DETERMINANTS OF MORTALITY AND SURVIVAL TIME AMONG HIV INFECTED PATIENTS ON ANTIRETROVIRAL THERAPY IN MACHAKOS COUNTY, KENYA

BY

KEVIN KING'UTU

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

I hereby declare that this thesis has not been presented to any institution either partially or in total for any academic award, publication, or other use. The works here in are original. Where the works of others are quoted, appropriate references has been given.

Author

Kevin King'utu Signature ……Date……...
EL/SMM/00890/2015

Supervisors:

Dr. Otumba E. Ouko, Signature: .....Date: .....
Department of Statistics and Actuarial Science,
Maseno University

Dr. Joyce A. Otieno, Signature: ..... Date: .....
Department of Statistics and Actuarial Science,
Maseno University

DEDICATION

This thesis is dedicated to my father Mr. Richmond Ng'ang'a and mother Mrs. Elizabeth Gikunju, who have always supported my endeavours.

ACKNOWLEDGEMENT

My most sincere appreciation to my mentors and supervisors Dr. Otumba E. Ouko and Dr. Joyce A. Otieno for the guidance and support they accorded me with patience.

I wish to thank the staff of the Mathematics, Statistics and Actuarial Science for the constructive comments, encouragement they showed while teaching us.

I would like to extend appreciation to the Machakos County CEC for health for allowing this study to take place within the County and respective subcouty health managers for allowing me to collect the data.

Lastly, I would like to thank, Dr Jamlick Mutugi, for the valuable contribution and moral and technical support he accorded me throughout the course.

ABSTRACT

Human Immunodeficiency Virus (HIV) remains one of the world's most significant public health concerns. Anti-retroviral treatment (ART) is meant to suppress replication of the virus, consequently improving health outcomes by improving immunity and delaying mortality. Despite policy interventions expanding antiretroviral treatment (ART) services rendering their provision in Kenyan public health facilities free, there exists limited data on survival outcomes in the ART program. Machakos having been listed as a medium epidemic county with no data on ART survival outcomes, it is worthwhile undertaking this study in the area. The aim of this study was to estimate survival time and identify survival predictors of patients on ART using data obtained from government hospitals drawn from five sub-counties in Machakos County. Records of 5,393 adult and paediatric patients initiating ART between January 2011 to December 31, 2015 were included in the retrospective cohort study. Kaplan-Meier methods were used to assess survival probabilities and patterns of HIV-infected patients receiving antiretroviral treatment and log rank tests to compare survival distributions. The Cox-proportional hazard model was used to identify predictors of mortality among ART taking HIV infected patients. The median survival time among ART receiving HIV infected patients was 55 months (95% CI: 51.046-59.986) months with 932 (17.3%) patients having died during the study period. Females exhibited better survival than the males 58months (95% CI: 51.790-59.546) vs. 48 months (95%CI=44.398-53.479) while young adults aged between 20 to 24 years registered lower survival time of 24 months(95%CI: 21.654-28.140). The most important determinants of survival at 5% significance were patients receiving ART in Machakos, patients aged between 20 to 24 years, BMI of less than 18.5, TB infection, advanced disease progression at initiation of therapy (WHO stage 4) and CD4 count less than 100 gm/ul. Based on the findings, it can be concluded that interventions around monitoring of patients receiving treatment in Machakos sub-county, aged between 20 to 24 years ,exhibiting low CD4 cell counts, advanced WHO stages, confection with TB should be carefully tailored to improve the survival of HIV infected patients. Similar studies in the future need to additionally consider ascertaining the status of the lost to follow up patients prior to undertaking survival analysis of HIV infected patients on antiretroviral.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
CD4	Cluster of differentiation 4
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
MLE	Maximum Likelihood Estimate
MOH	Ministry of Health
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organisation
TB	Tuberculosis
TDF	Tenofovir
ZDV	Zidovudine

BASIC CONCEPTS

This section provides certain concepts and definitions that are needed as a basis for the ensuing discussion.

ART: Antiretroviral treatment; comprises of antiretroviral medication which suppresses the HIV virus and hinders the progression of the disease a HIV infected person can live longer without the onset of AIDS-related diseases and occurrence of opportunistic infections (WHO, 2013).

Survival time: is the length of time that is measured from time origin to the time the event of interest occurs. To precisely estimate survival time, an unambiguous time of origin must be defined, a scale for measuring the passage of time must be agreed upon and the definition of event must be clear(Hosmer, Lemeshow, & May, 2008).

Censoring: A condition in which the value of a measurement or observation is partially known. In this context, it is where there exists some information about individual survival time, but the exact survival time is unknown(Hosmer et al., 2008)

There are generally three main reasons why censoring may occur: the HIV infected individual does not die before the study ends, a patient is lost to follow up during the study period or a patient withdraws from the study. The most basic distinction for censoring is between left censoring and right censoring. An observation on a time variable T is right censored if all that is known about T is that it is greater than some value c(where T is typically the time of occurrence for some event and cases are right censored because the observation terminates before event of interest(death due to HIV)occurs).

Left censoring on the other hand occurs when an observation on variable T is less than some value c. It occurs when you begin observing a sample at a time when some individuals may already have experienced the event. (Allison, 2010)

Mortality: a measure of deaths occurrence in a particular population, scaled to the size of that population, per unit of time (WHO, 2013)

Survival analysis: A set of methods for analysing data in which the time until the event is the variable of interest. This response variable is often referred to as failure time, survival time, or event time. It is the positive random variable T, time-to-event. (Hosmer et al., 2008)

Survival Function: the probability that a subject (in this case a HIV infected person) will survive beyond a specified time.

Let T denote the response variable, $T \geq 0$. The survival function is $S(t) = p(T > t) = 1 - H(t)$

The survival function $S(t)$, gives the probability that a subject will survive past time t. As t ranges from 0 to ∞, the survival function has the following properties:

- It is non increasing
- At time $t = 0, S(t) = 1$ meaning the probability of surviving past time 0 is 1
- At time $t = \infty, S(t) = S(\infty) = 0$ - as time goes to infinity, the survival curve goes to 0

The general shape of a survival function for a continuous random variable is as shown below

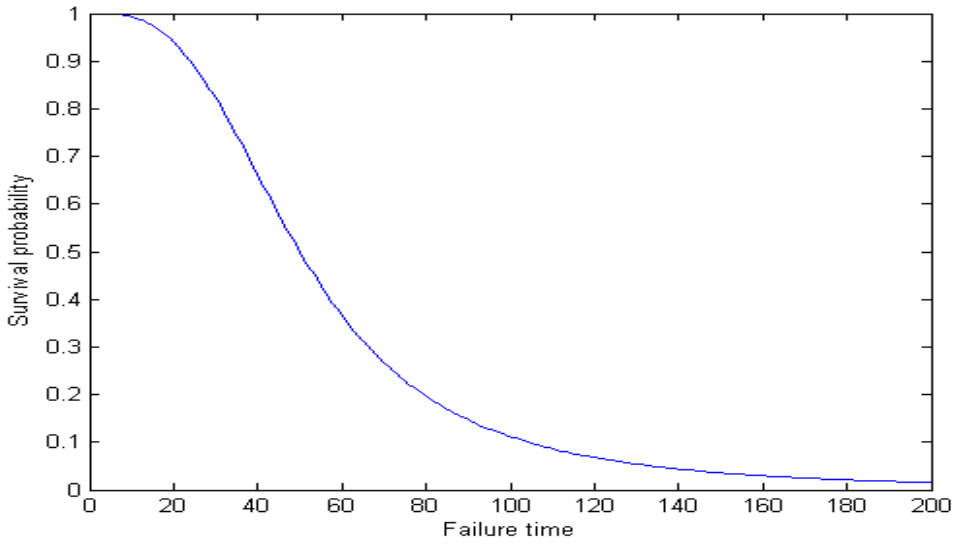


Figure 1: Survival Function

Hazard function: the rate at which a subject is likely to experience the event of interest in the next time interval given that the subject has not experienced the event up until that point (Hosmer et al., 2008). It is the instantaneous rate of occurrence of an event. It is the ratio of the probability density function to the survival function of random variable.

$$h(t) = \left(\frac{P(t < T \leq t + \Delta t | T > t}{\Delta t}\right)$$

For small t

$$P[t < T \leq t + \Delta t | T > t] \approx h(t).$$

Hence the hazard function simplifies to

$$h(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)}$$

$$= -\frac{dlog\{S(t)\}}{dt}$$

That is, the hazard is a conditional density given that death due to HIV has not occurred prior to time t. The hazard function, denoted as H(t) can also be represented in terms of the cumulative distribution function as follows;

$$H(t) = \int_0^t h(t)dt = -log\{S(t)\}$$

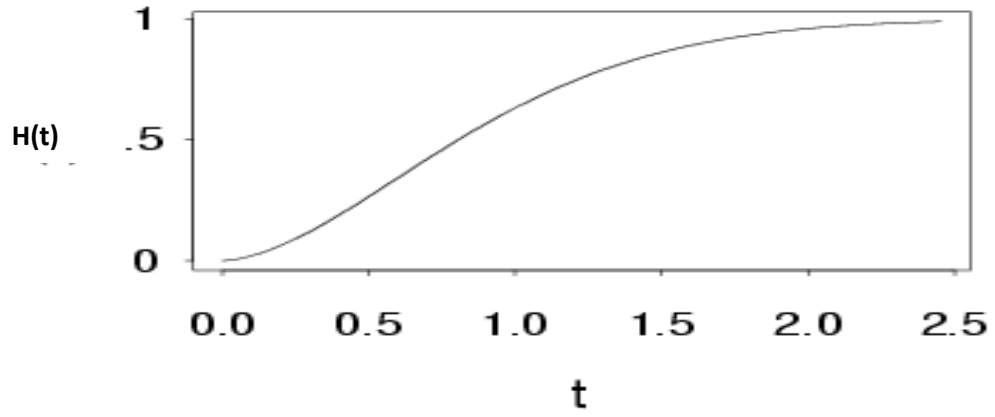


Figure 2: Cumulative Distribution function

As shown in figure 2, the cumulative distribution shows accumulated risk up to time t.

Kaplan-Meier estimator: an estimator of the survival functions from time to event data. It is used to measure the cohort of subjects living for a certain amount of time since commencement of treatment.

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Human Immunodeficiency disease (HIV) is a retrovirus that was first identified in 1983(Sharp & Hahn, 2011). When an individual is infected with the virus, they are termed to be HIV positive and this remains their lifetime status. Infection is mainly via unprotected sex, vertical transmission from mother to child, contaminated blood transfusion and needles. It attacks and kills CD4 lymphocytes leaving the immune system vulnerable to multiple disease infections which are otherwise known as opportunistic infections. Treatment using antiretroviral have shown to significantly slow disease progression(Schomaker et al., 2013)

Globally, it was estimated that in the year 2015, 36.9 million people lived with HIV with sub-Saharan Africa most affected accounting for 25 million people with an estimated 2 million new infections. An estimated 1.2 million HIV-related deaths occurred worldwide by the end of 2014 (WHO, 2015). However, these HIV related deaths showed a decreasing trend compared to 2.2 million deaths in 2005. The decline may be attributed to the increased availability of antiretroviral therapy, as well as care and support to people living with HIV, particularly in low and middle-income countries(UNAIDS, 2010)

Antiretroviral therapy is a combination of antiretroviral drugs used to slow down the rate at which the HIV multiplies in the body. It is a global response to the HIV pandemic. Since the introduction of Zidovudine (ZDV) as the first antiretroviral drug, remarkable strides in antiretroviral treatment have been witnessed. Between 1995 and 1996, the launch of ART combination (Highly Active Antiretroviral Treatment – HAART) was the milestone in the history of HIV treatment that converts HIV infection from what was an inevitable fatal condition into chronic manageable disease. There is notable improvement in ART coverage, attributable to the huge investment and efforts in expansion of ART programmes in low and middle-income countries. Though it is a well-studied and established fact that ART reduces mortality and prevents opportunistic infection among HIV/AIDS patients, less than adequate research has been done on ART programmes in

resource-limited countries such as Kenya to showcase any form of impact in the survival of HIV-infected patients on ART.

The Kenyan Ministry of Health seeks to strengthen HIV prevention, with the ultimate goal of reducing new HIV infections and deaths by 2030(NASCOP, 2014).The HIV epidemic in Kenya exhibits extreme geographical and gender disparities among other factors. National estimates and modelling indicate that 65% of new adult infections occur in nine of the 47 Counties. As at February 2015, over 773,629 patients were on ART (702,000 adults aged 15 years and 71,000 children aged less than 15 years) representing 55% coverage of those in need of ART across the country. The Kenya AIDS Strategic Framework (KASF) 2014/15 -2018/19 has targeted to have at least 90% of those HIV infected access ART by June 2019. Statistics show higher prevalence among women at 7.6% compared to men at 5.6%. There is a treatment gap of over 99,500 women and 64,900 men, in need of antiretroviral treatment but not currently receiving treatment.

Deaths due to HIV have been on the decline with approximately 58,465 HIV related mortalities in 2013 compared to 167,000 in 2003, and this is directly attributable to the wider access to ART and the ability of the National AIDS/STI Control Programme to cover treatment needs for HIV, co infections and provide care services (National AIDS and STI Control Programme, 2015).The deaths are estimated based on evidence on survival time, with and without treatment, and globally recognized methods and models are used to calculate this. HIV related mortality estimates are based on several assumptions among them: estimates of the number of adults and children who are living with HIV and estimates of survival from the time of initiation to treatment to the time of death.

1.2 Statement of the problem

Survival of HIV infected patients receiving Antiretroviral therapy (ART) is dependent on a range of factors that vary with economic, demographic, behavioural and health aspects. These factors affect the treatment outcome over a period of time. Time plays a key role in the outcome of the HIV infected individuals since it's a life-long disease whose treatment outcome is influenced by time. However, there exists less than adequate information regarding treatment outcomes and determinant factors in patient survival time with most research biased on the prevention and factors that increase the likelihood of contracting HIV. Machakos was recently listed as a medium

epidemic county and despite this, the determinant factors of survival among HIV infected on ART have not been studied in the county.

This study investigated the survival time and its predisposing factors among HIV infected patients taking antiretroviral treatment in Machakos County. Survival analysis was performed using the non-parametric survival analysis since the estimate of the hazard function offers great flexibility than most parametric approaches. The nonparametric Kaplan-Meier (K-M) product limit method estimator informed the best predictors for this study and this has been evidenced by Guvercinet et al., (2017). In addition, it has been proven that studies with a follow up period of less than 15 years have a greater power than the log-normal models (Gamel & Vogel, 1997). Thus, the current study was not only able to estimate the survival time but also provided the best predictors of survival among patients who are on ART in Machakos county relative to other studies that did not take time into consideration.

1.3 Objectives of study

Overall Objective

The aim of this study is to estimate survival time and to identify survival predictors of patients on ART in Machakos County.

Specific Objectives

1. To estimate the time to death for HIV infected persons on antiretroviral treatment in Machakos County.
2. Identify the factors associated with mortality among HIV-infected patients receiving ART.

1.4 Significance of the Study

This study provides information about the most influential covariates that have significant impact on survival of HIV infected persons on ART and also identified the risk of mortality of patients under these significant factors at different times during antiretroviral therapy.

1.5 Limitations of the study

- The study was conducted based on secondary data which had incomplete information.

- Part of the information on individuals was missed because of censored observations.
- The study was largely reliant on baseline values of the variables of interest

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Antiretroviral therapy (ART) for HIV-infected patients has been used in the Sub-Saharan African settings for more than a decade. Early analyses of treatment cohorts in this region focused on showcasing the viability and subsequently, the expansion of ART provision in high-burden but resource-constrained setting.

Despite there being apparent evidence of effective use of antiretroviral drug therapy (ART) for treatment of human immunodeficiency virus (HIV) infection, many doubt the feasibility of ART treatment programs in resource-poor settings. Several studies have sought to quantify the survival benefits attributable to antiretroviral therapy (ART) for instance, April et al., (2014) simulated eight cohorts of human immunodeficiency virus (HIV)–infected patients initiating ART each year between 2004 and 2011. Lifetime per capita survival benefits ranged from 9.3 to 10.2 life-years across the 8 cohorts. Total estimated population lifetime survival benefit for all persons starting ART between the study period was 21.7 million life years, with 2.8 million life-years (12.7%) being realized by December 2012. 17.9 million life-years benefited under current policies by 2013, 21.7 million life-years with universal second-line ART implying dramatic past and potential future survival benefits attributable to ART which is comparable to resource poor settings as well as indicating the effectiveness of ART even in future interventions.

The purpose of statistical modelling is to extract information that is meaningful from data collected in surveys and experiments. Critical facets in modelling include description of the data, an explanation of relationship between outcome and predictor variables and the prediction of outcomes. Different models have been used by different authors to study survival. Choice of model is dictated by the type of data collected and, in this section, we review literature by different authors in studying survival.

2.2 Analysis Methods of factors associated with HIV Mortality

Quite a number of studies have been conducted to determine the factors associated with HIV mortality. Getting its treatment has proven to be elusive thus the need to understand what factors are associated with HIV. The study of such factors has been done through use of varying methodological approaches and analytical procedures.

Previous researchers have employed the use logistic regression to assess the factors associated with HIV mortality. Logistic regression is an appropriate regression analysis method to conduct when we have a binary outcome e.g. death. It is used to explain the relationship between one dependent binary variable and other variables that can be nominal, ordinal or interval. The logistic regression models the chance of an outcome based on the individual characteristics and is given by:

$$\log \left(\frac{\pi}{1-\pi}\right) = \beta_0+\beta_1 x_1+\beta_2 x_2+ \ldots + \beta_m x_m \qquad (2.1)$$

where:

π is the probability of an event such as death

β_i are the regression coefficients associated with the reference group

x_i are the explanatory variables.

Since chance is a ratio, the associations are usually reported by use of odds ratios (OR) which is the ratio of probabilities of an event favourable to an outcome and an event against the same outcome. Th application of logistic regression has been widely done in health-related studies that had cohort, case-control and cross-sectional study designs.

For instance, a prospective cohort study was conducted at two urban sites in Lilongwe and Blantyre, Malawi, whereby all patients initiating second-line treatment were included in the study (McNairy et al., 2018). The study utilised logistic regression to identify the socio-demographic, socio-behavioural and clinical factors that were associated with HIV mortality. Factors considered in the model included age, gender, CD4 cell count, WHO stage, duration of first-line ART before presentation, haemoglobin and body mass index. The study found that WHO stage (Odds Ratio (OR) =3.47, 95% CI: 1.14–10.59) and body mass index of 18.5 (OR =4.43, 95% CI: 1.15–17.12)

were risk factors for death. Similar studies that had cohort designs and used the same statistical approaches found similar/ contradicting results.

A retrospective 4 year cohort study was conducted to estimate morbidity and mortality patterns among HIV infected individuals (Ogoina et al., 2012). Potential clinical and demographic variables associated to describe mortality patterns of hospitalised adult HIV/AIDS patients on the ART were compared according to survival status (i.e., deed or survived). An unconditional binary logistic regression analysis checked for model fitness and interactions. Associations were presented by use of odds ratio (OR) with 95% confidence interval (CI) to determine independent predictors of mortality.

Logistic regression analysis has also been used in cross-sectional studies that determined factors associated with HIV mortality. Kamenju & Aboud, (2011) used logistic regression analysis to assess the clinical presentation and estimate the TB/HIV mortality among in-patients. They found out that of 300 TB patients tested for HIV, 175 (58.3%) were HIV infected and 97 (55.4%) of these were already on antiretroviral therapy (ART) at the time of admission. From this, 104 (26.9%) of the TB patients admitted died with two thirds of patients died having PTB. Proportions of death were higher among 13 HIV-infected TB patients (29.1% versus 15.2%) in comparison with the HIV uninfected TB patients ($p = 0.005$).

Conditional logistic regression has also been used to assess factors associated with HIV mortality in case-control studies. This is a specialized type of logistic regression that is usually employed when the case subjects with a certain condition e.g. HIV positive individuals are matched to control subjects without the condition e.g. HIV negative individuals. Assuming there are *k* strata of matched sets and *p* independent variables, the model is given by:

$$logit(p) = \alpha_1 + \alpha_2 z_2 + \ldots + \alpha_s z_s + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2)$$

where:

z_i is the binary indicator variables foe each stratum

α_i is the regression coefficients associated with the indicator variables for the stratum

x_i is the covariate

β_i *is* the population regression coefficients.

Of note, the conditional logistic regression model estimates the population regression coefficients (β_i) but not the stratum indicator regression coefficients (α_i). Associations are also reported by use of OR of each covariate adjusted for the others.

Méda et al., (2013) conducted a case control study to identify the risk factors of active pulmonary TB among HIV infected patients using multiple logistic regression model. The final multivariate model was obtained by a forward and backward variables selection procedure. The results revealed that, after adjustment for potential confounders, an initial weight less than 18.5 kg (OR=4.1; 95% CI: 2.3, 7.4), a CD4 lymphocyte count less than 200 cells/mm3 (OR=9.8‘95% CI: 5.5, 17.5), a WHO clinical stage IV (OR=4.3;95% CI: 2.6, 6.8) and not taking antiretroviral treatment (OR=3.1; 95%CI: 1.9,4.9), were independently associated with the development of active tuberculosis in people living with HIV/AIDS.

Some authors choose to use Loglinear models to do the analysis. For instance (Crellen et al., 2018) used loglinear models to determine factors influencing mortality among HIV infected patients in the Central African Republic. The study found that male sex, age of more than 40 years, and WHO stage 3 and 4 to be significant factors of mortality among HIV infected patients.

HIV/AIDS is a disease that progresses over a time period thus the need to assess the explanatory variables that influence the waiting time. The mentioned studies relied on statistical models that ignore time to event and therefore fail to include the exposure to the risk of the event over time. Though the independent factors were almost similar across the various studies with various population settings, it is important to understand how the time-factor plays a role for an event to be experienced based on individual characteristics, more so for HIV related studies.

2.3 Survival Analysis and its Application to factors associated to HIV mortality

Better estimates of HIV mortality can be found when considering time as an influence on the time-to-event outcomes. Cox proportional hazard model by Cox (1972), are widely used to deal with time to event data and their relevancy on research in survival analysis in demography and related

fields has increased over the years. The Cox model has several advantages and some of them are; (a) ability to include analysis of censored and truncated data (b) ability to include analysis of time varying covariate effects and lastly (c) the extensions of the Cox regression models with the inclusion of random effects and flexible modelling through semi-parametric and non-parametric approach.

The cox proportional hazard model assesses the impact of life-time related factors on the hazard function. The hazard function for a given individual *i* given a set of *p* covariates $xi = (x_{1i}, x_{2i},.x_{pi})$ at time *T* will be given as:

$$h_i(t, x_i) = h_0(t)e^{(b^T x_i)} \quad (2.3)$$

Where

$$b^T x_i = b_1 x_{1i} + b_2 x_{2i} + \cdots + b_p x_{pi}$$

The probability of surviving longer than the defined time period are estimated by use of Kaplan-Meier plots in survival analysis. Survival analysis has been used to assess the determinants of mortality among HIV infected patients. Socio-demographic, behavioural and clinical factors have been found to be associated with HIV mortality.

2.4 Factors associated with HIV mortality

2.4.1 Sociodemographic Factors

Sociodemographic characteristics such as age, gender and education have been found to have a strong association with HIV survival. A retrospective study done in India assessed the survival among adult HIV-infected patients on ART (Bajpai et al., 2016). Cox regression was done, and higher mortality was observed in older aged patients relative to younger aged patients (adjusted HR (aHR) =1.43 for 30-39 years, aHR = 1.86 for 40-49 years and aHR =1.93 for ≥50 years). Higher proportion of deaths were observed among males compared to females (aHR =12.1, 95% CI:11.6-12.6). Limitations in this study can be associated with the study design which used only demographic characteristics ignoring the potential clinical confounders such as CD4 cell count, HIV staging, viral load, co-morbid coinfections that could have an influence on the survival of individuals. Other studies examining the determinants of mortality among adult HIV patients in

Nigeria and Tanzania used a similar methodology and found that significant independent predictors of mortality include age more than 45 years and being male (Eguzo et al., 2014;Gunda et al.,2017).Alvarez-Uria et al., (2013), further found that patients living in or near an urban centre had a reduced risk of mortality.

2.4.2 Clinical Factors

Some clinical variables like haemoglobin level, advanced HIV disease, malnutrition, TB and other co-morbidities have been shown to be important predictors of mortality among patients on ART. Studies by Kebebew, (2012) , Betre & Ameni, (2016) and Seyoum et al., (2017)aimed at estimating mortality rate and to identify survival predictors of patients taking ART in Ethiopia. Using Cox regression model, the investigators identified the most important predictors of mortality at 0.05 level of significance as; low CD4 cell count at baseline, employment status, functional status, WHO clinical stages III and IV, TB coinfection, and opportunistic infections cohort approaches. A retrospective cohort study conducted on children in Ethiopia using Kaplan Meier survival and Cox proportional hazard model to identify independent predictors of children's mortality on ART found significant contribution by anaemia to HIV mortality (Gebremedhin et al., 2013).

Mageda et al., (2012) used Cox regression analysis to identify predictors of mortality among HIV/AIDs patients attending a care and treatment clinic in Tanzania. The study found males having three times higher mortality compared to females, body weight below 45 kgs, WHO stage 4 disease, and CD4 cells below 50 as significant predictors of mortality.

Time to ART initiation is another main covariate associated with patient mortality in HIV. Wada et al., (2014) examined the relationship between time from diagnosis to time of HAART initiation and mortality using parametric mixed models to compare proportions of HIV related and non-HIV related death stratified by age as well as cox proportion hazards for cause specific hazards of mortality at ART initiation. Findings indicated that, clients initiated early into ART had higher likelihood of death from non-HIV causes, and an increased life span, compared to late initiators. Also, results pointed towards non-HIV mortality hazards among early initiators may approximate those faced by comparable HIV-uninfected individuals. Adjusted results from mixture models revealed lower ages at non-HIV related death for even early HAART initiators relative to HIV-uninfected individuals. However, comparative to HIV-uninfected subjects, HAART-treated

individuals had characteristics associated with higher mortality. Bhatta et al., (2013) used Kaplan-Meier and Cox-regression models to estimate survival and explore determinants of mortality among adult HIV-infected patients on ART in Far-western region of Nepal. Among his findings was that risk of motarlity was three times higher in the first three months after ART initiation.

In South Africa, a study was done to determine the relationship between mortality risk and the CD4 cell response to antiretroviral therapy with patients observed for up to 5 years of ART. Findings showed that despite swift immune recovery, high mortality in the first year of therapy was related to the large proportion of failures accrued within CD4 cell-strata less than 200 cells/ml. Furthermore, patients with baseline CD4 cell counts below 100 cells/ml had much higher cumulative mortality at one and 4 years (11.6 and 16.7%) as opposed to those of patients with baseline CD4 cell counts of at least 100 cells/m (Lawn, et al., 2009).

Cotrimoxazole prophylaxis has been found to significantly improve the survival probability of for patients with advanced HIV (CD4<200 cells/ mm^3 or WHO stage 3 or 4) on HAART. To assess the impact that cotrimoxazole has on HIV infected clients, CJ Hoffmann et al, (2011) used cox proportional hazards and modelled mortality rates in the first year of ART initiation in 14097 clients. Findings indicated a 36% reduction in mortality (aHR = 0.64, 95% CI: 0.57, 0.72) among HIV infected clients receiving cotrimoxazole with CD4 values of less than 200 cells/ mm^3or WHO stage 3 or 4. However, cotrimoxazole prophylaxis did not reduce the hazards for mortality for patients with CD4 counts above 200 or WHO stage 1 or 2.

Summary of Literature Review

Better estimates of HIV mortality can be found when considering time as an influence on the time-to-event outcomes with Cox proportional hazard model by Cox (1972) being widely used to deal with time to event data. The investigators in the literature reviewed have attributed age, gender, education, proximity to ART clinic as some of the demographic factors influencing mortality while TB/HIV coinfection, baseline nutrition status, baseline WHO clinical stage, baseline CD4 count, baseline ART regimen, adverse drug reactions, cotrimoxazole prophylaxis, opportunistic infections, haemoglobin status, and adherence to treatment as some of the main clinical factors influencing overall efficacy of treatment and eventual survival and retention in antiretroviral therapy.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter describes the methods used to collect the data, the study area, study population, study design, sampling procedure, data collection methods, data management and ethical considerations.

3.2 Study Area

The study was conducted in five sub-County hospitals (Machakos Level five, Mwala, Kathiani, Kangundo and Athi River) in Machakos County. The study area was chosen because it hosts more than 80% of the HIV infected persons receiving antiretroviral treatment and care and offer various HIV services inclusive of HIV testing and counselling, prevention of mother to child transmission, family planning and reproductive health, tuberculosis, and other opportunistic infections.

3.3 Study Population

This study population included all HIV infected patients enrolled in the 5 sub County hospitals between 2011 and 2015.

3.4 Study Design

This was a retrospective cohort analysis of HIV infected patients on ART who were enrolled between 2011-2015. Characterization of patients on ART was conducted from health facilities with an electronic medical record system in place.

3.4.1 Inclusion criteria

- Subjects under study must be HIV positive adults and children on antiretroviral therapy
- Must have been initiated on ART and receiving ART for a period not less than 1 year.

3.4.2 Exclusion criteria

- Missing HIV infection year and/or month
- Missing ART start year and/or month

- Missing HIV patients' death year and or month
- ART start date being before 2011 or after 2015

3.5 Sampling procedure

All patients enrolled in antiretroviral therapy between 2011 and 2015 and met the inclusion criteria were included in the study. These patients were then retrospectively followed up through to the end of 2015 maximum follow up time being five years.

3.6 Data Collection and Management

Preceding the analysis, participating facilities cross checked their patient records to verify accuracy to the best of their knowledge. Data was extracted from facility electronic medical records during the study and included patient level characteristics such as age, sex, location etc. This data was used in the analysis to determine survival time and determinants of mortality.

3.6.1 Study Variables

Since this study aims to identify the determinants of mortality among HIV infected patients on ART and estimate their survival time, the following variables were collected from patient records;

Outcome variables*:* These include; death while on antiretroviral treatment, and patient follow-up status (Active, dead)

Independent variables include;

Demographic characteristics: This include; location of ART centre(urban/rural), age at initiation of treatment and sex.

Clinical characteristics: TB infection during ART, nutritional status (measured via body mass index, baseline WHO clinical stage, baseline CD4 count, baseline ART regimen, cotrimoxazole prophylaxis and haemoglobin levels.

3.6.2 Data Security and Confidentiality

This data was stripped off all patient identifiers and stored in a password encrypted computer whose access was restricted to the principal investigator.

3.7 Survival Analysis

Survival analysis was used to assess the factors associated with HIV mortality. Survival analysis refers to a set of statistical procedures for analysis of data where the outcome variable of interest is time until an event occurs. The study time variable was quantified in months and the study's outcome was death

3.7.1 Censoring

The use of survival analysis, as opposed to the use of other statistical methods, is most important when some subjects under investigation do not experience the event of interest over the study period. These incomplete observations are referred to as being censored and can be left, right or interval censored. In this study, the collected data was right censored since some of the patients had experienced the event (death) before the lapse of the study.

3.7.2 Survival Modelling

Survival modelling is concerned with models for data that have three main characteristics which include the dependent variable, censoring and the explanatory variables. This study had death as the dependent variable, it was right censored and the explanatory variables consisted of the socio-demographic, behavioural and clinical characteristics of the patients

3.7.3 Descriptive methods for survival data

As an initial step in the analysis of survival data, graphical summaries of survival time were analysed and presented since routine measures of central tendency and variability will not yield estimates of the desired parameters when the data includes censored observations. The two key functions in analysis of survival data used were the survivor function and the hazard function. Assuming that T is the survival time, non-negative valued random variable, the value of T for this study was the time from start of treatment up to the experiencing of an event (i.e. death or censored) occurred. The study survival time, T, was quantified in months.

3.7.4 Survivor function

The survivor function refers to be the probability that the survival time of a person is greater than or equal to some specified time(t) and is denoted by S(t). It gives the probability of an individual surviving beyond a specified time. Suppose that the random variable T has a cumulative probability distribution, H*(t)*, with underlying probability density function *f (t)*. The survival function *S(t)* is then given by:

$$S(t) = p(T > t) = 1 - H(t) \qquad \ldots\ldots. (3.1)$$

The Cumulative distribution function H(t) is given as;$H(t) = p(T \leq t)$,then the survival function

$$S(t) = p(T > t) = \int_t^{\infty} h\,(u)du$$

$$=1 - \int_0^t h\,(u)du$$

$$=1 - H(t)$$

Theoretically, as *t* ranges from 0 to infinity, the survivor function can be graphed as a smooth curve. The study data collected had a time *t* measured in months and ranged from of $0 \leq t \leq 60$.

3.7.5 The Hazard function

The hazard function, denoted by *h* (*t*) is the instantaneous potential for failing at time *t*, given that the individual has survived up to time *t.* As opposed to the survivor function (which focuses on failure) the hazard function focuses on not failing.

The hazard function $h(t) \geq 0$, is given as:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t}{\Delta t}$$

and the survival function in terms of the hazard function is given by;

$$S(t) = e^{-\int_0^t h(u)du} = e^{-H(t)} \quad \ldots\ldots (3.2)$$

3.7.6 Non-Parametric Estimation of survival function

The study used the non-parametric survival analysis method as they are distribution-free since they do not require specific assumption to be made about the underlying distribution of the survival times.

3.7.7 Kaplan-Meier estimator

The Kaplan-Meier (K-M) estimator of the survivorship function, also called the product limit estimator, is the estimator used by most software packages. It is mostly used to estimate the survivor and hazard functions and incorporates information from all of the observations available(both censored and uncensored) by considering any point in time as a series of steps defined by the observed survival and censored times(Hosmer et al., 2008).

Suppose $t_1, t_2.., t_n$ denotes the survival times of an independent observations and $\delta_1, \delta_2.., \delta_n \; where \; \delta_i = 1$ event/death and $\delta_i = 0$ for censored observations. To obtain the K-M estimator of the survival function ,we first order the survival times as $t_1 < t_2.. < t_n$ .Assume that among the n observations there are $m \leq n$ failures occurred at distinct m times. Then the ranked failure times $t_1 < t_2.. < t_m$.The Kaplan-Meier estimator of the survivorship function (or survival probability) at time t, $S(t) = p(T \geq t)$ is defined as

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i} = \prod_{t_{(i)} \leq t} (1 - \frac{d_i}{n_i}) \quad \ldots\ldots (3.3)$$

where d_i =number of failure(death) at time t_i and n_i = number of individual at risk of failure(death) at t_i and with the convention that $S(t) = 1$ for $t < t_1$

The variance of the Kaplan-Meier estimator of the survival function (which is also known as the greenwood formula) is given as;

$$Var(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \quad \ldots\ldots (3.4)$$

Confidence intervals can be computed based on (3.4) .These however may extend above one or below zero. A more satisfying approach is to find confidence intervals for the complementary log-log transformation of $\hat{S}(t)$ as follows,

$$Var\left(log\ log\left[-log\,\widehat{S}(t)\right]\right) = \frac{1}{[log\ \hat{S}(t)]^2} \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)} \quad \ldots. (3.5)$$

3.7.8 Median Survival and the Confidence Interval for the Median

The median survival time is defined as $\hat{t}_{med} = inf\ \{\ t: \hat{S}(t) \leq 0.5\}$; that is, it is the smallest t such that the survival function is less than or equal to 0.5

To find a $1-\propto$ confidence interval for the median, consider the inequality;

$$-z_{\frac{\propto}{2}} \leq \frac{g\{\hat{S}(t)\} - g(0.5)}{\sqrt{var[g\{\hat{S}(t)\}]}} \leq z_{\frac{\propto}{2}}$$

where g ($\widehat{S}(t)$)= $log\ log\left[-log\,\widehat{S}(t)\right]$ and var [g $\{\hat{S}(t)\}$] is given by (3.5)

The study used the formula to obtain the 95% confidence interval for the median survival time. To obtain a 95 % confidence interval, we searched for the smallest value of t such that the middle of the expression is at least -1.96 (for the lower limit) and the largest value of t such that the middle expression does not exceed 1.96 (for the upper limit)

3.7.9 Comparison of Survivorship Functions

After obtaining statistics which provide a description of the overall survival experience, the study compared the survivorship experience between sub groups in the survival data. These groups might be defined by the values of a covariate which are thought to influence survival times. The K-M estimator of the survival function was plotted for each group. That is, plotting the corresponding

estimates of the two (or more) survivor functions on the same axes of the K-M estimator. For plots exhibiting the pattern of one survival function lying above another, then the group defined by the upper curve had more favourable survival experience than the group defined by the lower curve.

Due to censored observations, standard tests such as the one-way analysis of variance, the t -test for the comparison between groups could not be applied directly. There however exits tests for comparing survival functions. Some of these are the log-rank test, generalized Wilcoxon test, Taron-Ware test, Peto-Prentice test, and Harrington-Fleming test. All these are dependent on ratios of weighted survival experience.

The calculation of each test is based on a contingency table of group by status at each observed survival time. The contribution to the test statistic depends on which of the various tests is used, but each may be expressed in the form of a ratio of weighted sums over the observed survival time. For comparison of survival functions between two groups the test statistic Q may be defined as;

$$Q = \frac{(\sum_{i=1}^{m} w_i(d_{1i} - \hat{e}_{1i}))^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}}, where\ \hat{e}_{1i} = \frac{n_{1i}d_i}{n_i}\ and\ \hat{v}_{1i} = \frac{n_{0i}n_{1i}d_i(n_i - d_i)}{n_i^2(n_i - 1)} \quad \ldots\ldots(3.6)$$

m the number of rank-ordered failure (death) times
n_{0i} is the number of individuals at risk at observed survival time t_i in group 0
n_{1i} is the number of individuals at risk at observed survival time t_i in group 1
d_{0i} is the number of observed deaths in group 0
d_{1i} is the number of observed deaths in group 1
n_i is the total number of individuals or risk prior to time t_i
d_i is the total number of deaths at time t_i
w_i is the weight for censor adjustment at failure time t_i

With the null hypothesis that the two survivorship functions are the same, and assuming that the censoring experience is independent of group, and that the total number of observed events and the sum of the expected number of events is large, then the significance level for Q follows a chi-square distribution with one degree-of- freedom.

The log rank test, a special case of Q that is based on weights equal to one,($W_i = 1$),was used to make interpretation in this study. It is given by;

$$Q = \frac{(\sum_{i=1}^{m}(d_{1i} - \hat{e}_{1i}))^2}{\sum_{i=1}^{m} \hat{v}_{1i}} \qquad (3.7)$$

3.7.10 Regression Models for Survival Data

An objective for modelling survival data is to determine which combinations of potential covariates affect the form of the hazard function. Modelling the hazard function also helps to obtain an estimate of the hazard function for an individual from a set of covariates.

In survival analysis the risk of hazard of failure at any time after the time origin of the study is where the focus lies and as such, the hazard function is thus modelled directly in survival analysis. The model for modelling survival data in this study is the proportional hazard model.

3.7.11 The Cox Proportional Hazards Regression Model

The Cox Proportional Hazard (PH) Model is a multiple regression method and is used to evaluate the effect of multiple covariates on the survival. The covariates might be time-dependent or independent and hence, it is also referred to as a semi-parametric model. Semi-parametric models parametrically specify the functional relationship between the lifetime of an individual and their characteristics but leave the actual distribution of lifetimes arbitrary and thus, does not require the imposition of the probability density function of a parametric distribution. This means that Cox's semi-parametric modelling allows for no assumptions to be made about the parametric distribution of the survival times, making the method considerably more robust(Smith & Ryan, n.d.). However, it the ascertainment that the assumption that the hazards are proportional over time must be done. The proportional hazards assumption refers to the fact that the hazard functions are multiplicatively related. That is, their ratio is assumed constant over survival time, which means that the risk of failure is the same no matter how long the subject has been followed.

It allows the addition of covariates, while keeping the baseline hazards unspecified and can take only positive values. It is defined as;

$$h(t, X, \beta) = h_0(t)e^{\beta' X} \qquad \dots (3.8)$$

which is the hazard function at time t with covariates $X = (X_1, X_{2\dots\dots}X_P)'$

$h_0(t)$ is the arbitrary baseline hazard function that characterizes how the hazard function changes as a function of survival time.

$\beta = (\beta_1, \beta_{2\ldots\ldots}\beta_P)'$is a column vector of p regression parameters associated, with explanatory variables?

$e^{\beta' X}$characterizes how the hazard function changes as a function of subject covariates and *t* is the failure time.

With everyone having independent hazard function of survival time, the model becomes;

$$h(t, x_i, \beta) = h_0(t)e^{\beta_1 x_{i1} + \beta_2 x_{i2} \ldots\ldots\ldots + \beta_p x_{ip.}}, \text{ i} = 1,2\ldots\ldots n \qquad \ldots(3.9)$$

with *n* being the total number of observations in the study and $x = (x_{i1}, x_{i2\ldots\ldots}x_{iP})'$ is a column vector of measured covariates for the i^{th} individual (patient) which are assumed to affect the survival probability.

3.7.11.1 Assumptions of Cox proportional hazard model

1. The baseline hazard function $h_0(t)$ depends on *t*, but not on covariates $x_1, x_{2\ldots\ldots}x_P$.
2. The hazard ratio, e^{β} , depends on the covariates $X = (X_1, X_{2\ldots\ldots}X_P)'$,not on time.
3. The covariates *X i* are time-independent.

To demonstrate the assumption that the hazard ratio depends on the covariates and not on time (as stated in assumption 2), consider two distinct values,x_{i1} and x_{i2} ,of a continuous covariate *X*;

$$h(t, X, \beta) = h_0(t)e^{\beta' X}$$

Then, the hazard ratio, which is independent on time becomes;

$$\frac{h(t,x_1,\beta)}{h(t,x_1,\beta)} = \frac{h_0(t)e^{\beta_i x_{i1}}}{h_0(t)e^{\beta_i x_{i1}}} = e^{\beta_i(x_{i1} - x_{i2})} \qquad \ldots (3.10)$$

which reveals that the ratio of the hazard functions for two individuals with different covariate values does not vary with time.

3.7.11.2 Estimation of Parameters in proportional hazard model

Regression coefficients in the proportional hazards model can be estimated using the maximum likelihood method. The likelihood function, denoted as L ,is a mathematical expression which describes the joint probability of obtaining the data observed on the subjects in the study as a function of the unknown parameters (β) in the model being considered.

L is written notational as L(β) where β denotes the collection of unknown parameters. In Cox proportional hazards model, we can estimate the vector of parameters β without having any assumptions about the baseline hazard $h_0(t)$.

Consider an independent individual, the data for the Cox proportional hazard model is represented by;

$(t_i, \delta_i, x_i), i$=1,2……n, *where*;

t_i is the survival time for i^{th} individual

δ_i an indicator of censoring for the i^{th} individual given by 0 for censored and 1 for event/death

x_i a vector of covariates for individual i $(x_{i1}, x_{i2……}x_{iP})$

Then the full maximum likelihood is given as;

$$L(\beta) = \prod_{i=1}^{n} h(t_i, x_i, \beta)^{\delta_i} S(t_i, x_i, \beta) \qquad ...(3.11)$$

where $h(t_i, x_i, \beta) = h_0(t_i)\, e^{(\beta' x_i)}$ is the hazard function for individual i

$S(t_i, x_i, \beta) = (S_0(t_i))e^{\beta' x_i}$ is the survival function for individual i

Thus, the full maximum likelihood function becomes

$$L(\beta) = \prod_{i=1}^{n} (h_0(t_i)\, e^{(\beta' x_i)})^{\delta_i} (S_0(t_i)) e^{\beta' x_i} \qquad \ldots.(3.12)$$

The full maximum likelihood requires that we maximize (3.11) with respect to the unknown parameter of interest β and unspecified baseline hazard and survival functions. This indicates that to obtain the maximum likelihood estimators for the full likelihood, the baseline hazard, $h_0(t)$,needs to be explicitly specified. However, using the partial likelihood function that depends only on the parameter of interest is recommended (Cox, 1972). The partial likelihood is used because the likelihood formula considers probabilities only for those subjects who fail, and does not explicitly consider probabilities for those subjects who are censored

3.7.11.3 Partial likelihood

Suppose *m* failures occur out of *n* subjects with $m \leq n$ and that $t_1 \leq t_2 .. \leq t_m$ are the m distinct ordered failure times observed, and let R_i be the set of subjects at risk prior to time t_i. Assume that there is only a single failure at time t_i

Then P (individual *i* has experienced an event at time t_i| one event at time t_i)

$$= \frac{h(t,x_i)}{\sum_{j \in R_{t(i)}} h(t,x_j)} = \frac{h_0(t) e^{\beta' x_i}}{\sum_{j \in R_{t(i)}} h_0(t) e^{\beta' x_j}} = \frac{e^{\beta' x_i}}{\sum_{j \in R_{t(i)}} e^{\beta' x_j}} \qquad \ldots (3.13)$$

When that there are no tied times assumed, the partial likelihood is defined over all failure time t_i that *i= 1, 2,..,m* and given as

$$L_p(\beta) = \prod_{i=1}^{m} \frac{e^{\beta' x_i}}{\sum_{j \in R_{t(i)}} e^{\beta' x_j}} \qquad \ldots.. \ (3.14)$$

where the product is over *m* distinct ordered failure times and x_i denotes the value of the covariate for the subject with ordered survival time t_i. The log partial likelihood function is;

$$L_p(\beta) = \sum_{i=1}^{m} [\ \beta' x_i - \ln\ (\sum_{j \in R_{(ti)}} e^{\beta' x_j})\ ] \quad \ldots(3.15)$$

We obtain the maximum partial likelihood estimator by differentiating the right-hand side of (3.15) with respect to the component of β, setting the derivative equal to zero and solving for the unknown parameters. However, this partial likelihood function methods assume that there are no tied values among the observed survival times. In most real situations however, tied survival times are more likely to occur. In addition to the possibility of more than one death at a time, there might also be more than one censored observation at a time of death.

Several approaches to handle tied data have been suggested and, of these, three are used by software packages: an exact expression that is derived in Kalbfleisch & Prentice, (1980) and approximations due to Breslow (1979) and Efron (1977). However, approximations derived by Breslow and Efron are designed to provide expressions that are more easily computed than the exact partial likelihood, yet that still account for the fact that ties are among the observed values of survival time. In many applied settings there is little or no practical difference between the estimators obtained from the two approximations and since the Breslow approximation is more commonly available and popular it is mostly used (Hosmer et al., 2008).

3.7.11.4 The Breslow approximation

This approximation is proposed by Breslow and Peto to modify the partial likelihood and has the form

$$L_B(\beta) = \prod_{i=1}^{m} \frac{e^{\beta' s_i}}{[\sum_{l \in R_{t(i)}} e^{\beta' x_l}]^{d_i}} \quad \ldots(3.16)$$

where d_i is the number of deaths occurred at time t_i

s_i the sum of covariates over d_i subjects at time t_i

Then, the partial log of (3.16) thus given as

…(3.17)

$$L_B(\beta) = \sum_{i=1}^{m}[\beta' s_i - d_i ln \sum_{l \in R_{t(i)}} e^{\beta' x_l}]$$

The study used the Breslow maximum partial likelihood estimator, adjusted for tied observation. It is obtained, by differentiating equation (3.17) with respect to the components of β and setting the derivative equal to zero and solving for the unknown parameters.

3.7.12 Model development

It is likely that to have data on more covariates can reasonably be included in a model. This calls for a method to select a subset of the total number of covariates. During this selection, issues such as clinical importance and statistical significance need to be considered. There are three methods of selection of influential covariates. These are purposeful selection, stepwise selection (forward selection and backward elimination) and best subset selection. Survival analysis using Cox regression method begins with a thorough univariable analysis of the association between survival time and all-important covariates.

Hosmer et al.,(2008) recommends the inclusion of all variables that are significant in the univariable analysis at the 25 percent level and any other variables which are presumed to be clinically important to fit the initial multivariable model. Survival analysis using cox regression model begins with bivariate analysis of the association between survival time and all-important covariates.

All variables that were significant at 25% level are included, the modest level of significance from one explanatory single covariate regression model are then taken into initial multivariable model. The stepwise statistical method is used to select significant explanatory variables. In forward variable selection method, variables are added successively (the most significant at each step) until no variable adds significant information. The current study added all explanatory variables that were significant at 0.25 into the multivariable model. Statistical significance of the explanatory variables was considered for $p < 0.05$.

3.7.13 Model diagnostics for Cox PH model

The preliminary final model which fulfils the model development stages will not be identified as the final model until it has been critically examined for the adherence to key assumptions (i.e. proportional hazards and the presence of undue influence or outliers on the fitted model).

3.7.14 Checking Cox Proportional Hazard Assumption

In order to use the Cox model, it must be checked that the assumption of whether the effects of covariates on hazard ratio remain constant over time. This is a vital assumption of proportional hazards model and must be assessed for each covariate. The Schoenfeld residuals graphical technique can be used to assess Cox model proportionality assumption.

The technique is based on individual contributions to the log partial likelihood and measures the difference between the covariate for the i^{th} individual and a weighted average of the covariate over the risk set at each event. To check the proportionality assumption for each covariate, we plot the scaled Schoenfeld residuals against log of survival time. If the proportional hazards assumption is satisfied, the distribution of residuals over time is random, that is, does not show a particular trend, and the smoothed plot called Locally Weighted polynomial regression line summarizing the residuals should be a straight line and close to the horizontal reference line. Formal tests need to detect any time dependency in particular covariates, after allowing for the effects of explanatory variables that are known. Testing the dependency of covariates on time is equivalent to testing for a non-zero slope in a generalized linear regression of the scaled Schoenfeld residuals on functions of time. A non-zero slope is an indication of a violation of the proportional hazard assumption. The Grambsch-Therneau test of non-proportionality uses partial residuals for the test of proportional hazards assumption. To use this test for the i^{th} covariate, Grambsch & Therneau, (1994) proposed a time-varying coefficient as;

$$\beta_i(t) = \beta_i + \gamma_i g_i(t) \qquad (3.18)$$

where $\beta_i(t)$ is time varying coefficient β_i is constant and $g_i(t)$ is some specified function of time, usually $g_i(t) = ln\,(t)$

Then, the Cox proportional hazard model for time varying coefficient with $g_i(t) = ln\,(t)$ is defined as;

$$h(t, x_i\beta_i(t)) = h_0(t)e^{\beta_i(t)x}$$

by substituting for $\beta_i(t)$ and $g_i(t)$;

$$= h_0(t)e^{\beta_i+\gamma_i ln\,(t)x}$$

$$= h_0(t)e^{\beta_i x+\gamma_i (ln\ t)x} \qquad (3.19)$$

Equation (3.19) is the proportional hazards model with the interaction term, $x \ln(t)$ and main effect x_i.

To test the significance of the interaction term $x \ln(t)$, the study performed the test:

$H_0: \gamma=0$ versus: $H_0: \gamma \neq 0$ and used the Wald test.

3.8 Statistical Analysis

Data exploration, manipulation, organization and statistical analysis was done using excel and statistical software R 3.2.2. Independent-sample T-test (for continuous variables) and Chi-square test (for categorical variables) were used to explore significant difference and associations in patient's characteristics. Kaplan-Meier methods were used to assess survival probability and patterns of HIV-infected patients receiving antiretroviral treatment and log rank tests to compare survival curves. The Cox-proportional hazard model was used to analyse the hazard ratio determinants of survival for patients on ART over the study period.

3.9 Ethical Consideration

All data collection and review procedures observed strict human subject protection measures. Ethical clearance was obtained from the Ethical Review Board of Maseno University. Permission to use patient level electronic medical data from select facilities was obtained from the office of the chief executive for health Machakos County. Verbal informed consent was also obtained from responsible bodies of the hospitals and ART clinic prior to data collection. Encrypted hard disks shall be used in the collection of datasets from the health facilities. The EMR databases had all patient identifiable information de-identified hence confidentiality and privacy of the information was guaranteed and maintained.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Introduction

Here we present results from data exploration, analysis, interpretation and discussion. We first explored summary statistics of various variables considered in this research study. Descriptive analysis was used in comparing survival time in different strata and finally the adequacy of the survival model is investigated. Based on these findings, results are discussed and interpreted.

4.2 Summary Statistics

Between January 1 2011 to December 31 2015, 5,393 adult and paediatric patients who were initiated on ART in Machakos County met the inclusion criteria for this study among them, 3,289(61.0%) were female. Majority of the patients 3,871 (71.8%) were >25 years old at ART Start with 2,643(49.0%) of the cohort being from Machakos sub county. Among patients who had CD4 count data available 51.8% had CD4 counts less than 100 cells/µl at ART initiation with 987 (24.1%) of the patients starting treatment while in advanced disease stages at WHO stage 3 and 4. Among patients with BMI results available,1,831(52.5%) were within the normal range of 18.5-25.0. About 504(9.3%) had active TB or were suspected to have TB. The most prescribed ART regimen at treatment initiation were the TDF based regimes at 3,834(71.1%) patients. (Appendix A, Table 1)

4.3 Descriptive Survival Analysis

The median survival time among ART receiving HIV infected patients was 55 months (95% CI: 51.046-59.986) months. Among the patients, 932(17.3%) died during the study period (within 5 years of ART initiation), with 498 (53.4%) of the deaths occurring in the first 12 months (Appendix A, Table 2).

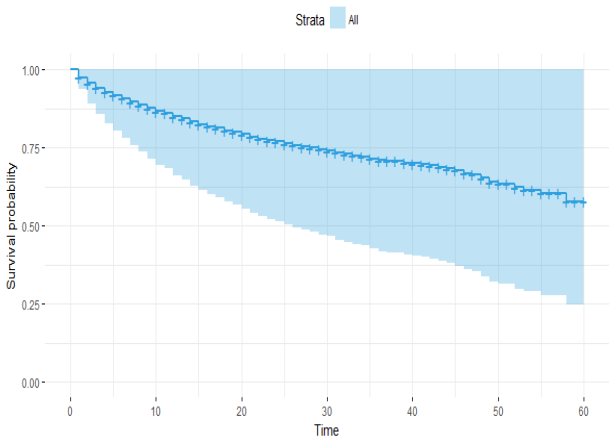


Figure 3: Overall Plot of Kaplan-Meier survivor function among HIV infected patients on ART in Machakos County

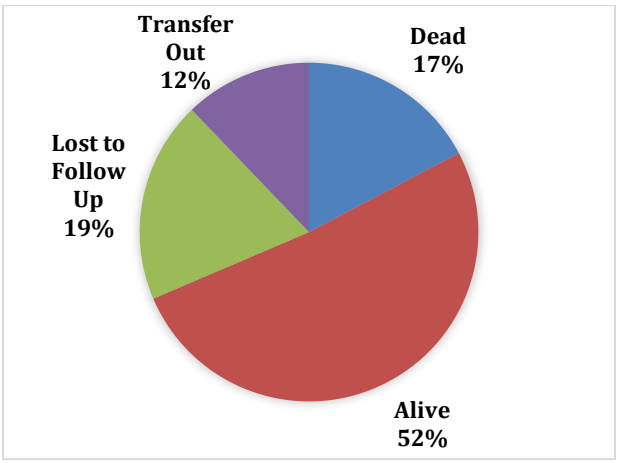


Figure 4: Follow up status among HIV infected patients on ART in Machakos County

A total of 2,766(72%) of the patients survived to the end of the study period. However, outcomes for 1,039 (19%) of the patients could not be accounted for due to lost to follow-up and 656 (12%) transferred out (figure 4).

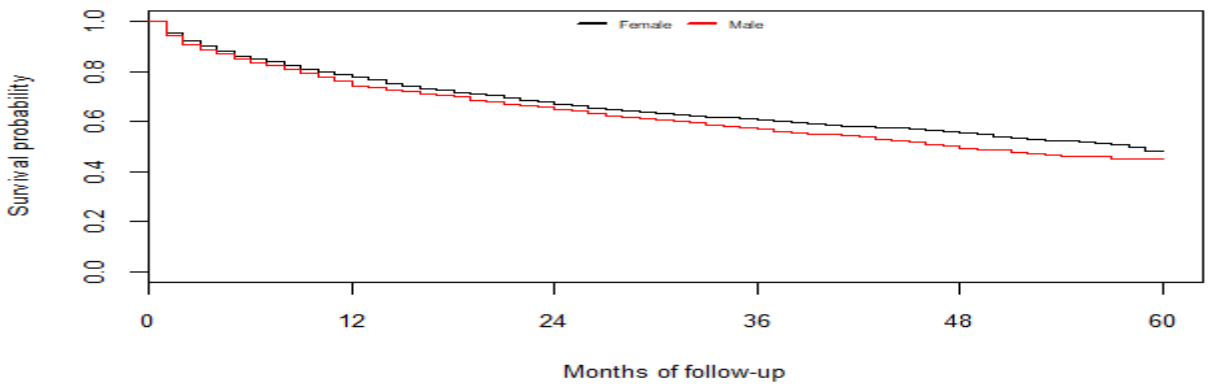


Figure 5: Plot of Kaplan-Meier survivor function by sex among HIV infected patients on ART in Machakos County

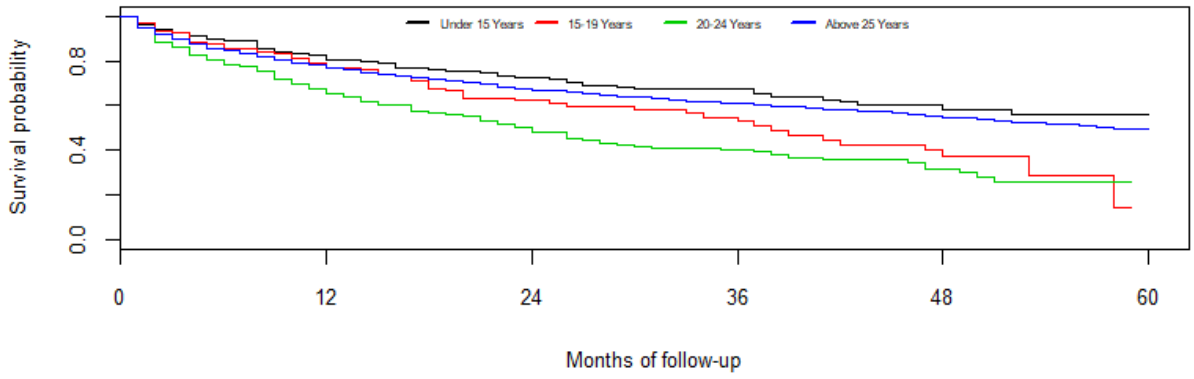


Figure 6: Plot of Kaplan-Meier survivor functions by age category among HIV infected patients on ART in Machakos County

Kaplan-Meier analysis of survival status showed that female sex exhibited better survival than the male i.e. estimated median survival time was 58months (95% CI: 51.790-59.546) vs. 48 months, (95%CI:44.398-53.479) respectively. Young adults aged between 20 to 24 years had lower survival time at 24 months (95%CI: 21.654-28.140) compared to other age groups while patients from Kangundo showed better survival at 57 months (95%CI: 52.490-55.404)compared to other sub-counties in Machakos.

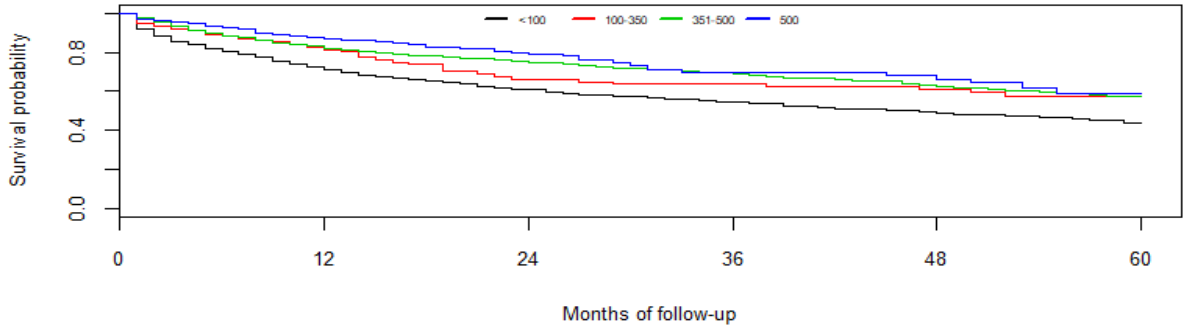


Figure 7: Plot of Kaplan-Meier survivor functions by CD4 Count among HIV infected patients on ART in Machakos County

Analysis of clinical characteristics revealed patients with BMI below 18.5 and CD4 counts below 100 cells/µl showed lower survival at 48 months (95%CI=41.483-58.822) and at 47 months (95%CI=40.108-54.034) respectively. Patients presenting with advanced disease (WHO stage 4) had lower the median survival time of 15 months (95%CI: 8.115-15.987) respectively compared to other WHO staging categories. Patients who were on a D4T based ART regimen displayed poorer prognosis compared to the other drug regimens (38months, 95%CI: 29.964-57.479).

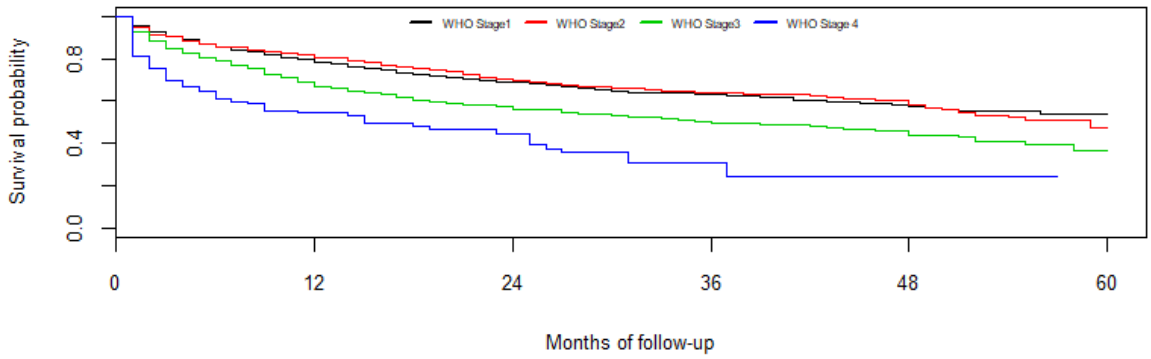


Figure 8: Plot of Kaplan-Meier survivor functions by WHO Staging among HIV infected patients on ART in Machakos County

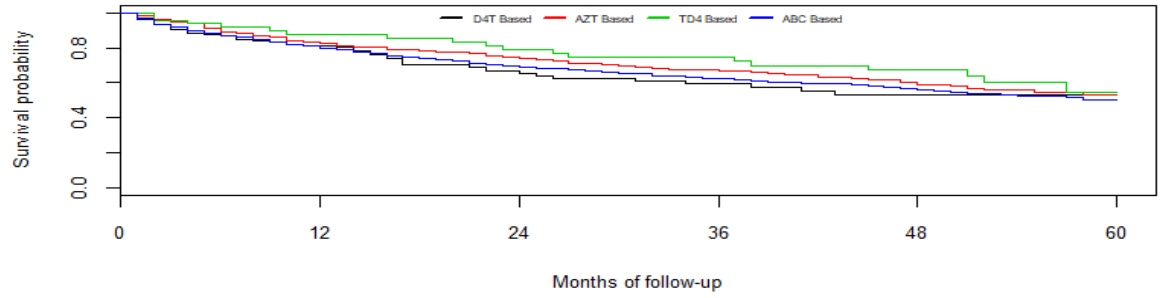


Figure 9: Plot of Kaplan-Meier survivor functions by baseline ART regimen among HIV infected patients on ART in Machakos County

Further still, patients who were suspected or confirmed to have TB infection were found to have lower survival time 35 months (95%CI: 43.169-56.001) and 31 months (95%CI:19.486-50.974) respectively compared to those with no signs of TB at 58 months (95%CI: 46.907-58.333). Patients with haemoglobin levels below 10 g/dl displayed lower survival times at 51 months (95%CI: 41.917-55.355) compared to those with HB levels above 10 g/dl at 56 months (95%CI: 46.847-57.331).

.

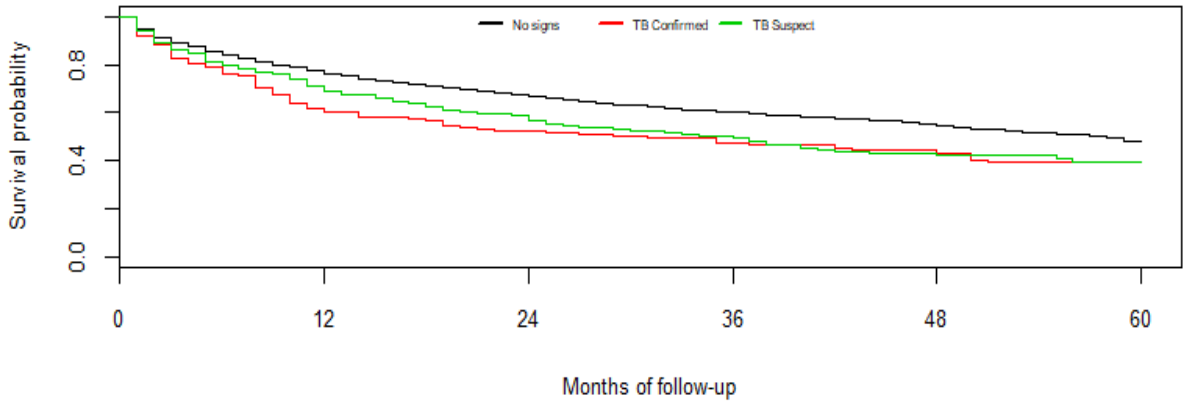


Figure 10:Plot of Kaplan-Meier survivor functions by TB Status among HIV infected patients on ART in Machakos County

Using log rank tests, we investigated whether the observed differences in data summaries among different factors were statistically significant with the null hypothesis being that there is no difference between survival curves. Results in Appendix A, Table 4 indicate that age at ART initiation, subcounty, baseline WHO stage, baseline CD4 count, TB infection in course of therapy and nutritional status manifest significant difference (at $p<.001$) in log rank test. While survival for sex, baseline haemoglobin and ART regimen showed no significant difference.

4.4 Results from Cox Proportional Hazards Model

Univariate analysis was carried out to identify variables that had a modest level of significance for inclusion in the multivariate analysis. This is to identify statistically significant factors that

influence the survival of HIV patients on antiretroviral therapy using the Cox proportional hazard model.

From the analysis, all variables except baseline haemoglobin and the category of regimen the patient was initiated were significant. In this case, significant predictors of mortality (at 5% significance level) were sex, sub county from which patients accessed services, age at which patients started therapy, baseline CD4 count ,patients nutritional status (measured through the variable BMI),WHO staging, and TB infection during treatment.(Appendix A, Table 5)

From the Cox regression model, patients from Machakos were more likely to die (hazard ratio ,HR 1.266;95% Cl 1.128-1.419) compared to patients from Kangundo while those from Mwala sub county displayed better prognosis (HR 0.832;95% Cl 0.715-0.968). Patients aged 20 to 24 years at initiation of ART were twice more likely to die compared to patients under 15 years (HR 2.013; 95% CI 1.551-2.364),while patients above 25 years displayed lower risk of dying compared to those under 15 years.(HR 0.896 ;95% CI 0.723-0.986). Patients with a baseline CD4 counts of less than 100 were twice at risk of death (HR 2.010; 95% CI 1.083–17.923) as opposed to patients initiating with CD4 cell counts greater than 500.Those with CD4 counts between 100-350 were also at high of risk of mortality (HR 1.619; 95% CI 1.547-1.700). Those patients with BMI less than 18.5 were at risk of death(HR 1.672; 95% CI 1.587-1.768) as compared to clients who had BMI >30 whereas those who initiate ART with BMI between 18.5-25.0 and 25.1-30.0 were at lower risk of death(HR 0.779; 95% CI 0.641-0.947 and HR 0.487; 0.395-0.599 respectively).Patients with advanced disease progression at initiation of therapy WHO stage 4 were most at risk of death HR 2.558; 95% CI 1.340 -2.816) compared to those in WHO stage 1. Patients with confirmed TB more at risk of death (HR 1.200; 95% CI 1.010-1.847) compared to those with no signs of TB

Table 6: Summary statistics from Cox proportional hazards of HIV patients on ART in Machakos County, Kenya

Characteristics	***Hazard Ratio***	***95% CI for the Hazard Ratio***		***P-Value***
		Lower	**Upper**	
Subcounty				
Kangundo(ref)	1			
Kathiani	0.869	0.713	1.058	0.1894
Machakos	1.266	1.128	1.419	**0.0442**
Athi River	1.092	0.782	1.524	0.0897
Mwala	0.832	0.715	0.968	**0**
Age Category at ART start				
Under 15(ref)	1			
15-19	1.401	1.039	1.89	**0.0377**
20-24	2.013	1.551	2.364	**0.0092**
>25	0.896	0.723	0.986	**0.013**
Baseline CD4 Count				
>500(ref)	1			
350-500	0.498	0.382	0.648	**0.025**
100-350	1.619	1.547	1.7	**0.007**
<100	2.01	1.083	17.923	**0**
BMI at ART initiation				
Bmi_>30.0(ref)	1			
Bmi 18.5-25.0	0.779	0.64	0.947	**0.0003**
Bmi 25.1-30.0	0.487	0.395	0.599	**0.0021**
Bmi <18.5	1.672	1.587	1.768	**0**
Baseline WHO stage				
WHO stage 1(ref)	1			
WHO stage 2	1.065	0.787	1.121	0.4725
WHO stage 3	1.52	1.332	1.733	**0.004**
WHO stage 4	2.558	1.34	2.816	**0**
TB Status				
No Signs(ref)	1			
TB Suspected	1.423	0.891	1.939	0.16
TB Confirmed	1.2	1.01	1.847	**0.0062**
Sex				
Female(ref)	1			
Male	1.3307	0.5402	1.453	0.922

4.5 Cox Proportional Hazards Model Diagnostics

We performed diagnostics for the model identified and whether it adequately described the data. Here we checked whether there was any violation of the assumptions of proportional hazards .To determine whether the fitted Cox regression model adequately describes the data, model diagnostics were conducted to assess for violation of the assumptions of proportional hazards. Test for violation of the assumption of proportional hazards was done. The proportional hazards assumption asserts the hazard ratios are constant overtime which is vital to the interpretation and use of a fitted proportional hazards model. This means, the risk of failure must be the constant no matter how long subjects have been followed. The test result showed the Wald chi-square value and corresponding p-values for each covariate. Since the p-value of the Wald test is greater than 0.05 for all covariates there no evidence of violation of the proportionality of hazard assumption.

Table 7: Cox Proportional Hazards model diagnostics for the Determinants of mortality and survival analysis among HIV infected patients in Machakos County, Kenya

Characteristics	**rho**	**Wald chisq**	**P-Value**
SubCounty			
Kangundo	0.0592	1.29E+00	0.0732
Kathiani	-0.0352	3.07E+00	0.272
Athi River	-0.10852	5.32E+00	0.7863
Machakos	0.0546	1.70E+00	0.1921
Mwala	0.0858	3.94E+00	0.0571
Age Category			
Under 15	-0.0552	1.74E+00	0.1869
15-19	-0.0241	3.28E-01	0.567
20-24	-0.0469	1.25E+00	0.2639
>25	-0.0334	6.13E-01	0.4337
Nutrition Assessment			
>30.0	0.0431	9.68E-01	0.3252
18.5-25.0	0.016	1.39E-01	0.7094
25.1-30.0	0.0742	2.82E+00	0.0932
<18.5	0.0429	1.05E+00	0.1869
Baseline WHO stage			
WHO stage 1	-0.014	1.12E+00	0.2878
WHO stage 2	0.119	3.90E-01	0.8223
WHO stage 3	-0.0496	1.21E+00	0.9741
WHO stage 4	0.0445	1.01E+00	0.3149

Baseline CD4 Count			
>500	-0.0065	2.40E-02	0.8768
100-350	0.0223	3.07E-01	0.5796
350-500	0.0185	1.81E-01	0.6706
<100	0.0141	1.09E-01	0.7407
TB Status			
TB Confirmed	0.1948	8.36E-07	0.9993
No Signs	-0.00657	5.07E-03	0.9432
TB Suspected	-0.20976	6.54E+00	0.8105
Sex			
Female	0.1265	2.35E+00	0.125
Male	-0.08003	7.73E-01	0.3794

4.6 Discussion

These findings were interpreted while making considerations on the nature of the study design. As a retrospective cohort study, data on exposure status at a specific earlier time-point (potential risk factors ART initiation) was required and also the identification of outcomes of interest during the study period.(i.e. Death or censored)

HIV Mortality Rate

Median survival following ART initiation in Machakos County was 55 months with mortality within the first year of follow up at 48% .This is consistent with findings from a retrospective cohort study done in Tanzania and Ethiopia where majority of patients died within the first year after ART initiation(Wada et al., 2014).In addition, (Bhatta et al., 2013) found similar results indicating higher risk of mortality in the first three months of antiretroviral treatment initiation among adults.

Factors Associated with mortality

The significance of gender of the patients in determining the survival time is variable in many studies. In this study, we did not find an association between gender and the survival time until HIV related death. This in contrast to a retrospective study done in India where there was a strong association between gender and survival probability (Bajpai et al.,2014). A similar study in

Tanzania also found that HIV infected males on ART were more likely to die in comparison to their female counterparts. Some of the possible reasons for the difference may include better health seeking behaviour among women hence tend to know about their HIV status at an earlier stage and begin antiretroviral therapy with better CD4 cell counts relative to males(Gunda et al., 2017).

This study found an association between TB infection and death hazard rate. Similarly, Kamenju and Aboud (2011) found that proportions of death were higher among HIV-infected TB patients (29.1% versus 15.2%) in comparison with the HIV uninfected TB patients ($p = 0.005$) while a retrospective cohort study done in the Armed Forces General Hospital in Ethiopia found that TB is a significant cause of mortality among HIV infected patients Hazard ratio(HR 1.734 95% CI: 1.039 - 2.893), Kebebew, (2012). This we may attribute to improved TB care and treatment in line with DOTs treatment guidelines. This study found that young adults aged between 20-24 years receiving ART are at highest risk of mortality. The Kenyan national HIV estimates according to the National AIDS Control Council of Kenya indicate a slightly different situation where adolescents and young adults aged between 15-24 years have been seen to have higher risk of mortality. A study conducted in Nigeria found that age was not significant predictor of mortality(Eguzo et al., 2014)

Patients living in Machakos sub-county had higher risk of HIV related mortality compared to the counter parts in the rather rural town settings. This is contrary to a study that found that patients living in or near urban centres had reduced risk of mortality (HR 0.82, 95% CI 0.7-0.999) Alvarez-Uria et al., (2013).

Findings in this study showed that individuals who began the ART treatment while malnourished (BMI below 18.5) exhibited a greater risk of death. These results are in tandem with findings from a study conducted by Mageda et al.,(2012), which found that despite the apparent benefits of HAART use on HIV-related survival, malnutrition is a significant independent predictor of mortality. Similarly, studies in Ethiopia and Tanzania found that presence of moderate to severe malnutrition at ART initiation was a significant independent predictor of death(Gunda et al., 2017; Kebebew, 2012)

Several studies revealed that advanced clinical stage at the initiation of ART is a significant predictor of mortality among HIV infected patients on ART(Betre & Ameni, 2016,Seyoum et al.,

2017) thus corroborating our findings that patients initiating on ART in WHO stages and 4 are at higher risk of mortality. In Tanzania, a retrospective cohort found that patients initiated on ART with advanced disease progression (WHO clinical stage 3 or 4) were four times more at risk of dying (HR,4.99 (95% CI :1.91–7.29) than patients in less chronic stages(WHO clinical stage 1 or 2)(Mageda et al., 2012).

CD4 counts at the beginning of ART was a significant predictor of mortality in this study. Patients starting ART with a baseline CD4 counts less than 100 compared those with CD4 counts of more than 500 copies were twice as at risk of dying .A similar finding from a study in India which found patients initiating on ART with CD4 counts less than 100 cells were thrice as likely to die (Alvarez-Uria et al., 2013).

This study did not find significant association between risk of mortality and haemoglobin level (a marker for anaemia).This is inconsistent with findings by Gebremedhin et al., (2013) who found that haemoglobin level less than 7 g/dl increased the risk of death four fold.

Strengths and Limitations

The main strength of this retrospective cohort study is the large sample size since all the eligible HIV-infected patients in the five subcounty hospitals were included. Follow up time of five years was also long enough to estimate survival and its determinants. Critical also is the fact that we used routine HIV treatment program data, which is cost effective and the findings give a vital insight for Machakos county to develop an effective and efficient HIV care and treatment program

The main limitation of this study, as with most implementing a retrospective study design ,was the exclusion of cases for variables that exhibited missing data may have biased our findings.Also,the contribution to improved prognosis or otherwise by cotrimoxazole prophylaxis was excluded from this study. This is due to the fact that the proportion of patients on ART not taking cotrimoxazole was too small to make meaningful comparison. Furthermore, information on two important potential risk factors was not available for analysis, these were any viral load test results and adherence .Viral load was largely unavailable since very few patients had received a viral load test and Kenya has only recently introduced it as the main diagnostic test for patient monitoring as well

as a measure for treatment failure. Another important limitation to this study was the high lost to follow up patient rates whose outcomes could not be accounted for and hence could bias the estimation of true mortality rates.

CHAPTER FIVE

SUMMARY CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents a summary of the main findings, conclusions drawn from the study and recommendations to help Machakos County streamline strategy in addressing factors associated with increased risk of mortality and improve survival time among HIV infected patients on antiretroviral therapy.

5.2 Summary of the findings

In the analysis of the factors associated with survival of HIV patients on antiretroviral treatment, analysis was done. Majority of the patients (71.8%) were above 25 years old at ART Start with 49.0% of the cohort being from Machakos sub county. Among patients who had CD4 count data available ,51.8% had CD4 counts less than 100 cells/µl at ART initiation with 24.1% of the patients starting treatment while in advanced disease stages (WHO stage 3 and 4). More than half of those with BMI results available (52.5%) were within the normal range of 18.5-25.0 with 9.3% having either active TB or were suspected to have TB. Majority of the patients (71,1%) had TDF based regimes prescribed at ART initiation.

Using Kaplan Meir methods, this study gave insights into the overall survival time as well as survival among group of covariates. The study found that after initiation of the treatment, HIV infected patients on antiretroviral therapy had an estimated 55 months' median survival time. Among patients who died while on antiretroviral therapy, majority of the deaths occurred within the first year of ART initiation (53. 5%).The study could however not account for the outcomes of 19.27% of the patients in the study due to lost to follow up.

Bivariate analysis (at 5% significance level) found that sex, sub county, age at ART start, baseline CD4 count ,patients nutritional status (measured through the variable BMI),WHO staging and TB infection during treatment and haemoglobin levels had modest levels of significance for inclusion in the multivariable analysis.

At multivariate analysis, a Cox proportional hazards model was fit. Significant effects in the survival of patients was noted with the following variables; Baseline BMI <18.5, CD4 cell count less than 100 cells at ART initiation, advanced disease stage 4 at ART initiation, TB infection and Machakos Sub-County. HIV patients on ART with either of these characteristics are at high risk of mortality. This study did not find an association between sex, haemoglobin levels and ART regimen and the survival time until HIV related death.

5.3 Conclusions

1. Mortality among HIV infected patients is highest in the first 12 months of ART initiation.
2. A high number of patient outcomes (19.27%) could not be ascertained since they were lost to follow up during the period of review (between 2011 and 2015) most 20% of patient.
3. Young adults aged 20-24 ,malnutrition (BMI<18.5)at ART initiation, CD4 cell count less than 100 cells at ART initiation, advanced disease stage 4 at ART initiation, TB infection and location of ART centre are the main contributors of mortality among HIV infected patients in Machakos county.
4. Patients residing in Machakos Sub-County, an urban centre within Machakos county had higher risk of HIV related mortality compared to the counter parts in the rather rural town settings.

5.4 Recommendations

1. The Machakos County Health management team should implement policy that places emphasis on intensified care and support of HIV infected patients in the first year of antiretroviral therapy.
2. This study recommends research to ascertain outcomes among lost to follow up patients. This is because the unaccounted for outcomes may to skew mortality estimates and associated factors among HIV infected ART taking patients.

REFERENCES

Alvarez-Uria, G., Naik, P. K., Pakam, R., & Midde, M. (2013). Factors associated with attrition, mortality, and loss to follow up after antiretroviral therapy initiation: Data from an HIV cohort study in India. *Global Health Action*, *6*(1), 1–8. https://doi.org/10.3402/gha.v6i0.21682

April, M. D., Wood, R., Berkowitz, B. K., Paltiel, A. D., Anglaret, X., Losina, E., … Walensky, R. P. (2014). The survival benefits of antiretroviral therapy in South Africa. *Journal of Infectious Diseases*, *209*(4), 491–499. https://doi.org/10.1093/infdis/jit584

Bajpai, R. C., Raj, P. Y., Jha, U. M., Chaturvedi, H. K., & Pandey, A. (2014). Demographic Correlates of Survival in Adult HIV Patients Registered at ART Centers in Andhra Pradesh , India : A Retrospective Cohort Study, *4*(1), 31–38. https://doi.org/10.5923/j.phr.20140401.06

Bajpai, R., Chaturvedi, H., Jayaseelan, L., Harvey, P., Seguy, N., Chavan, L., … Pandey, A. (2016). Effects of Antiretroviral Therapy on the Survival of Human Immunodeficiency Virus-positive Adult Patients in Andhra Pradesh, India: A Retrospective Cohort Study, 2007-2013. *Journal of Preventive Medicine & Public Health*, *394*, 49. https://doi.org/10.3961/jpmph.16.073

Betre, E. T., & Ameni, G. (2016). Survival and predictors of mortality among HIV patients on anti-retroviral treatment at Jinka hospital, South Omo, Ethiopia. *Epidemiology and Health*, *38*, e2016049. https://doi.org/10.4178/epih.e2016049

Bhatta, L., Klouman, E., Deuba, K., Shrestha, R., Karki, D. K., Ekstrom, A. M., & Ahmed, L. A. (2013). Survival on antiretroviral treatment among adult HIV-infected patients in Nepal: a retrospective cohort study in far-western Region, 2006–2011. *BMC Infectious Diseases*, *13*(1), 604. https://doi.org/10.1186/1471-2334-13-604

Breslow, N. (1979). Statistical methods for censored survival data. *Environmental Health Perspectives*, *32*, 181–192. https://doi.org/10.1289/ehp.7932181

CJ Hoffmann et al. (2011). Reducing mortality with co-trimoxazole preventive therapy at initiation of antiretroviral therapy in South Africa, *4*(164), 1709–1716. https://doi.org/10.1126/scisignal.2001449.Engineering

Cox, D. R. (1972). *Regression Models and Life-Tables*. *Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 34). Retrieved from http://www.biecek.pl/statystykaMedyczna/cox.pdf

Crellen, T., Ssonko, C., Piening, T., Simaleko, M. M., St. Calvaire, D. H., Gieger, K., & Siddiqui, M. R. (2018). What drives mortality among HIV patients in a conflict setting? A prospective cohort study in the Central African Republic. *BioRxiv*, 437103. https://doi.org/10.1101/437103

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, *72*(359), 557–565. https://doi.org/10.1080/01621459.1977.10480613

Eguzo, K. N., Lawal, A. K., Eseigbe, C. E., & Umezurike, C. C. (2014). Determinants of Mortality among Adult HIV-Infected Patients on Antiretroviral Therapy in a Rural Hospital in Southeastern Nigeria: A 5-Year Cohort Study. *AIDS Research and Treatment*, *2014*, 867827. https://doi.org/10.1155/2014/867827

Gebremedhin, A., Gebremariam, S., Haile, F., Weldearegawi, B., & Decotelli, C. (2013). Predictors of mortality among HIV infected children on anti-retroviral therapy in Mekelle Hospital, Northern Ethiopia: a retrospective cohort study. *BMC Public Health*, *13*, 1047. https://doi.org/10.1186/1471-2458-13-1047

Grambsch, P. M., & Therneau, T. M. (1994). *Proportional Hazards Tests and Diagnostics Based on Weighted Residuals* (Vol. 81). Retrieved from https://www.jstor.org/stable/2337123

Gunda, D. W., Nkandala, I., Kilonzo, S. B., Kilangi, B. B., & Mpondo, B. C. (2017). Prevalence and Risk Factors of Mortality among Adult HIV Patients Initiating ART in Rural Setting of HIV Care and Treatment Services in North Western Tanzania: A Retrospective Cohort Study. *Journal of Sexually Transmitted Diseases*, *2017*, 7075601. https://doi.org/10.1155/2017/7075601

Hosmer, D., Lemeshow, S., & May, S. (2008). *Applied survival analysis : regression modeling of time-to-event data*. Wiley-Interscience. Retrieved from https://www.wiley.com/en-ke/Applied+Survival+Analysis:+Regression+Modeling+of+Time+to+Event+Data,+2nd+Edition-p-9780471754992

K Kebebew, E. W. (2012). Survival analysis of HIV-infected patients under antiretroviral treatment at the Armed Forces General Teaching Hospital, Addis Ababa, Ethiopia | Kebebew | Ethiopian Journal of Health Development. *Ethiopian Journal of Health Development*, *26*. Retrieved from https://www.ajol.info/index.php/ejhd/article/view/115703

Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data*. J. Wiley. Retrieved from https://books.google.co.ke/books?hl=en&lr=&id=BR4Kq-a1MIMC&oi=fnd&pg=PR7&ots=xDom7ETO5-&sig=1Rk8lcB5a4HzmUxUxMP_FNsRlMw&redir_esc=y#v=onepage&q&f=false

Kamenju, P., & Aboud, S. (2011). Tuberculosis-HIV co-infection among patients admitted at Muhimbili National Hospital in Dar es salaam, Tanzania. *Tanzania Journal of Health Research*, *13*(1), 21–26. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24409643

Mageda, K., Leyna, G. H., & Mmbaga, E. J. (2012). High Initial HIV/AIDS-Related Mortality and -Its Predictors among Patients on Antiretroviral Therapy in the Kagera Region of Tanzania: A Five-Year Retrospective Cohort Study. *AIDS Research and Treatment*, *2012*, 1–7. https://doi.org/10.1155/2012/843598

McNairy, M. L., Jannat-Khah, D., Pape, J. W., Marcelin, A., Joseph, P., Mathon, J. E., … Evans, A. (2018). Predicting death and lost to follow-up among adults initiating antiretroviral therapy in resource-limited settings: Derivation and external validation of a risk score in Haiti. *PLOS ONE*, *13*(8), e0201945. https://doi.org/10.1371/journal.pone.0201945

Méda, Z. C., Sombié, I., Sanon, O. W. C., Maré, D., Morisky, D. E., & Chen, Y.-M. A. (2013).

Risk factors of tuberculosis infection among HIV/AIDS patients in Burkina Faso. *AIDS Research and Human Retroviruses*, *29*(7), 1045–1055. https://doi.org/10.1089/aid.2012.0239

NASCOP. (2014). Kenya HIV Prevention Revolution Road Map. *Kenya HIV Prevention Revolution Road Map*, 24.

National AIDS and STI Control Programme. (2015). *Kenya Hiv Estimates 2015*.

Ogoina, D., Obiako, R. O., Muktar, H. M., Adeiza, M., Babadoko, A., Hassan, A., … Tabi-Ajayi, E. (2012). Morbidity and Mortality Patterns of Hospitalised Adult HIV/AIDS Patients in the Era of Highly Active Antiretroviral Therapy: A 4-year Retrospective Review from Zaria, Northern Nigeria. *AIDS Research and Treatment*, *2012*, 1–10. https://doi.org/10.1155/2012/940580

Schomaker, M., Egger, M., Maskew, M., Garone, D., Prozesky, H., Hoffmann, C. J., … IeDEA Southern Africa. (2013). Immune recovery after starting ART in HIV-infected patients presenting and not presenting with tuberculosis in South Africa. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *63*(1), 142–145. https://doi.org/10.1097/QAI.0b013e318288b39d

Seyoum, D., Degryse, J.-M., Kifle, Y. G., Taye, A., Tadesse, M., Birlie, B., … Speybroeck, N. (2017). Risk Factors for Mortality among Adult HIV/AIDS Patients Following Antiretroviral Therapy in Southwestern Ethiopia: An Assessment through Survival Models. *International Journal of Environmental Research and Public Health*, *14*(3). https://doi.org/10.3390/ijerph14030296

Smith, T., & Ryan, M. (n.d.). *Survival Analysis Using Cox Proportional Hazards Modeling For Single And Multiple Event Time Data*. Retrieved from https://pdfs.semanticscholar.org/6f13/c884643fac0bb99ee984cab6ff27d50177d2.pdf

UNAIDS. (2010). *GLOBAL REPORT*.

Wada, N., Jacobson, L. P., Cohen, M., French, A., Phair, J., & Muñoz, A. (2014). Cause-specific mortality among HIV-infected individuals, by CD4(+) cell count at HAART initiation, compared with HIV-uninfected individuals. *AIDS (London, England)*, *28*(2), 257–265. https://doi.org/10.1097/QAD.0000000000000078

APPENDICES

Appendix A

Table 1: Summary of HIV related mortality by different baseline characteristics of HIV patients in Machakos County, Kenya.

Characteristics	***Frequency***	***%***
Age at ART start		
Under 15	360	6.7%
15-19	409	7.6%
20-24	753	14.0%
>25	3,871	71.8%
Sex		
Female	3,289	61.0 %
Male	2,104	39.0%
Subcounty		
Athi River	1,018	18.9%
Kangundo	384	7.1%
Kathiani	427	7.9%
Machakos	2,643	49.0%
Mwala	921	17.1%
BMI at ART Initiation		
<18.5	835	15.5%
>30.0	251	4.7%
18.5-25.0	1,831	34.0%
25.1-30.0	565	10.5%
Missing	1,911	35.4%
Baseline WHO stage		
1	1,996	37.0%
2	1,118	20.7%
3	681	12.6%
4	306	5.7%
Missing	1,292	24.0%
Baseline_CD4 Count		
<100	1,594	29.6%
100-350	1,058	19.6%
350-500	265	4.9%
>500	158	2.9%
Missing	2,318	43.0%
TB Status		
No Signs	4,351	80.7%

TB Suspected	197	3.7%
TB Confirmed	307	5.7%
Missing	538	10.0%
Regimen Category		
ABC_Based	197	3.7%
AZT_Based	976	18.1%
D4T_Based	227	4.2%
TDF_Based	3,834	71.1%
Missing	159	2.9%
Baseline HB Levels		
<10 g/dl	174	3.2%
≥10 g/dl	851	15.8%
Missing	4,368	81.0%

Table 2: Mortality of HIV infected patients over different follow up time intervals

Follow Up Time in Months	*Number of Deaths*	*%*
0-12	498	53.4%
0-24	216	23.2%
0-36	127	13.6%
0-48	53	5.7%
0-60	38	4.1%

Table 3: Kaplan-Meier Analyses of Survival Status for Patients on Antiretroviral Treatment in Machakos

	Survival time (in months)	***Confidence Interval(CI 95%)***
Overall	55	51.046-59.986
Age Category at ART start		
Under 15	55	48.897-57.052
15-19	38	30.948-53.090
20-24	24	21.654-28.140
>25	58	54.013-59.510
Sex		
Female	58	51.790-59.546
Male	48	44.398-53.479
Subcounty		
Kangundo	57	52.490-55.404
Kathiani	52	43.961-58.001
Machakos	49	46.847-56.581
Athi River	54	52.123-57.742
Mwala	56	43.382-59.854
BMI at ART initiation		
<18.5	48	41.483-58.822
18.5-25.0	56	45.210-58.650
25.1-30.0	53	42.736-59.642
>30.0	52	49.207-56.758
Baseline WHO Stage		
1	57	56.115-59.318
2	55	49.604-59.070
3	34	29.451-48.896
4	15	8.115-15.987
Baseline CD4		
<100	47	40.108-54.034
100-350	50	48.152-56.338
350-500	55	52.236-57.300
>500	58	49.177-59.444
Regimen Category		
ABC_Based	52	32.490-55.404
AZT_Based	56	46.847-59.331
D4T_Based	38	29.964-57.479
TDF_Based	56	36.847-58.471
TB Status		
No Signs	58	46.907-58.333
TB Suspected	35	43.169-56.001

TB Confirmed	31	19.486-50.974
Baseline Hb		
<10 g/dl	51	41.917-55.355

Table 4: Results of the log-rank tests -Determinants of mortality and survival analysis among HIV infected patients in Machakos County, Kenya.

Characteristics	***Chi-Square***	***P-value***
Sex	7.3	0.007
Age	99.7	0.0000
Subcounty	1.5	0.0000
Baseline WHO Stage	97.6	0.0000
Baseline CD4 count	78.4	0.0000
BMI at ART initiation	71.3	0.0000
Baseline ART regimen	6.4	0.0920
TB Status	31.9	0.0000
Baseline HB	1.1	0.8280

Table 5: Results from univariate Cox regression analysis - HIV patients on ART in Machakos County, Kenya

Characteristics	***Hazard Ratio***	***95% CI for the Hazard Ratio***		***P-Value***
		Lower	**Upper**	
Subcounty				
Kangundo(ref.)	1			
Kathiani	0.649	0.551	0.932	0.039
Machakos	1.561	1.067	3.669	0
Athi River	1.326	1.261	5.176	0.023
Mwala	0.465	0.25	0.854	0.004
Age Category at ART start				
Under 15(ref.)	1			
15-19	1.902	1.146	3.157	0
20-24	1.767	1.336	2.337	0.0013
>25	0.48	0.345	0.668	0.0034
Baseline CD4 Count				
>500(ref.)	1			
350-500	0.274	0.115	0.591	0.031
100-350	2.519	1.301	3.994	0.034
<100	4.37	1.194	13.52	0.012
BMI at ART initiation				
bmi >30.0(ref.)	1			
bmi 18.5-25.0	0.666	0.412	0.871	0.001
bmi 25.1-30.0	0.543	0.18	0.93	0
bmi <18.5	1.392	1.24	11.35	0.0381
Baseline WHO stage				
WHO stage 1(ref.)	1			
WHO stage 2	0.441	0.184	0.633	0.038
WHO stage 3	2.584	1.797	8.372	0.025
WHO stage 4	3.297	1.993	10.95	0
Sex				
Male (ref.)	1.6307	1.162	3.921	0.002
Female	1			
Regimen category				
TDF_Based(ref.)	1			
ABC_Based	1.18	0.76	1.82	0.564
AZT_Based	0.35	0.28	2.27	0.272
D4T_Based	1.13	0.84	1.54	0.313
TB Status				
No Signs(ref.)	1			

Tb suspected	1.204	1.149	1.364	0.016
TB confirmed	1.471	1.031	2.136	0.002
Baseline Hb				
≥10 g/dl(ref.)	1			
<10 g/dl	1.043	0.267	2.964	0.109

Appendix B

1. Letter of data collection approval Machakos County

REPUBLIC OF KENYA

GOVERNMENT OF MACHAKOS COUNTY
DEPARTMENT OF HEALTH & EMERGENCY SERVICES

Telephone: +254-44-20575
Fax: 254-44-20655
When replying please quote
DHES/GEN/CORR/VOL. I/99

Machakos Highway
P.O. Box 2574-90100
Machakos, Kenya.
16th March, 2017

Medical Superintendents:-
- Machakos Level 5
- Mwala Level 4
- Kangundo Level 4
- Matuu Level 4
- Kathiani Level 4

Sub County Medical Officers of Health:-
- Athi River
- Mwala

MACHAKOS COUNTY

RE: KING'UTU KEVIN APPROVAL LETTER TO CONDUCT A RESEARCH STUDY IN MACHAKOS COUNTY

It is my understanding that Kevin King'utu will be conducting a research study in select Machakos County on ***"Survival Analysis of HIV infected patients on Antiretroviral Therapy in Machakos County".***

Mr. King'utu is a student at Maseno University and has informed me of the study objectives, design of the study as well as the targeted population.

I support this effort and will provide any assistance necessary for the successful implementation of this study.

Sincerely,

Dr. Jacks Nthanga
Ag: Chief Officer
Department of Health and Emergency Services
MACHAKOS COUNTY

2. Letter of study apporval, Maseno University

MASENO UNIVERSITY
SCHOOL OF GRADUATE STUDIES

Office of the Dean

Our Ref: EL/SMM/00890/2015

Private Bag, MASENO, KENYA
Tel:(057)351 22/351008/351011
FAX: 254-057-351153/351221
Email: sgs@maseno.ac.ke

Date: 1st September, 2017

TO WHOM IT MAY CONCERN

RE: PROPOSAL APPROVAL FOR KING'UTU KEVIN NG'ANG'A— EL/SMM/00890/2015

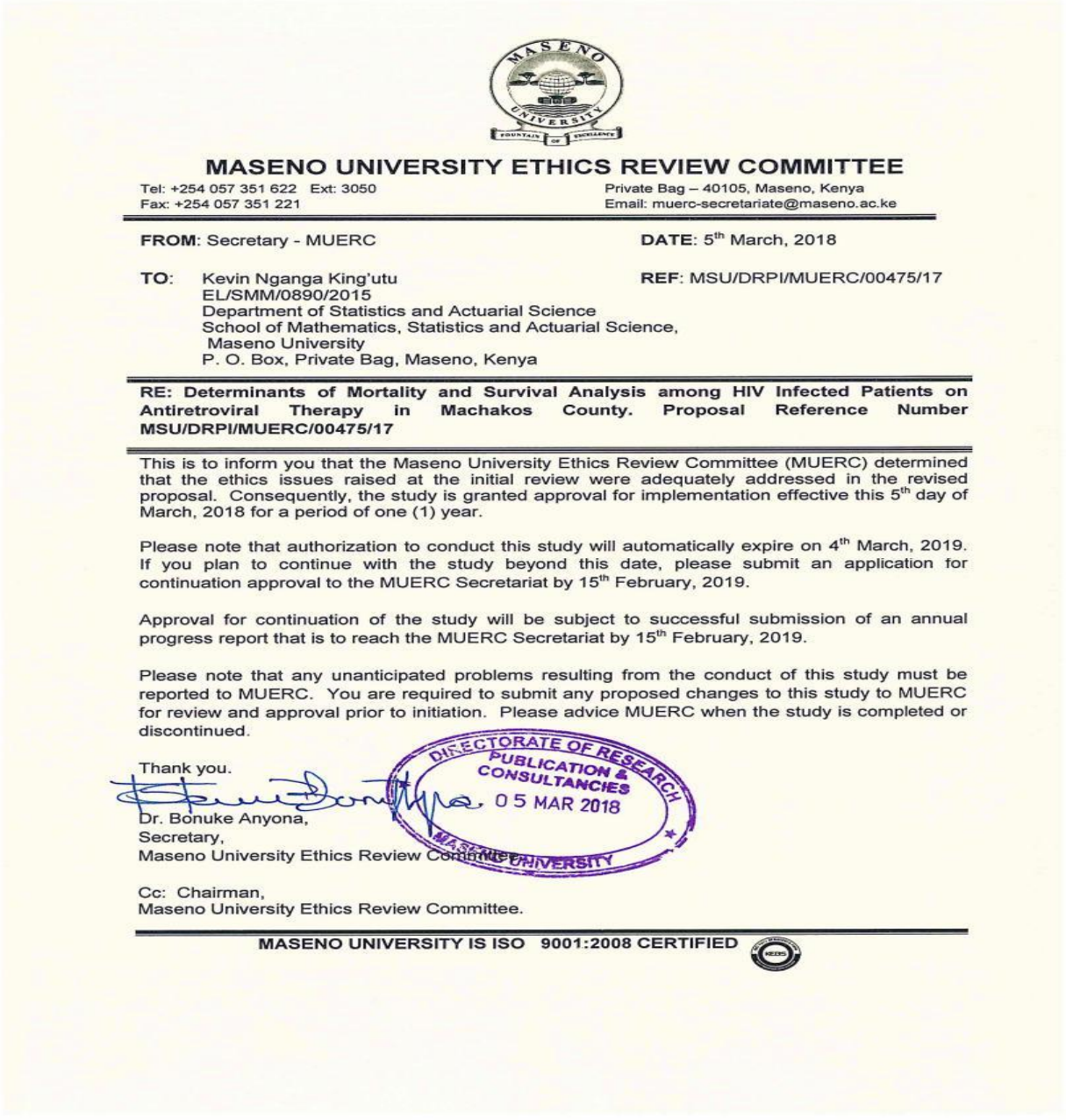
The above named is registered in the Master of Science Programme of the School of Mathematics, Statistics and Actuarial Science, Maseno University. This is to confirm that his research proposal titled "**Determinants of mortality and Survival Analysis among HIV Infected patients on Antiretroviral therapy in Machakos County, Kenya**" has been approved for conduct of research subject to obtaining all other permissions/clearances that may be required beforehand.

Dr. Patrick Onyango
ASSOCIATE DEAN, SCHOOL OF GRADUATE STUDIES

SCHOOL OF GRADUATE STUDIES
MASENO UNIVERSITY
11 SEP
P.O. BOX PRIVATE BAG, MASENO

Maseno University *ISO 9001:2008 Certified*

3. Letter of ethical approval, Maseno University Ethics review committee



MASENO UNIVERSITY ETHICS REVIEW COMMITTEE

Tel: +254 057 351 622 Ext: 3050
Fax: +254 057 351 221

Private Bag – 40105, Maseno, Kenya
Email: muerc-secretariate@maseno.ac.ke

FROM: Secretary - MUERC

DATE: 5th March, 2018

TO: Kevin Nganga King'utu
EL/SMM/0890/2015
Department of Statistics and Actuarial Science
School of Mathematics, Statistics and Actuarial Science,
Maseno University
P. O. Box, Private Bag, Maseno, Kenya

REF: MSU/DRPI/MUERC/00475/17

RE: Determinants of Mortality and Survival Analysis among HIV Infected Patients on Antiretroviral Therapy in Machakos County. Proposal Reference Number MSU/DRPI/MUERC/00475/17

This is to inform you that the Maseno University Ethics Review Committee (MUERC) determined that the ethics issues raised at the initial review were adequately addressed in the revised proposal. Consequently, the study is granted approval for implementation effective this 5th day of March, 2018 for a period of one (1) year.

Please note that authorization to conduct this study will automatically expire on 4th March, 2019. If you plan to continue with the study beyond this date, please submit an application for continuation approval to the MUERC Secretariat by 15th February, 2019.

Approval for continuation of the study will be subject to successful submission of an annual progress report that is to reach the MUERC Secretariat by 15th February, 2019.

Please note that any unanticipated problems resulting from the conduct of this study must be reported to MUERC. You are required to submit any proposed changes to this study to MUERC for review and approval prior to initiation. Please advice MUERC when the study is completed or discontinued.

Thank you.

DIRECTORATE OF RESEARCH PUBLICATION & CONSULTANCIES 05 MAR 2018 MASENO UNIVERSITY

Dr. Bonuke Anyona,
Secretary,
Maseno University Ethics Review Committee

Cc: Chairman,
Maseno University Ethics Review Committee.

MASENO UNIVERSITY IS ISO 9001:2008 CERTIFIED