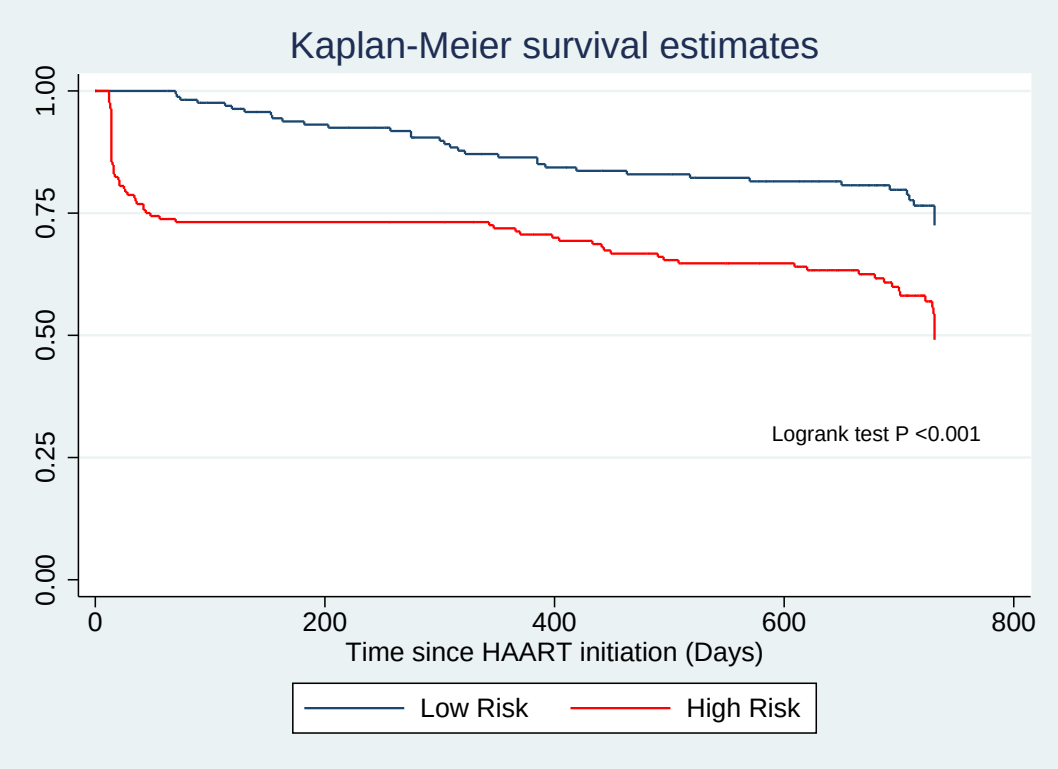

Figure 4.7: KM AIDS free survival estimates by Risk group

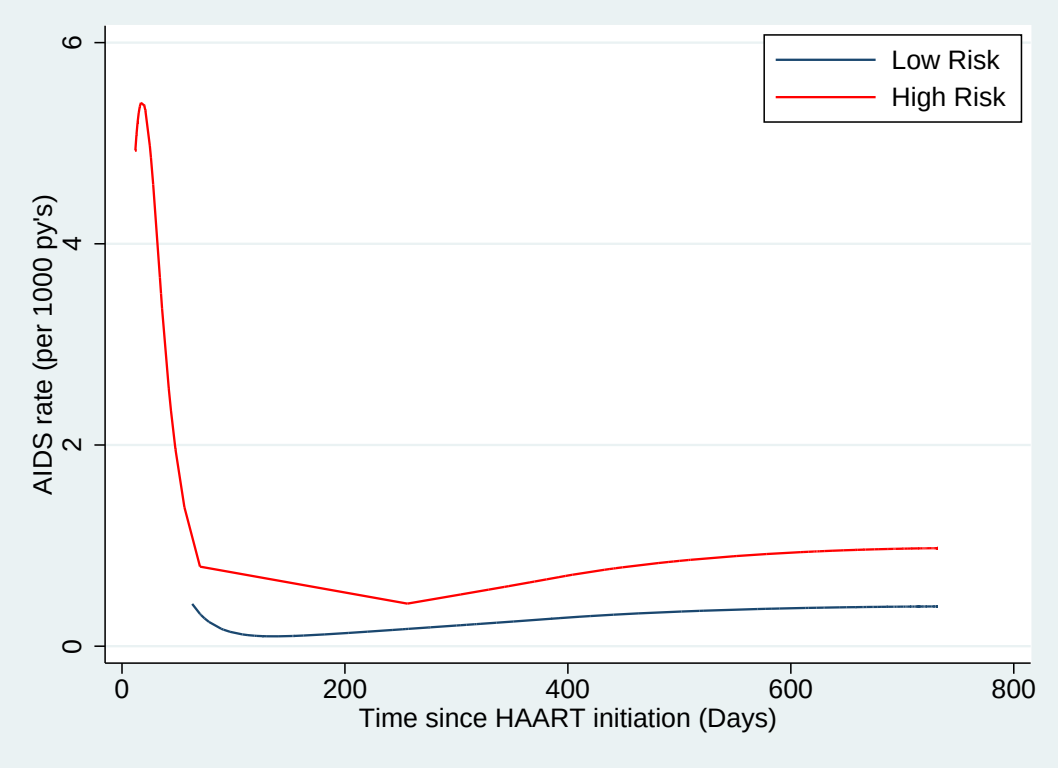

Figure 4.8: Rate of progression to AIDS by Risk group

4.6 Scoring System

The risk score was translated from the prognostic risk score model using the z values in table 4.7, to derive a simple score by rounding off to the nearest whole number. The score was assigned as follows; If the client was widowed the score was 2 and 0 if otherwise. For age we had a 1 point increase in risk for each 5 year increase in age up to the age of 35 years (capped at risk score of 4 for age 15 to 19 year and risk core of 0 for age 35 and above). Clients with a BMI of less than 15 were high risk were a score of 4 while BMI of 15 to 18 attracted a score of 2 while those with BMI greater than 18 scored 0. Haemoglobin value of less than or equal to 10 g/dL was assigned a score of 2 while haemoglobin value of more than 10 g/dL was assigned a score of 0. Clients with WHO stage II at baseline got a score of 3 while those with WHO stage I got a score of 0. Gender and CD4 did not contribute meaningfully to the prognostic model hence scored at 0 for either. Clients with a history of TB scored 3 while those without assigned a score of 0. The total maximum score is 18; the risk score grading classification is High risk (Score

of 10 and above) and Low risk (Score of 9 and below).

Based on this scoring system a female (Score = 0) , aged 32 years (Score = 1) , widowed (Score = 2), BMI of 12 (Score = 4), haemoglobin value of 9 (Score = 2) , WHO stage II (Score = 3) , CD4 of 400 (Score = 0) and without any TB treatment prior to HAART initiation (Score = 3): The total score for this client is 15, classified as high risk.

Table 4.7: Scoring system based on Multivariable prognostic model's Z values

Variables	Z value	Risk Score
Demographics		
Gender		
Female	0.00	0
Male	0.00	0
Age(per 5 year increase)	-0.96	
15 - 19		4
20 - 24		3
25 - 29		2
30 - 34		1
35 +		0
Marital Status		
Widowed	1.91	2
Not widowed		0
Clinical characteristics		
Body Mass Index (kg/m2)	-2.01	
< 15		4
15 - 18		2
> 18		0
Haemoglobin (g/L)	-2.00	
<=10 g/dL		2
>10 g/dL		0
WHO Stage		
Stage I	2.40	0
Stage II		3
CD4 (per 100 cells/l)	0.15	0
Medical History		
Prior TB treatment	3.11	
Yes		3
No		0

Chapter 5

Summary, Conclusion and Recommendations

5.1 Introduction

This chapter cover the summary of the thesis accomplishments, conclusion and recommendations drawn.

5.2 Summary

The aim of the study was to develop a prognostic risk score model for predicting progression to AIDS from a set of associated variables. The study utilized flexible parametric models developed by Royston and Parmer and implemented in Stata's *stpm2* routine. The prognostic risk score model specifically utilized demographic, clinical and medical variables that are easily available within a clinic or hospital setting in Sub Saharan Africa. The prognostic model included deomgraphic, lab and clinical variables.

The study found out that the strong predictors were BMI, haemoglobin, WHO staging and TB treatment prior to HAART initiation. Other factors trending towards statistical significance only included marital status, specifically being widowed or not. Age, CD4

count and gender did not have a statistically significant effect on risk of progression to AIDS in this study sample.

The study was able to develop a two group risk categorization based on the prognostic risk score model developed. The discriminative ability of the risk model was moderately strong (Harrell's c-index of 0.69). The rate of progression to AIDS between the high and low risk groups was well defined. The rate of progression was 0.38 and 0.93 per thousand person-years of followup for the low risk and high risk groups respectively representing more than twofold risk of progression to AIDS among high risk group, (HR= 2.47 95% CI: 1.66 - 3.69; p <0.001). The risk groups were finally translated into an easy to use 18 point prognostic index where a score of 10 and above was considered high risk and score below 10 considered low risk.

5.3 Conclusions

Baseline haemoglobin, Body Mass Index (BMI), WHO staging and TB treatment prior to HAART initiation were among the factors strongly associated with progression to AIDS in the prognostic risk score model developed. Baseline haemoglobin was significantly associated with 7% reduction in the risk of progression to AIDS in the prognostic model developed. This result is consistent with what was reported in a previous study [13].

Increasing age has been associated with the increased risk of progression to AIDS in several studies [9, 29, 24]. In one study when age was fitted as a continuous variable and adjusted for exposure group, the relative risk of developing AIDS by any time after seroconversion was 34% for a 10-year increase difference [24]. However, in this multivariable prognostic risk score model the positive significant effect of age at univariable analysis was diminished and was not statistically significant. The effect changed from 8% increase in risk of AIDS per 5-year increase difference in age at univariable analysis to 4% decrease in risk of AIDS per five year increase in age, however without statistical significance. The change in direction of age effect as a predictor in the prognostic model could be the due to HIV stigma leading to late HAART initiation among the younger HIV clients. This

in turn leaves them exposed to a high risk of progression to AIDS.

An increased baseline body mass index (BMI) has been associated with reduced risk of progression to AIDS [13]. Consistent with our results there was a 1% significant reduction in risk of progression to AIDS for increased BMI values. Even though weight loss is only noted in about one-third of the HIV-infected individuals who develop AIDS it is highly predictive for AIDS. The higher WHO staging has been previously associated with an increased risk of progression to AIDS. The results from this study indicated a 64% increase in risk of progression to AIDS among those initiating HAART at WHO stage II in comparison to WHO stage I. We would naturally expect those in WHO stage II to progress easily to WHO stage III or IV.

History of Tuberculosis (TB) among HIV is usually associated with increased risk of progression to AIDS according to studies conducted previously [15] . Clients who had received TB treatment prior to HAART initiation had a two fold increase in the risk of progression to AIDS, hazard ratio of 2.39. This is consistent with the results from the studies done which indicate that prior TB has been reported to increase the risk of subsequent TB illness among HIV infected clients [6]. Opportunistic infections occur in the course of increasing HIV immunosuppression and are largely an indication of decreased immunity. The baseline CD4 has been pointed as a significant predictor of progression to AIDS among HIV clients by several earlier studies [40, 29, 23]. In the univariable prognostic model there was a trend towards reduction of risk for higher CD4 counts, about 4% reduction in risk, however this effect is lost in the multivariable model. This could be due to challenges

Male gender has been identified as a risk factor for death outcome, however not clearly identified in previous research as risk factor for progression to AIDS [29]. This study did not find any association of risk with gender. There have been some reports that HIV disease progresses at different rates by gender, however, the majority of studies have not found significant differences. Bereavement, depression, stress and other psychological factors have been hypothesized to affect disease progression, even though majority of studies have shown no association of psycho-social aspects of the clients and overall HIV disease progression [44]. This was able to highlight the social and emotional impact of

HIV on clients' risk of progression to AIDS. However some studies have reported an association with depressive symptoms and more rapid loss of CD4 lymphocytes. Our results, 64% increase in risk, indicated a trend toward rapid progression to AIDS among those who were widowed compared to those who were not [18].

The data used was obtained from adult (15 years and above) male and female HIV clients in Nyanza region of Western Kenya. The client's demographic characteristics could be homogeneous but they could still differ in other unmeasured ways. However, it is clear risk score really predicts progression very well in the first 90 days. The prognostic model built in this study may also not be well prognostic in all HIV patients including those from different countries, races, ethnicity because of the geographic setting from which the data was acquired. However we can be able to generalize the prognostic model to some extent. It is well known that fitting and validating the prognostic model on a same set of data can lead to over-optimistic estimates of predictive value [27, 30]. However we tried to remedy the situation using formal bootstrapping approaches to evaluate the prognostic ability of the prognostic model developed. This study was able to incorporate demographic variables such as marital status and test the prognostic ability of such a variable. This was able to highlight the social and emotional impact of HIV on clients' risk of progression to AIDS.

Overall, the prognostic risk score model developed had a good predictive value and discriminative ability and can be applied in predicting progression to AIDS in similar settings in Sub Saharan Africa. The tool has utilised the flexible parametric model which is able to model the baseline and use it in the final risk score which is an advantage over the traditional Cox proportional hazards regression. The risk score model should however be used with caution in HIV-infected clients as it has not been externally validated in other populations where their performance may be unreliable.

5.4 Recommendations

This simple and clinically sensible prognostic risk score model can be used for timely prognosis (especially in the first 90 days of ART initiation) of HIV clients at risk of

progression to AIDS, timely change in regimen, change in clinic schedule and appropriate medical action to slow down progression to AIDS. It may also be used to study and monitor the health status of a group of patients on HAART in comparison to other cohorts in assessing changes in the short term risk of clinical progression. The risk score may also be used by the individual and their clinician in future planning; to assess if the risk of disease progression was sufficiently low that the patient could be seen in 6 months or later rather than 3 months. Overall, it is of such a great importance to prevent progression to AIDS among HIV infected clients.

There are however limitations of the prognostic risk score model developed that could be addressed by future studies. It is clear that the risk score really predicts progression very well in the first 90 days but would need further calibration for longer periods. However, the prognostic risk score model built in this study may not perform uniformly in all HIV patients including those from different countries, races, ethnicity because of the geographic setting from which the data was acquired, genetic response and other factors at play. However we can be able to generalize the risk score model to some extent. It is also well known that fitting and validating the prognostic model on same set of data can lead to over-optimistic estimates of predictive value [27, 30]. The prognostic model was developed and internally validated through bootstrap re-sampling of relevant statistics. Further research could fit the model on data from different populations and study how the model performs. In this instance, given absence of an external population data for validation, the situation was remedied using formal bootstrapping approaches to evaluate the prognostic ability of the prognostic risk score model developed. Even though the results of the validation suggested good precision, external validation of the prognostic risk score models to different or similar populations would be a better way of objectively testing the validity of this prognostic risk score model. Further modelling could be done to look at a combined outcome of AIDS or death, notwithstanding the challenge of obtaining accurate death records, to be able to come up with a modified prognostic risk score model of progression to AIDS or death for such purposes as a next step.

References

[1] Mocroft A., Ledergerber B., Zilmer K., and et al. Short-term clinical disease progression in HIV-1-positive patients taking combination antiretroviral therapy: the EuroSIDA risk-score. *AIDS*, 21:1867–1875, 2007.

[2] Mocroft A., Kirk O., Barton S.E., and et al. Anaemia is an independent predictive marker for clinical prognosis in HIV-infected patients from across Europe. EuroSIDA study group. *AIDS*, 13:943–950, 1999.

[3] Babiker A.G., Peto T., Porter K., and et al. Age as a determinant of survival in hiv infection. *J Clin Epidemiol*, 54:16–21, 2001.

[4] Phillips A.N., Lee C.A., Elford J., and et al. More rapid progression to aids in older hiv-infected people: the role of cd4>+ t-cell counts. *J Acquir Immune Defic Syndr*, 4:970–975, 1991.

[5] Miladinovic B., Kumar A., Mhaskar R., and et al. A Flexible Alternative to the Cox Proportional Hazards Model for Assessing the Prognostic Accuracy of Hospice Patient Survival. *PLoS ONE*, 7:1371–1375, 2012.

[6] Seyler C., Toure S., Anglaret X., and et al. Risk factors for active tuberculosis after antiretroviral treatment initiation in Abidjan. *Am J Respir Crit Care Med*, 172:123–7, 2005.

[7] Rubin D. *Multiple imputation for non response in surveys*. Wiley, 1987.

[8] Cox D.R. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220, 1972.

[9] Malvy E., Thiebaut R., Marimoutou C., and et al. Weight loss and body mass index as predictors of HIV disease progression to AIDS in adults. Aquitaine Cohort, France 1985–1997. *J Am Coll Nutr*, 20:609–615, 2001.

[10] Kaplan E.L. and Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

[11] Centers for Disease Control and Prevention. Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Morb Mortal Wkly Rep*, 41:1–19, 1993.

[12] Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. *TAC*, 1:716–723, 1974.

[13] Byakwaga H., Petoumenos K., Ananworanich J., and et al. Predictors of Clinical Progression in HIV-1-Infected Adults Initiating Combination Antiretroviral Therapy with Advanced Disease in the Asia-Pacific Region: Results from the TREAT Asia HIV Observational Database. *J Int Assoc Provid AIDS Care*, 12:270–277, 2013.

[14] Christian H.O., Jose G., Bruno L., and et al. Risk of AIDS and death at given HIV-RNA and CD4 cell count, in relation to specific antiretroviral drugs in the regimen. *AIDS*, 19:319–330, 2005.

[15] Zhou J., Elliott J., Li P.C., and et al. Risk and prognostic significance of tuberculosis in patients from The TREAT Asia HIV Observational Database. *BMC Infect Dis*, 9:46, 2009.

[16] Zhou J. and Kumarasamy N. Predicting short-term disease progression among HIV-infected patients in asia and the pacific region: preliminary results from the Treat Asia HIV Observational Database (TAHOD). *HIV Med*, 6:216–223, 2005.

[17] Lundgren J.D., Mocroft A., Gatell J.M., and et al. A clinically prognostic scoring system for patients receiving highly active antiretroviral therapy: results from the EuroSIDA study. *J Infect Dis*, 185:178–187, 2002.

[18] Burack J.H., Barrett D.C., Stall R.D., and et al. Depressive symptoms and CD4 lymphocyte decline among HIV-infected men. *JAMA*, 270:2568–73, 1993.

[19] Harrell F.E. Jr., Lee K.L., and Mark D.B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, 15:361–387, 1996.

[20] National AIDS Control Council Kenya. Maisha newsletter : Empowering communities in the fight against HIV & AIDS, 2012.

[21] May M., Porter K., Sterne J.A.C., and et al. Prognostic model for HIV-1 disease progression in patients starting antiretroviral therapy was validated using independent data. *J Clin Epidemiol*, 58:1033–1034, 2005.

[22] May M., Royston P., Egger M., and et al. Development and validation of a prognostic model for survival time data: application to prognosis of HIV positive patients treated with antiretroviral therapy. *Stat Med*, 23:2375–2398, 2004.

[23] Brinkhof M.W.G., Boulle A., Weigel R., and et al. Mortality of HIV-Infected Patients Starting Antiretroviral Therapy in Sub-Saharan Africa: Comparison with HIV-Unrelated Mortality. *PLoS Med*, 6:1371–1375, 2009.

[24] Carre N., Deveau C., Belanger F., and et al. Effect of age and exposure group on the onset of AIDS in heterosexual and homosexual HIV-infected patients. *AIDS*, 8:797–802, 1994.

[25] World Health Organization. WHO 2012: World Health Statistics. Technical report, World Health Organization, 2012.

[26] Braitstein P., Brinkhof M.W., Dabis F., and et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet*, 367:817–824, 2006.

[27] Royston P. and Parmar M.K.B. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21:2175–2197, 2002.

[28] Royston P. and Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med*, 23:723–748, 2004.

[29] Srasuebkul P., Lim P.L., Lee M.P., and et al. Short-Term Clinical Disease Progression in HIV-Infected Patients Receiving Combination Antiretroviral Therapy:

Results from the TREAT Asia HIV Observational Database. *Clinical Infectious Diseases*, 48:940–950, 2010.

[30] Lambert P.C. and Royston P. Further development of flexible parametric models for survival analysis. *The Stata Journal*, 9:265–290, 2009.

[31] United Nations Development Programme. South Africa Human Development Report 2003: The challenge of Sustainable Development in South Africa; Unlocking People's Creativity. Technical report, United Nations Development Programme, 2003.

[32] United Nations Development Programme. Botswana Human Development Report 2000: Towards an AIDS-free generation. Technical report, United Nations Development Programme, 2005.

[33] United Nations Development Programme. Malawi Human Development Report 2005: Reversing HIV/AIDS in Malawi. Technical report, United Nations Development Programme, 2005.

[34] United Nations Development Programme. Kenya Human Development Report 2006: Human security and human development; A deliberate choice. Technical report, United Nations Development Programme, 2006.

[35] United Nations Development Programme. Mozambique Human Development Report 2007: Challenges and opportunities: the response to HIV/AIDS. Technical report, United Nations Development Programme, 2007.

[36] United Nations Development Programme. Zambia Human Development Report 2007: Enhancing household capacity to respond to HIV and AIDS. Technical report, United Nations Development Programme, 2007.

[37] United Nations Development Programme. Zimbabwe Human Development Report 2003: Redirecting our responses to HIV and AIDS. Technical report, United Nations Development Programme, United Nations Development Programme.

[38] Gilbert R. Modelling Time in Medical Education Research: The Potential of New Flexible Parametric Methods of Survival Analysis. *SciRes*, 3:916–922, 2012.

[39] Moh R., Danel C., Messou E., and et al. Incidence and determinants of mortality and morbidity following early antiretroviral therapy initiation in HIV-infected adults in West Africa. *AIDS*, 21:2483–2491, 2007.

[40] Zachariah R., Fitzgerald M., Massaquoi M., and et al. Risk factors for high early mortality in patients on antiretroviral treatment in a rural district of Malawi. *AIDS*, 20:2355–2360, 2006.

[41] Bennet S. Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2:273–277, 1983.

[42] Lawn S.D., Harries A.D., Anglaret X., and et al. Early mortality among adults accessing antiretroviral treatment programmes in sub-Saharan Africa. *AIDS*, 22:1897–1908, 2008.

[43] ART Cohort Collaboration groups The Antiretroviral Therapy in Lower Income Countries Collaboration. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet*, 23:2375–2398, 2006.

[44] Coates T.J., McKusick L., Kuno R., and et al. Stress reduction training changed number of sexual partners but not immune function in men with HIV. *Am J Public Health*, 79:885–7, 1989.

[45] USAID. HIV/AIDS Health Profile: Sub–Saharan Africa, July 2012. http://www.usaid.gov/our_work/global_health/aids/Countries/africa/hiv_summary_africa.pdf.

[46] WHO, UNICEF, and UNAIDS. Global HIV/AIDS response: epidemic update and health sector progress towards universal access. Technical report, WHO, UNICEF and UNAIDS, 2011.

Appendix

This section contains the various Stata code snippets that were used for the descriptive, regression, imputation and graphing analyses.

/* Analysis Codes. The symbols or characters * and */ */ denote comments about the code */ use "C:\data\Kev thesis\Thesis data clean\Merged_Data.dta", clear

/*set survival data and describe */

stset time_event, failure(AIDS_followup==1)

stdes

stptime, per(1000)

/* Descriptive Statistics*/

* tabulate the categorical variables

foreach var of varlist gender marital_status prev_tb discsp deceased who_baseline1 hgb_category hgb_categ10 BMI_category cd4base_category {

tab2 'var' case, col

}

/* Summarize continuous variables */

foreach var of varlist age weight height BMI hgb_gL hgb_baseline cd4base {

tabstat 'var' , by(case) stats(mean sd median p25 p75)

}

/* CD4 tabulation */

tabstat cd4base , by(case) stats(mean sd median p25 p75)

```
/* sensitivity to knots */

* /Investigate the optimum number of degrees of freedom for the baseline.

forvalues i = 1/5 {

    stpm2 i.hgb_categ1, scale(hazard) df(‘i’) eform

    estimates store df‘i’

    predict h_df‘i’, hazard per(1000)

    predict s_df‘i’, survival

}

/* Univariable and Multivariable Flexible parametric analysis */

/* Flexible parametric model */

/* gender age marital_status height weight prev_tb who_baseline1 cd4base */

xi: stpm2 i.gender , scale(hazard) df(2) eform

stpm2 age , scale(hazard) df(2) eform

/* age per 5 year units */

stpm2 age_5 , scale(hazard) df(2) eform

xi: stpm2 i.marital_status , scale(hazard) df(2) eform

xi: stpm2 weight , scale(hazard) df(2) eform

xi: stpm2 height , scale(hazard) df(2) eform

xi: stpm2 i.prev_tb , scale(hazard) df(2) eform

xi: stpm2 i.who_baseline1 , scale(hazard) df(2) eform

/* CD4 per 100 units */

xi: stpm2 cd4per100, scale(hazard) df(2) eform
```

```
xi: stpm2 i.cd4base_category, scale(hazard) df(2) eform

xi: stpm2 hgb_gL, scale(hazard) df(2) eform

xi: stpm2 i.hgb_category, scale(hazard) df(2) eform

xi: stpm2 i.hgb_categ10, scale(hazard) df(2) eform

xi: stpm2 i.BMI_category, scale(hazard) df(2) eform

xi: stpm2 BMI, scale(hazard) df(2) eform

/* Change reference of BMI category to Normal */

char BMI_category [omit] 2

xi: stpm2 i.BMI_category, scale(hazard) df(2) eform

**

/* Missing data analysis (for AIDS_followup==1 |NO_AIDS_followup==1)*/

misstable tree cd4base if AIDS_followup==1 |NO_AIDS_followup==1

misstable tree BMI if AIDS_followup==1 |NO_AIDS_followup==1

misstable tree age_5 if AIDS_followup==1

misstable tree who_baseline1 if AIDS_followup==1 |NO_AIDS_followup==1

misstable tree hgb_gL if AIDS_followup==1 |NO_AIDS_followup==1

misstable tree _Imarital_s_4 if AIDS_followup==1 |NO_AIDS_followup==1

misstable tree marital_status_coded if AIDS_followup==1 |NO_AIDS_followup==1

misstable tree prev_tb if AIDS_followup==1 |NO_AIDS_followup==1

/** Multivariable Analysis*/

xi: stpm2 i.gender age_5 i.marital_status4 BMI i.prev_tb hgb_gL i.who_baseline1 cd4per100
, scale(hazard) df(2) eform
```

**

/** Explained variation (D method)*/

xi: str2d stpm2 age_5 i.marital_status BMI i.prev_tb hgb_gL i.who_baseline1 cd4per100 , scale(hazard) df(2) eform

**

sts graph, by(stage) censored(single) risktable

/* Explore PH assumption for gender*/

stphplot, by(gender)

estat phtest, plot(gender)

***

/* stcstat2 calculates and reports Harrell's c-index and Somer's D after fitting a model with stpm2.*/

stcstat2

*run command after fitting model

/** Bootstrap procedure for Harrell's c-index and Somer's D */

bootstrap , reps(50) seed(12345): stpm2 i.BMI_category, scale(hazard) df(2) eform

stcstat2

/* Generate quintiles of Hazard score (3 quintiles, appears approriate and well distinguished)*/

xtile hr_quintile3 = h1 , nq(3)

sts graph, by(hr_quintile3)

strate hr_quintile3 , per(1000)

```
/* Predict the baseline survival */

predict surv_baseline, survival ci zeros

/* Perform for HGB category, marital status (1,4) */

/* Obtain predicted values of the survival and hazard functions and plot them */

predict s1, survival

predict h, hazard

predict h1, hazard per(1000)

**

strate marital_status_coded , per(1000)

strate who_baseline1 , per(1000)

stset time_event, failure(AIDS_followup==1)

stdes

stptime, per(1000)

/* Multivariate Analysis */

xi: stpm2 i.gender age_5 i.marital_status4 BMI i.prev_tb hgb_gL i.who_baseline1 cd4per100
, scale(hazard) df(2) eform

predict s1, survival

predict h, hazard

predict h1, hazard per(1000)

/* Setting up dataset and performing the Imputed Data Analysis */

use "C:\data\Kev thesis\Thesis data clean\Merged_Data - Copy.dta", clear
```

```
stset, clear

mi set flong

** vars to be imputed BMI hgb_gL cd4base

mi register imputed cd4per100 BMI hgb_gL marital_status4 prev_tb who_baseline1

mi misstable patterns, frequency

mi impute chained (reg) cd4per100 BMI hgb_gL (ologit) who_baseline1 (logit) marital_status4 prev_tb = age_5 gender_coded ///

, burnin(15) rseed(100) add(10) replace

/* set survival data and describe */

mi stset time_event, failure(AIDS_followup==1)

/* Fit model Survival model on imputed and original data */

* Original data (in imputed set , original data is represented by _mi_m ==0)

xi: stpm2 i.gender age_5 i.marital_status4 BMI i.prev_tb hgb_gL i.who_baseline1 cd4per100 if _mi_m==0 ///

, scale(hazard) df(2) eform

// Imputed data stpm2 model

mi estimate, dots cmdok sav(mi_stpm2,replace): stpm2 i.gender_coded age_5 i.marital_status4 BMI i.prev_tb hgb_gL ///

i.who_baseline1 cd4per100, scale(hazard) df(2) eform

/* Justification for the number of iterations used (15 in our case). Current literature suggests that in many practical

applications between 5-20 iterations (low number of burn-in iterations) are adequate for convergence.
```

```
Also taking consideration that we have few missing values */

/* Generate quintiles of Hazard score (2 quintiles, appeared appropriate and well distin-
guished)*

xtile hr_quintile3 = h1 , nq(3)

xtile hr_quintile2 = h1 , nq(2)

xtile hr_quintile4 = h1 , nq(4)

sts graph, by( hr_quintile3 )

sts graph, by( hr_quintile2 )

sts graph, by( hr_quintile4 )

strate hr_quintile3 , per(1000)

strate hr_quintile2 , per(1000)

strate hr_quintile4 , per(1000)

tabstat h1, by( hr_quintile3) stats(mean sd median p25 p75)

tabstat h1, by( hr_quintile2) stats(mean sd median p25 p75)

tabstat h1, by( hr_quintile4) stats(mean sd median p25 p75)

/* Quantile charts */

xi: stpm2 i.hr_quintile2 , scale(hazard) df(2) eform

predict hq2, hazard per(1000)

twoway (line hq2 _t if hr_quintile2==1, sort) ///

    (line hq2 _t if hr_quintile2==2, sort) ///

    , legend(order(1 "Low Risk" 2 "High Risk") ring(0) pos(1) col(1)) ///

    xtitle("Time since HAART initiation (Days)") ///
```

```
    ytitle("AIDS rate (per 1000 py's)")

/* Kaplan Meier Curve */

sts graph, by( hr_quintile2 )

sts test hr_quintile2, logrank

/* HR and KM by WHO */

xi: stpm2 i.who_baseline1 , scale(hazard) df(2) eform

predict hw1, hazard per(1000)

twoway (line hw1 _t if who_baseline1==1, sort) ///

    (line hw1 _t if who_baseline1==2, sort) ///

    , legend(order(1 "WHO Stage 1" 2 "WHO Stage 2") ring(0) pos(1) col(1)) ///

    xtitle("Time since HAART initiation (Days)") ///

    ytitle("AIDS rate (per 1000 py's)")

/* Kaplan Meier Curve */

sts graph if who_baseline1 <3, by(who_baseline1)

sts test who_baseline1 if who_baseline1 <3, logrank

* BMI_category

xi: stpm2 i.BMI_category , scale(hazard) df(2) eform

predict hb1, hazard per(1000)

twoway (line hb1 _t if BMI_category==1, sort) ///

    (line hb1 _t if BMI_category==2, sort) ///

    (line hb1 _t if BMI_category==3, sort) ///
```

, legend(order(1 "Underweight" 2 "Normal" 3 "Overweight") ring(0) pos(1) col(1)) ///

xtitle("Time since HAART initiation (Days)") ///

ytitle("AIDS rate (per 1000 py's)")

* Kaplan Meier Curve

sts graph , by(BMI_category)

sts test BMI_category, logrank

twoway (line s1 _t if hgb_categ1==1, sort) ///

(line s1 _t if hgb_categ1==2, sort) ///

, legend(order(1 "HGB <=10" 2 "HGB>10") ring(0) pos(1) col(1)) ///

xtitle("Time since HAART initiation (Days)") ///

ytitle("Survival")

**

* BMI_category

xi: stpm2 i.prev_tb , scale(hazard) df(2) eform

predict htb, hazard per(1000)

twoway (line htb _t if prev_tb==0, sort) ///

(line htb _t if prev_tb==1, sort) ///

, legend(order(1 "No TB treatment" 2 "Previous TB treatment") ring(0) pos(1) col(1)) ///

xtitle("Time since HAART initiation (Days)") ///

ytitle("AIDS rate (per 1000 py's)")

* Kaplan Meier Curve

```
sts graph , by(prev_tb)

sts test prev_tb, logrank

twoway (line s1 _t if hgb_categ1==1, sort) ///

    (line s1 _t if hgb_categ1==2, sort) ///

    , legend(order(1 "HGB <=10" 2 "HGB>10") ring(0) pos(1) col(1)) ///

    xtitle("Time since HAART initiation (Days)") ///

    ytitle("Survival")

**

strate marital_status_coded , per(1000)

strate who_baseline1 , per(1000)

twoway (line h _t if who_baseline1==1, sort) ///

    (line h _t if who_baseline1==2, sort) ///

    , legend(order(1 "WHO Stage 1" 2 "WHO Stage 2") ring(0) pos(1) col(1)) ///

    xtitle("Time since HAART initiation (Days)") ///

    ytitle("HIV/AIDS rate (per 1000 py's)") yscale(log)

    /* Baseline Hazard */

predict h0, hazard zeros per(1000) ci

line h0* _t, sort ///

    legend(on) ///

    xtitle("Time since HAART initiation (Days)") ///

    ytitle("Baseline Hazard ratio (Log scale)")

twoway (rarea h0_lci h0_uci _t, sort) ///

    (line h0 _t, sort) ///
```

```
    ,legend(on) yscale(log) ylabel(5 10 25 50 100 250) ///
    xtitle("Time since HAART initiation (Days)") ///
    ytitle("Baseline Hazard ratio (Log scale)")

/* hazard values and graphs on log scale */
predict h2, hazard per(1000)
twoway (line h2 _t if hr_quintile3==1, sort) ///
    (line h2 _t if hr_quintile3==2, sort) ///
    (line h2 _t if hr_quintile3==3, sort) ///
    , legend(order(1 "Q1" 2 "Q2" 3 "Q3") ring(0) pos(1) col(1)) ///
    xtitle("Time since HAART initiation (Days)") ///
    ytitle("HIV/AIDS rate (per 1000 py's)") yscale(log)

twoway (line h1 _t if who_baseline1==1, sort) ///
    (line h1 _t if who_baseline1==2, sort) ///
    , legend(order(1 "WHO Stage 1" 2 "WHO Stage 2") ring(0) pos(1) col(1)) ///
    xtitle("Time since HAART initiation (Days)") ///
    ytitle("HIV/AIDS rate (per 1000 py's)") yscale(log)

/** Measures of the predictive power of the survival model*/
gene invhr=1/h1
gene censind=1-_d if _st==1
somersd _t invhr if _st==1, cenind(censind) tdist transf(c)
somersd _t invhr if _st==1, cenind(censind) tdist transf(z)
***
```

```
** Difference in Hazard Ratios (WHO stage 1 vs. 2
predict hdiff_who2, hdiff1(_Iwho_basel_2 1) ci per(1000)
twoway (rarea hdiff_who2_lci hdiff_who2_uci _t, sort) ///
(line hdiff_who2 _t, sort) ///
,legend(off) /// or on
xtitle("Time since HAART initiation (Days)") ///
ytitle("Difference in hazard rate (WHO stage 1 vs. 2)")
/* Compare the hazard ratios, AIC and BIC from the different models */
estimates table df*, eq(1) se stats(AIC BIC)
/* Predict and plot the baseline hazard function for this model */
/* predict baseline hazard and graph it*/
predict h0, hazard zeros per(1000) ci
line h0* _t, sort ///
    legend(off) ///
    xtitle("Time since diagnosis (years)") ///
    ytitle("Cause specific mortality rate (per 1000 py's)")
/* end */
```