Modelling the Survival Time of HIV Positive Pediatrics using Cox PH and Accelerated Failure Time Models

By:

Amsayaw Tefera

Email: amsetefera@gmail.com

A thesis submitted to the Department of Statistics, College of Natural Science, and Jimma University as partial fulfillment of the requirements for the degree of Master of Science (MSc) in Biostatistics

January 2014

Jimma, Ethiopia

Modelling the Survival Time of HIV Positive Pediatrics using Cox PH and Accelerated Failure Time Models

MSc Thesis

Amsayaw Tefera

Advisor: Yehenew Getachew (PhD)

Co-advisor: Dinberu Seyoum (MSc)

January 2014

Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis research advisors, we hereby certify that we have read and evaluated the thesis prepared by **Amsayaw Tefera** under our guidance, which is entitled "**Modelling the Survival Time of HIV Positive Pediatrics using Cox PH and Accelerated Failure Time Models"**. We recommend that the thesis be submitted as it fulfills the requirements for the degree of Master of Science (MSc) in Biostatistics.

Yehenew Getachew (PhD)	_____	_____
Advisor	Signature	Date
Dinberu Seyoum (MSc)	_____	_____
Co-advisor	Signature	Date

As the members of the board of examiners of MSc. thesis, open defense examination of **Amsayaw Tefera**, we certify that we have read and evaluated the thesis and examined the candidate. We recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____	_____	_____
Name of chairman	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of Co-advisor	Signature	Date
_____	_____	_____
Name of internal examiner	Signature	Date
_____	_____	_____
Name of external examiner	Signature	Date

STATEMENT OF THE AUTHOR

First, I declare that this thesis is a result of my genuine work and all sources of materials used for writing it have been duly acknowledged. I have submitted this thesis to Jimma University in partial fulfillment for the Degree of Master of Science. The thesis can be deposited in the library of the university to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Amsayaw Tefera

Date: _____

Signature: _____

Jimma University, Jimma

ACKNOWLEDGMENT

Many people have given me their generous and kind assistance during this last year. Without their support, this thesis could not have been realized. I would like to thank them here for the roles they played, and the impact they have made on this research, and on me.

Firstly, I wish to express my special thanks to my principal advisor and instructor, Dr. Yehenew Getachew, for his excellent scientific guidance and tireless efforts to make this work a reality. My gratitude also goes to my co-advisor, Mr. Dinberu Seyoum, for his unreserved guidance and technical input.

I would like to give special thanks to Mr. Legesse Negash who did a great job from starting title selection to until final proposal and Mr. Bogale Tessema for give me the direction and enforce me to do my thesis on pediatrics HIV/AIDS. Even if doing research on pediatrics HIV/AIDS so difficult but he providing me with many helpful suggestions to overcome obstacles related to clinical and medical problems.

I am very grateful to my family for being there to encourage and endless support to me during the period of study.

My sincere appreciation also goes to my friends and colleagues whoever explicitly or implicitly backed my work in whatever means**.**

Above all, I am very grateful to God, the Almighty for keeping me in good health to enable me carryout this research project, to Him be the Glory and Honor forever and ever.

ABSTRACT

Background: The introduction of antiretroviral therapy in 1996 improved the longevity and wellbeing of peoples living with HIV in the industrialized world including children. This survival benefit of antiretroviral therapy (ART) in reducing HIV related deaths has been well studied in the developed world. In resource-poor settings, where such treatment was started recently, there is inadequate information about impact of ART on the survival of patients especially in children. Hence, this study aimed to modelling the overall survival (OS) after HAART in HIV/AIDS pediatrics by Applying Accelerated Failure Time (AFT), Cox Proportional Hazard (PH) and Cox with time varying Coefficient models.

Methods: Institution based retrospective follow up study was carried out among HIV-positive pediatrics from September 2005 to September 2013 at JUSH. Out of a population of HIV positive pediatrics who were taking antiretroviral therapy in the hospital during that period, data on 218 patients are included in this study. The study subjects were pediatrics in the age less than 15 years. Non-parametric methods, Kaplan-Meier method and log-rank tests, were employed to compare the survival between the different categories of the explanatory variables. Semi-parametric methods, Cox PH models with time-dependent covariates and parametric methods, parametric PH model and AFT model was applied with the objective of identifying potential predictors of mortality.

Results: After initiation of the antiretroviral treatment, HIV positive pediatrics lived for an average of 40.14 months; the median survival time was found to be 38 months. The baseline functional status, CD4 counts, WHO clinical stage, TB/HIV co-infected and opportunistic infections significantly influence the survival of the patients.

Conclusion: Although the proportional hazard assumption not holds for Cox PH model, the weibull parametric model fitted the data well and can be taken as an alternative for Cox PH model.

LIST OF ACRONYMS

AFT	The Accelerated Failure Time
AIC	Akaike Information Criterion
AIDS	Acquired Immune Deficiency Syndrome
ARV / ART	Antiretroviral Therapy
BMI	Body Mass Index
CD4 %	Cluster of Differentiation 4 percentage
CPT	Cotrimoxazole Prophylaxis Therapy
FHAPCO	Federal HIV/AIDS Prevention and Control Office
HAART	Highly Active Antiretroviral Therapy
JUSH	Jimma University specialized Hospital
Ols	Opportunistic Infections
PH	Proportional Hazard
SD	Standard Deviation
TB	Tuberculosis
TLC	Total Lymphocyte Count
UNAIDS	Joint United Nations Programme on HIV/AIDS
W/H	Weight for Height
WHO	World Health Organization

LIST OF CONTENTS

Contents

LIST OF TABLES

LIST OF FIGURE

CHAPTER ONE

1. INTRODUCTION

1.1. Pediatric HIV/AIDS

A decade ago, living with HIV/AIDS was almost equivalent to a death sentence. Since 1996, with the introduction of combined antiretroviral treatment AIDS patients have opportunities of being cared for. The introduction of antiretroviral therapy (ART) has improved survival and quality of HIV-infected patients in the developed countries. This survival benefit of highly active antiretroviral therapy (HAART) in reducing HIV related deaths has been well studied in the developed world. In resource-poor settings however, where such treatment was started recently, knowledge on treatment results is limited. Moreover, mortality has been high in the first month of ART initiation. But factors contributing to high mortality are poorly investigated (Gebremedhin et al., 2013). About 1,400 children under the age of 15 are infected with HIV every day and approximately 90% of these infections occur in sub-Saharan Africa. Without appropriate care and treatment more than 50% of newly infected children die before celebrating their second birthday (Abrams et al., 2011).

The use of ART has dramatically reduced HIV associated mortality and morbidity among children in resource-rich settings. Pediatric HIV has been transformed from rapidly fatal infection of early childhood to a manageable chronic disease. For this reason, many HIV infected children are aging into adolescence and early adulthood (WHO, 2011). But according to American Academy of Pediatrics, there might be differences from country to country in many ways such as nutritional status, racial or ethnic and gender differences in total lymphocyte count and CD4+cell count (Abrams et al., 2011).

95% of HIV infections in children are acquired through mother-to-child transmission during pregnancy, around the time of delivery, or during breastfeeding. While there are effective ways to prevent mother-to-child transmissions, fewer than 10 percent of women are being reached by prevention efforts (Koye et al., 2012). HIV/ AIDS affect children in three ways. It incapacitates and kills their parents. It incapacitates and kills them directly. And, finally, by infecting or orphaning them, it exposes them to stigma, discrimination, abuse and exploitation.

Studying the history of disease progression due to HIV/AIDS and the treatments are useful for the purpose of developing treatment guidelines, modelling the epidemic and prioritizing and allocating resources. Health care planning depends on good knowledge of prevalence, which requires an accurate understanding of survival patterns. Monitoring the length of survival after diagnosis is, therefore, an important component of the surveillance of AIDS. It provides a basis for evaluating individual prognostic factors. In addition, differences in survival may also reflect differences in access to health care (e.g. Access to testing, counseling, preventive treatment). The survival of patients with AIDS may depend on a variety of factors including host factors, the patterns of diseases present, access to health care, diagnostic routines and therapeutic interventions (Assefa et al., 2012).

Knowledge of the survival times of patients with AIDS and variables that influence survival are important both for increasing understanding of the patho-physiology of the disease, clinical decision making and planning health service interventions (Jerene et al., 1995). The main challenge here is the sustainability of providing ART and other medications at district hospitals. About one million people were living with HIV in Ethiopia in the year (Assefa et al., 2009).

1.2. Statement of the Problems

Identifying predictors of patients' survival time after HAART is important not only because it enables the physicians to detect the factors whose changes affect patients' survival time, but also helps them to make the best decision about patients' treatment. Several factors are known to predict long-term survival of pediatrics HIV/AIDS patients, including, body weight, TB co-infected, Hemoglobin level, functional status, stage of the disease at the time of HAART, opportunistic infections, CD4 count, age and drug type. The role of some of these predictors of mortality among children on HAART such as CD4 count, age, drug type, opportunistic infections and functional status are controversial (Alemu, 2010; Atnafu, 2011 and Koye et al., 2012). For instance CD4 count was reported a significant factor for survival after HAART (Koye et al., 2012), whereas it was not a predictive factor in other studies (Alemu, 2010 and Atnafu, 2011). These differences may be due to a methodological issue.

Predictors of mortality among children after HAART are already identified by using non-parametric survival methods such as Kaplan-Meier and Cox Proportional Hazard (PH) in some studies (Alemu, 2010; Atnafu, H. (2011) and Koye et al., 2012); the latter is used when the effect of covariates on the hazard ratio is desired. Review of literature shows the extensive use of the Cox PH regression model for hazard rate of a given event (Atnafu, 2011 and Koye et al., 2012). However, the basic and the most important assumption underlying this model is proportionality of hazard rates, which may not be held in some situations. Where PH assumption is not met, it is improper to use standard Cox PH model as it may entail serious bias and loss of power when estimating or making inference about the effect of predictors of mortality (Moran, 2008 and Kleinbaum, 1998). A review of survival analysis in cancer journals reveals that only 5% of all studies using the Cox PH model considered the underline assumption (Orbe, 2002).

Recently, AFT models as parametric models have attracted considerable attention, because not only they do not need PH assumption but also thanks to availability of

standard methods such as Maximum Likelihood parameter estimation and testing can be done readily (Altman, 1995). AFT models have not been used very often and the few usage of this model is found in kidney transplant and time-to-menarche studies (Gluckberg, 1974 and Getie, 2011). It has not been used to recognize the predictors of mortality among children after HAART in the world.

A lot of effort has been made to improve the health quality of HIV-positive patients in Ethiopia and extend the time interval from HIV-infection/AIDS-diagnosis to death. For instance, activities like prevention of the disease through effective use of prophylaxis, intervention strategies and awareness raising such as mass media campaigns, peer education about HIV transmission, treatment of other sexually transmitted infections, safe blood transfusion provision, and prevention from mother to child treatment have been taken.

Despite the availability of a large body of research evidence that addresses issues about AIDS in Ethiopia, the level of understanding about predictor variables associated with mortality rate as a result of HIV infection is low.

Findings from studies in Africa and other low income countries also showed that there was good result of the ART programs i.e. the mortality of children on ART has decreased (Eley et al., 2004). But in sub-Saharan Africa including Ethiopia, there is substandard information about treatment results and the impact of ART on mortality of children. The independent predictors of survival among children living with HIV and on ART remain poorly characterized in Ethiopia. Therefore, the available evidence in Ethiopia regarding the determinants of survival among children less than 15 years old is inadequate. The limited work on the area and lack of appropriateness of the model applied for data have generated interest to determine predictors of mortality among children after HAART by fitting a statistical model that can explain the data in a more meaningful manner. This study was undertaken with objectives to estimate mortality and identify predictors that have statistically significant impact on the survival of HIV-positive pediatrics using

Cox-PH and AFT models. By doing so, we are trying to provide empirical evidence/answer, in the case of Jimma Zone, for the following research questions:

- Does Cox PH model explain the survival time of HIV positive pediatrics in meaningful way as compared to AFT models, mainly when censoring is more than 80%;
- What are the determinant factors to survival time of patients in the study area; and
- Which factor(s) accelerate or decelerate the survival time of HIV positive pediatrics under the era of HAART.

1.3. Objectives of the study

1.3.1. General objective

The general objective of this study is modelling the Survival Time of HIV Positive Pediatrics under the era of HAART using Cox PH and Accelerated Failure Time Models and examines determinant factors for survival time of HIV/AIDS pediatrics.

1.3.2 Specific objectives

In the light of the above major objective, the specific objectives are:

- to identify and come-up with a workable model for pediatric HIV/AIDS data set from two commonly used modeling approaches in survival analysis; namely, Cox-PH and AFT models;
- to identify predictors other than HAART that have significant impacts on the survival time of HIV positive pediatrics in JUSH; and
- to estimate the survival probability pattern of HIV/AIDS patients who are enrolling under ART in the study area.

1.4 Significance of the study

The results of this study will be very useful in the development of an effective HIV care and antiretroviral therapy patient monitoring system. Specifically, this study will be helpful in:

- estimating survival probability and its predictors among HIV positive children would provide an input for policy makers to revise the guidelines of pediatrics ART initiation; and
- this will also have implication for program planners and decision makers at various stages of the HIV/AIDS care and support program.

1.5 Limitation of the study

- The study is conducted based on secondary data which might have incomplete information.
- The study presumed that all deaths are caused by HIV/AIDS.
- The study is based on baseline values of the variables of interest.

CHAPTER TWO

2. LITERATURE REVIEW

The main focus of this literature review is to bring to light relevant information from previous studies about the determinants of AIDS mortality, progress of HIV/AIDS and various statistical models used for survival data.

2.1. Literature in relation to specific variables of the study (ART, Functional status, CD4, WHO stage, TB, Gender, Opportunistic infections and Socio-demographic Characteristics)

The primary goals to initiate antiretroviral therapy (ART) are mainly to suppress plasma HIV viral load, to reduce HIV-associated morbidity and prolong survival, to improve quality of life, to restore and preserve immunologic function, and to prevent HIV transmission (Office of AIDS Research Advisory Council ;2009). The investigation based on a prospective cohort of 1691 HIV seropositive women who enrolled between October 1994 and November 1995 in the United States, indicated that the use of antiretroviral therapies led to improved immunological function, suppressed HIV disease activity, and dramatic declines in morbidity and mortality.

While the asymptomatic phase in adults often continues for around 10 years, the progression of the disease in children is usually much faster and consequently they experience much higher morbidity and mortality rates than adults (DUNN et al, 2008). In parentally infected children without ART, three categories of progression can be differentiated: approximately 25-30% of children are so called rapid progresses who develop AIDS and die within the first year of life. others (50-60%), develop symptoms early in life, but survive with a slow downhill course of disease until the age of three to five years; a few children (5-25%), who are categorized as long-term

survivors, show symptoms of immune suppression only later in life and survive beyond the age of eight years (Tindyebwa, 2006).

All together around 80% of all children infected parentally with HIV die before the age of five years, unless treated with ARVs (Cook and Zumla 2003). A more recent study in rural Uganda shows that even 50% of HIV infected children die before reaching the age of 24 months without ART. Nevertheless, in spite of this generally fast progression of HIV in children and the obvious need for timely initiation of ART, access to the drugs is still insufficient for children and far behind when compared to services for adults (Prendergast et al., 2008). In 2008, the median age when children started ART was five and nine years, which is far too late, given the high mortality rate in the younger age groups (Unite for Children, 2008).

A further problem in providing services for HIV infected children is the lack of good and widely available pediatrics drug formulations as well as poor knowledge and policies (DEBaets, 2007), leading to concerns about possibly poor adherence resulting in the risk of the emergence of drug resistant virus strains . Consequently, children are usually treated with ARVs only in high level referral health facilities which are often difficult to access for the rural community because of long distances and insufficient transport facilities.

A retrospective cohort study was designed to assess clinical factors associated with growth in HIV infected children on ART in Uganda. Height and weight measurements were taken pre ART and post-ART initiation for at least 6 months from 749 children included in the study. Descriptive and logistic regression analyses were conducted to identify covariates associated with risk of either stunting or being underweight. Children in World Health Organization (WHO) clinical-stages II, III, and IV at baseline were 1.5 times more likely to become underweight from that of Clinical stage I, but Initiation of ART resulted in improvement in mean standardized weight-for-age. Weight-for-age Z score improved significantly after initiation of ART. This pediatric

population gained weight more rapidly than height after initiation of ART (Kabue et al., 2008).

In a study to evaluate changes and risk factors for death among HIV-infected children in pediatrics AIDS Clinical Trials Group 219/219c in the US, 3553 HIV-infected children were followed up for a median of 5.3 years. The study shows that increased risk of death was significantly associated with low CD4, pneumonia and AIDS-defining illness at entry. Whereas, decreased risks of mortality were identified for children who timely began highly active antiretroviral therapy (Brady et al., 2010).

The study conducted with a data from Boston hospital employed the Kaplan-Meier method to calculate median survival time and the Cox proportional hazards regression techniques to develop multivariable regression models. The result of the study indicated functional status and recent opportunistic diseases as the major predictors of survival time (Subbaraman et al., 2007).

A study on 272 HIV/AIDS patients on ART in Shashemene and Assela Hospitals employed Kaplan Meier method to construct survival curves and the Cox proportional hazards model to determine predictors of mortality (Alemu, 2010). The median survival time of the study was 104.4 weeks. The findings of the study showed WHO clinical stage IV, hemoglobin 510 g/dL, and cotrimoxazole prophylaxis therapy (CPT) initiation as the independent determinants of mortality. By the same token, (Jerene, 2006) based on 162 patients who were enrolled and treated between August 2003 and January 2005, ascertained that advanced disease stage (WHO clinical stage IV) and having total lymphocyte count (TLC) of up to 750/mcL were the major prognostic factors of mortality. The study also recommended identifying and treating patients early through improved counseling and testing strategies.

A study conducted in Sub-Saharan Africa based on data from 18 published cohort studies containing 39,536 HIV/AIDS patients had employed the Kaplan-Meir

method to assess the proportion of survival time and random-effects model to find hazard ratio of prognostic variables (Lutalo et al.,2006). Thus, a result of the study suggested advanced WHO clinical stage and lowCD4 cell count as indicator of high mortality. Similarly, a study in Malawi based on 1308 patients employed Kaplan-Meier method to assess the probability of survival and the Cox proportional hazards model to assess the potential predictors of death. The study found low body-mass index, WHO clinical stage IV, male gender, and baseline CD4 count lower than 50 cells/ml as independent determinants of death (Ferradini et al., 2006).

Another study from Cape Town on children known to be vertically infected with HIV has revealed that the risk of death was significantly associated with age less than 6 months and severity of disease at time of diagnosis. The median survival for all the children from time of diagnosis was 32 months. Infants diagnosed before 6 months of age had significantly shorter median survival (10 months) compared with 36 months for those diagnosed at 7-12 months of age. For the children over the age of 12 months the cumulative proportion surviving 48 months was 78%. Children with severe disease (category C) had a median survival of 21 months, significantly lower than that in category B (32 months). For the children in category A the cumulative proportion surviving at 48 months was 66% (Hussey et al., 19 98).

2.2. Cox PH versus AFT models

A review of literature on survival analysis used in different journals reveals that the Cox PH model is the most widely used way of analyzing survival data in clinical research. Researchers in medical sciences often tend to prefer semi-parametric instead of parametric models because of fewer assumptions. However, in recent time, AFT models as parametric models have attracted considerable attention, because not only they do not need PH assumption but also thanks to availability of standard statistical software ML, parameter estimation and testing can be done readily (Bradburn et al., 2003).

The main drawback of parametric models is the need to specify the distribution that most appropriately mirrors that of the actual survival times. This is an important requirement that needs to be verified and an appropriate distribution may be difficult to identify. When a suitable distribution can be found, the parametric model is more informative than the Cox model. It is straightforward to derive the hazard function and to obtain predicted survival times when using a parametric model, which is not the case in the Cox framework. Additionally, the appropriate use of these models offers the advantage of being slightly more efficient; they yield more precise estimates (i.e. smaller standard errors) and that in the parametric models we often use maximum likelihood procedures to estimate the unknown parameters in which this technique and its interpretation are familiar for researchers(Bradburn et al., 2003).

The parametric approach offers more in the way of predictions, and the AFT formulation allows the derivation of a time ratio, which is arguably more interpretable than a ratio of two hazards in Cox PH models. However, AFT models are relatively unfamiliar and seen rarely in medical research (Ersoy et al., 2003)

Performance comparison between Cox PH and parametric survival models have been made on other data sets other than time to death such as survival patients with different disease for example cancer. Accordingly, the analysis of Survival of Patients with Gastric Carcinoma of data from a historical cohort study of southern Iran with a diagnosis of stomach cancer has been made using Cox PH and parametric Lognormal, Exponential, Gompertz, Weibull, Log logistic and Gamma regression models in which all-parametric survival models were performed better than the Cox model (Pourhoseinghol et al., 2011). In this study, the proportional assumption is checked and found to be hold, but the model diagnostic for the parametric case has not been made yet. The comparison of parametric and semi-parametric models were made based on AIC. The study shows that, there may not be a single model that is substantially better than others in univariate analysis. The

data strongly supported the lognormal regression among parametric models and it can lead to more precise results as an alternative to Cox PH model.

Similarly, a study was conducted on prognostic factors of survival time after hematopoietic stem cell transplant in acute lymphoblastic leukemia on 206 patients that were enrolled after HSCH in Shariati Hospital between 1993 and 2007 so that the performance among AFT and Cox's models was assessed using explained variation and goodness of fit methods. Accordingly, predictive power of Weibull AFT models was superior to Cox PH model (Sayehmiri et al., 2008). Cox-Snell residual shows Weibull AFT fitted to data better than other distributions in multi variable analysis. In a similar fashion, a study was done on the survival of 1236 tuberculosis patients admitted in randomized controlled clinical trial in India. The result for this study showed that AFT model gave smaller deviance showing that AFT models seem to be more appropriate models than the Cox PH model (Ponnuraja et al., 2010).

Furthermore, AFT model was used to analyze data from 16 survivorship experiments in aging research experiments that evaluated the effects of one or more genetic manipulations on mouse lifespan. According to this study, AFT model deceleration factors also provided a more intuitive measure of treatment effect than the hazard ratio, and were robust to departures from modeling assumptions (Swindle, 2009).

In contrast, in analysis of survival in acute severe illness, AFT models identified the same predictors as the Cox model. But they did not demonstrate convincingly superior overall fit than the Cox PH model does (Ponnuraja et al., 2011) and the analysis of survivals of breast cancer relapse time with different treatments consistent results were obtained from accelerated failure time model and Cox proportional hazard model. But they Cox PH is also chosen over accelerated failure time model to calculate the appropriate survival curves of relapse time for patients in different treatment groups. i.e., with respect to predicting survival curve, Cox-PH model gives better fit than AFT models (Conge, 2010)

According to the studies above, the survival time of HIV infected patient after initiation of ART is a function of baseline variables like CD4 count, clinical stage of the disease, weight, age, drug types and so on. This study is aimed at modeling the Survival Time of HIV Positive Pediatrics under HAART using Cox PH and Accelerated Failure Time Models and examining determinant factors for survival time of HIV/AIDS pediatrics as a case study in Jimma University Specialized Hospital ART unit at Jimma.

CHAPTER THREE

3. METHODOLOGY

3.1 Data source and study site

Survival data for this study were obtained from Jimma University Specialized Hospital, which is a teaching referral hospital in Jimma. The ART unit in the hospital provides general voluntary counseling and testing services, follow up, pre-antiretroviral and ARV service for people living with HIV/AIDS. At the Hospital's ART clinic the data were recorded using the standardized data collection formats and registers prepared by the Ministry of Health. Health officers and nurses working in the clinic did data recording. The examining medical doctors also recorded follow up information about their patients. Data recording starts from the date patients started HIV regular care in the clinic till it was confirmed that patients have experienced one of the events - "death", "lost to follow-up", "dropped from the clinic", "stopped", and "transferred out to other health centers". Information on patients who had been transferred to the clinic were also recorded after reviewing their past history from the referring hospitals or heath centers.

Study Design: A historical institutional based retrospective cohort study was conducted in Jimma University Specialized Hospital to assess predictors of survival in pediatrics HIV/AIDS those who are receiving antiretroviral therapy.

3.2 Study Population

All 218 HIV + patients who are less than 15 years old and placed under HAART in between September 1, 2005 to September 1, 2013 in Jimma University Specialized Hospital were included in the study.

3.3 Variables in the study

The response variable: survival time of HIV patients in JUSH is the response (dependent) variable in this study. This dependent variable (in months) is measured as the length of time from ART start date until the date of death or censor.

Explanatory variables/factors: the predictor variables related to the social, demographic, medical and clinical background of the patients having these respective classifications. These covariates are described with their values or codes in Table 3.1.

3.4 Data collection

The data were collected by reviewing Pre-ART register, lab request, monthly cohort form, follow up form, ART intake form, patient's card, and death certificate complemented by registration by home visitors or calling by drug adherence supporters. The most recent laboratory results before starting ART are used as base line values. Five advanced ART nurses, who were involved for data collection and one ART physician have, supervised the data collectors. A one-day intensive training was given for supervisor and for five data collectors. The principal investigator of the study controls the overall activity. Data qualities were controlled by designing the proper data collection materials and pre-testing, through continuous supervision. All completed data collection forms were examined for completeness and consistency during data management, storage, cleaning and analysis. The data was entered and cleaned by trained data clerk and principal investigator respectively before analysis. To perform a task 5 days were taken for data collection and 5 days for entering, cleaning and coding the data.

Table 3.1: Explanatory variables used in the study

Variable	Description	Values/Codes
Sex	Gender	Male=1, Female=0
Age	Age	in year
weight	body weight	in kg
Family hist	Family history of HIV/AIDS	Yes=1, No=0
WHO Stage	WHO Clinical Stage	Clinical Stage- IV/III = 1 Clinical Stage- I/II = 0
Drug	Type of ART drug MmHg	D4t-3TC-NVP&D4t-3TC-EFV = 0 AZT-3TC-NVP&AZT-3TC-EFV = 1
Function	functional status	Working = 0 Ambulatory = 1 Bedridden = 2
Residence	place residence	Urban=0 Semi-urban=1 Rural =2
Ols	Opportunistic infections	Yes=1, No=0
CD4	CD4 count	$\geq$200 mm^3= 0 <200 mm^3 = 1
TB	TB screen	Yes=1, No=0
TBRx	TB preventive	Yes=1,No=0

3.5 Statistical methods of survival data analysis

Basic survival analysis: In follow-up studies the exact survival time is only known for those study participants or units who show the event of interest during the follow-up

period. For the others, what one can say is that they did not experience the event of interest during the follow-up period. These study participants or units are called censored observations. Individuals can be right censored, left censored or interval censored.

Censoring: subjects are right censored if it is known that the event of interest occurs some time after the recorded follow-up time whilst left censoring is when it is known that the event of interest happened sometime before the recorded follow up time. Interval censoring is when the exact time when the event occurred is not known precisely, but an interval bounding this time is known. If the interval is very short, it is common to ignore this form of censoring and pick one end point of the interval consistently. Interval-censored survival data frequently arise in clinical trials and follow-up studies such as AIDS and cancer studies.

The statistical method used in this thesis is known as survival data analysis, which involves modeling of data that have a principal end point, which is the time until an event occurs (time-to-event data). Survival analysis considers conditional information on the remaining time of a subject's survival given current survival time. Survival data were censored in the sense that they did not provide complete information since, for a variety of reasons, subjects of the study may not have experienced the event of interest. The existence of variables that change over time is also a distinguishing feature in survival analysis.

Basic Concepts in Survival Data Analysis: Survival data appear in various settings. In general, let T be the time until some event of interest occurs. This event can be death, the recurrence of a disease, the occurrence of being unemployed, or committing a second crime, and so forth. Clearly, T is a nonnegative random variable. Three well known functions characterizing the distribution of T are (1) the survival function, (2) the hazard rate (function) and (3) the probability density (or probability mass) functions.
The survival function, $S(t)$, is the probability of an individual surviving to time T , which is given by

$$S(t) = \Pr(T > t) = 1 - F(t),$$

Where $F(t) = \Pr(T \le t)$ is the c.d.f. of T and $f(t) = -dS(t)/dt$ is the density of T. The hazard rate is roughly the probability per time unit that an individual will fail in an interval given that the individual has survived to the beginning of the respective interval. It is denoted by $h(t)$ and defined as

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t / T \ge t)}{\Delta t} \\ &= f(t)/S(t) = -d \ln\left[S(t)\right]/dt \end{aligned}$$

The cumulative hazard function is defined as $H(t) = \int_0^t h(u)du = -\ln\left[S(t)\right]$, which implies that,

$$S(t) = \exp\left[-H(t)\right] = \exp\left[-\int_0^t h(u)du\right].$$

3.5.1 Non-parametric survival methods

Preliminary analysis of the data using non-parametric methods provides insight into the shape of the survival function for each group and get an idea of whether or not the groups are proportional, i.e., if the estimated survival functions for two groups are approximately parallel (do not cross).

The Kaplan-Meier Estimator is a non-parametric estimator of the survival function, which is not based on the actual observed event and censoring times, but rather on the order in which events occur. This principle of non-parametric estimation of the survival function is to assign probability to and only to event failure times. The log-rank test is utilized to test whether observed differences in survival experience between/among the groups are significant or not.

3.5.2 Semi-parametric survival models

The non-parametric method does not control for covariates and it requires categorical predictors. When we have several prognostic variables, we must use multivariate approaches. But we cannot use multiple linear regression or logistic regression because they cannot deal with censored observations. We need another method to model survival data with the presence of censoring. One very popular model in survival data analysis is the Cox proportional hazards model, which is proposed by (Cox, 1972).The Cox Proportional Hazards model is given by:

$$h(t,x) = h_0(t)\exp(\beta' x)$$

Where $h_0(t)$ is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero; $x = (x_1, x_2, ..., x_p)'$ is the values of the vector of explanatory variables for a particular individual, and $\beta' = (\beta_1, \beta_2, ..., \beta_p)$ is a vector of regression coefficients.

The corresponding survival functions are related as follows;

$$S(t,x) = S_0(t)^{\exp\left(\sum_{i=1}^{p}\beta_i x_i\right)}.$$

Where $S_0(t)$ the baseline survival functions. This model, also known as the Cox regression model, makes no assumptions about the form of $h_0(t)$ (non-parametric part of model) but assumes parametric form for the effect of the predictors on the hazard (parametric part of model). The model is therefore referred to as a semi-parametric model.

The beauty of the Cox approach is that this vagueness creates no problems for estimation. Even though the baseline hazard is not specified, we can still get a good estimate for

regression coefficients. The measure of effect is called hazard ratio. The hazard ratio of two individuals with different covariates x and x^* is

$$\hat{HR} = \frac{h_0(t)\exp\left(\hat{\beta}' x\right)}{h_0(t)\exp\left(\hat{\beta}' x\right)} = \exp\left(\sum \beta' \left(\hat{x} - x^*\right)\right)$$

This hazard ratio is time-independent, that is why this is called the proportional hazards model.

3.5.3 Methods of estimation

For this study, partial likelihood estimating were methods employed for Cox proportional hazards model. Fitting the Cox proportional hazards model, we wish to estimate $h_0(t)$ and β. One approach is to attempt to maximize the likelihood function for the observed data simultaneously with respect to $h_0(t)$ and β. It proposes a more popular approach, in which a partial likelihood function that does not depend on $h_0(t)$ is obtained for β. Partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameters ($h_0(t)$ in the Cox PH model). In this section, we will construct the partial likelihood function based on the proportional hazards model (Cox, 1975).

Let $t_1, t_2, ..., t_n$ be the observed survival time for n individuals. Let the ordered death time of r individuals be $t_{(1)} < t_{(2)} <, ..., < t_{(r)}$ and let $R(t_{(j)})$ be the risk set just before $t_{(j)}$ and r_j for its size. So that $R(t_{(j)})$ is the group of individuals who are alive and uncensored at a time just prior to $t_{(j)}$.The conditional probability that the i^{th} individual dies at $t_{(j)}$ given that one individual from the risk set on $R(t_{(j)})$ dies at $t_{(j)}$ is;

$$p\left(\text{individual i dies at } t_{(j)} \Big/ \text{one death from the risk set dies at } R\left(t(j)\right) at\ t_{(j)}\right)$$

$$= p\left(\frac{\text{individual i dies at } t_{(j)}}{\text{One death at } t_{(j)}}\right)$$

$$= p\left(\frac{\text{individual i dies at } t_{(j)}}{\sum_{k \in R(t(j))} p\left(\text{individual k dies at } t_{(j)}\right)}\right)$$

$$\approx \frac{p\left\{\text{individual i dies at } (t_{(j)}, t_{(j)} + \Delta t)\right\}/\Delta t}{\sum_{k \in R(t(j))} p\left(\text{individual k dies at } (t_{(j)}, t_{(j)} + \Delta t)\right)/\Delta t}$$

$$= \frac{\lim_{\Delta t \to 0} p\left\{\text{individual i dies at } (t_{(j)}, t_{(j)} + \Delta t)\right\}/\Delta t}{\lim_{\Delta t \to 0} \sum_{k \in R(t(j))} p\left(\text{individual k dies at } (t_{(j)}, t_{(j)} + \Delta t)\right)/\Delta t}$$

$$= \frac{h_i(t_{(j)})}{\sum_{k \in R(t_{(j)})} h_k(t_{(j)})}$$

$$= \frac{h_0(t)\exp(B' x_i(t_{(j)}))}{\sum_{k \in R(t_{(j)})} \exp\left(B' x_k(t_{(j)})\right)}$$

Then the partial likelihood function for the Cox PH model is given by

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp\left(\beta' x_i(t_{(j)})\right)}{\sum_{k \in R(t_{(j)})} \exp\left(\beta' x_k(t_{(j)})\right)}, \quad \text{.....(3.1)}$$

In which $x_i(t_{(j)})$ is the vector of covariate values for individual i who dies at $t_{(j)}$; the general method of partial likelihood was discussed by Cox (19 75).

Note that this likelihood function is only for the uncensored individuals. Let $t_1, t_2, ..., t_n$ be the observed survival time for n individuals and δ_i be the event indicator, which is zero if the i^{th} survival time is censored, and unity otherwise. The likelihood function in equation (3.1) can be expressed by

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{\exp\left(\beta' x_i(t_i)\right)}{\sum_{k \in R(t_i)} \exp\left(\beta' x_k(t_i)\right)} \right]^{\delta_i} \quad \text{.....(3.2)}$$

where $R(t_i)$ is the risk set at time t_i.

The presence of tied data points further complicates the derivation of the partial likelihood. Kalbfleisch and Prentice (1973) derive an exact partial likelihood for survival data with tied observations.

3.5.4 Proportional hazard assumption checking

The main assumption of the Cox proportional hazards model is proportional hazards. Proportional hazards means that the hazard function of one individual is proportional to the hazard function of the other individual, i.e., the hazard ratio is constant over time. There are several methods for verifying that a model satisfies the assumption of proportionality.

Graphical method: we can obtain Cox PH survival function by the relationship between hazard function and survival function

$$S(t, x) = s_0(t)^{\exp\left(\sum_{i=1}^{p} \beta_i x_i\right)}$$

Where $x=(x_1,x_2,...,x_p)'$ is the value of the vector of explanatory variables for a particular individual. When taking the logarithm twice, we can easily get

$$\log\left[-\log S(t,x_1)\right]-\log\left[-\log S(t,x_2)\right]=\sum_{i=1}^{p}\beta_i\left(x_{1i}-x_{2i}\right)$$

This does not depend on t. This relationship is very helpful to help us identify situations where we may have proportional hazards. By plotting estimated log (-log (survival)) versus survival time for two groups we would see parallel curves if the hazards are proportional. This method does not work well for continuous predictors or categorical predictors that have many levels because the graph becomes "cluttered". Furthermore, the curves are sparse when there are few time points and it may be difficult to tell how close to parallel is close enough.

However, looking at the K-M curves and log (-log (survival)) is not enough to ascertain of proportionality since they are univariate analysis and do not show whether hazards will still be proportional when a model includes many other predictors. But they support our argument for proportionality. We will show some other statistical methods for checking the proportionality.

Adding time-dependent covariates in the Cox model: we create time-dependent covariates by creating interactions of the predictors and a function of survival time and including them in the model. For example, if the predictor of interest is x_j , then we create a time-dependent covariate $x_j(t), x_j(t)=x_j\times g(t)$ where g (t) is a function of time, e.g., t, log t or Heaviside function of t. The model assessing PH assumption for x_j adjusted for other covariates is

$$h\left(t,x\left(t\right)\right)=h_0\left(t\right)\exp\left[\beta_1x_1+\beta_2x_2+...+\beta_jx_j+...+\beta_px_p+\delta\ \mathrm{x}_j\times g\left(t\right)\right],$$

Where $x(t) = (x_1, x_2, ..., x_p, x_j(t))'$ is the value of the vector of explanatory variables for a particular individual. The null hypothesis to check proportionality is that $\delta = 0$. The test statistic can be carried out using either a Wald test or a likelihood ratio test. In the Wald test, the test statistic is

$$W = \left(\hat{\delta} \Big/ \operatorname{se}\left(\hat{\delta}\right)^2 \right).$$

The likelihood ratio test calculates the likelihood under null hypothesis, L_0 and the likelihood under the alternative hypothesis L_a. The LR statistic is then $LR = -2\ln(L_0 / L_a) = -2(L_a - L_0)$ where L_0, L_a are log likelihood under two hypotheses respectively. Both statistics have a chi-square distribution with one degree of freedom under the null hypothesis. If the time-dependent covariate is significant i.e., the null hypothesis is rejected, and then the predictor is not proportional. In the same way, we will also assess the PH assumption for several predictors simultaneously.

Tests based on the Schoenfeld residuals: the other statistical test of the proportional hazards assumption is based on the Schoenfeld residual Schoenfeld (1982).The Schoenfeld residuals are defined for each subject who is observed to fail. If the PH assumption holds for a particular covariate then the Schoenfeld residual for that covariate will not be related to survival time. So this test is accomplished by finding the correlation between the Schoenfeld residuals for a particular covariate and the ranking of individual survival times. The null hypothesis is that the correlation between the Schoenfeld residuals and the ranked survival time is zero. Rejection of null hypothesis concludes that PH assumption is violated.

3.6 Cox proportional hazards model diagnostics

After a model has been fitted, the adequacy of the fitted model needs to be assessed. The model checking procedures below are based on residuals. In linear regression methods, residuals are defined as the difference between the observed and predicted values of the dependent variable. However, when censored observations are present and partial likelihood function is used in the Cox PH model, the usual concept of residual is not applicable. A number of residuals have been proposed for use in connection with the Cox PH model. For this study, three major residuals in the Cox model were used: the Cox-Snell residual, the deviance residual, and the Schoenfeld residual. Then we will talk about influence assessment.

3.6.1 Cox-Snell residuals and deviance residuals

The Cox-Snell residual is given by Cox and Snell (Klein, 1997). The Cox-Snell residual for the i^{th} with observed survival time t_i is defined as

$$\gamma c_i = \exp\left[\hat{\beta}' X_i\right] \quad \hat{H}_0(t_i) = \hat{H}_i(t_i) = -\log\left(\hat{S}_i(t_i)\right)$$

Where $\hat{H}_0(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i, which was derived by Kalbfleisch and Prentice (1973). This residual is motivated by the following result:

Let T has continuous survival distribution $S(t)$ with the cumulative hazard $H(t) = -\log(S(t))$. Thus, $S_T(t) = \exp(-H(t))$

Let $Y = H(t)$ be the transformation of T based on the cumulative function. Then the survival function for Y is

$$S_Y(y) = P(Y > y) = P(H(t) > y)$$
$$= P\left(T > H_i^{-1}(y)\right) = S_T\left(H_i^{-1}(y)\right)$$
$$= \exp\left(-H_T\left(H_T^{-1}(y)\right)\right) = \exp(-y)$$

Thus, regardless of the distribution of T, the new variable $y = H(t)$ has an exponential distribution with unit mean. If the model is be well fitted, the value $\hat{S}_i(t_i)$ would have similar properties to those $S_i(t_i)$. So $\gamma c_i = -\log\left(\hat{S}_i(t_i)\right)$ will have a unit exponential distribution with $f_r(\gamma) = \exp(-\gamma)$.

Let $S_R(\gamma)$ denote the survival function of Cox-Snell residual γc_i. Then

$$S_R(\gamma) = \int_{\gamma}^{\infty} f_R(x)dx = \int_{\gamma}^{\infty} \exp(-x)dx = \exp(-\gamma).$$

and

$$H_R(\gamma) = -\log(S_R(\gamma)) = -\log(\exp(-\gamma)) = \gamma$$

Therefore, we use plot of $H(\gamma c_i)$ versus γc_i to check the fit of the model. This gives a straight line with unit slope and zero intercept if the fitted model is correct. Note the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be negative.

The deviance residual (Subbaraman, 2007) is defined by

$$r_{Di} = sign(r_{m_i})\left[-2\left\{r_{m_i} + \delta_i \log\left(\delta_i - r_{m_i}\right)\right\}\right]^{1/2}$$

Where the function sign (.) is the sign function, which takes the value, 1 if γ_{m_i} is positive and -1 if γ_{m_i} negative; $\gamma_{m_i} = \delta_i - \gamma_{c_i}$ is the martingale residuals for the i^{th} individual and $\delta_i = 1$ for uncensored observation $\delta_i = 0$ for censored observation.

The martingale residuals take values between negative infinity and unity. They have a skewed distribution with mean zero. The deviance residuals are a normalized transform of the martingale residuals (Sayehmiri, 2008).They also have a mean of zero but are approximately symmetrically distributed about zero when the fitted model is appropriate. Deviance residual can also be used like residuals from linear regression. The plot of the deviance residuals against the covariates can be obtained. Any unusual patterns may suggest features of the data that have not been adequately fitted for the model. Very large or very small values suggest that the observation may be an outlier in need of special attention. In a fitted Cox-PH model, the hazard of death for the i^{th} individual at any time depends on the value of $\exp\left(\beta' x\right)$ which is called the risk score. A plot of the deviance residuals versus the risks core is a helpful diagnostic to assess a given individual on the model. Potential outliers will have deviance residuals whose absolute values are very large. This plot will give the information about the characteristic of observations that are not well fitted by the model.

3.6.2 Schoenfeld residuals

All the above three residuals are residuals for each individual. We will describe covariate wise residuals by Schoenfeld (1982). The Schoenfeld residuals were originally called partial residuals because the Schoenfeld residuals for i^{th} individual on the j^{th} explanatory variable x_{ij} is an estimate of the i^{th} component of the first derivative of the logarithm of the partial likelihood function with respect to β_j: From equation (3.2), this logarithm of the partial likelihood function is given by

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum_{i=1}^{p} \delta_i \{x_{ij} - a_{ij}\},$$

Where x_{ij} is the value of the j^{th} explanatory variable $j = 1,2,..,p$ for the i^{th} individual and

$$a_{ij} = \frac{\sum_{l \in R(t_i)} x_{jl} \exp(\beta' x_i)}{\sum_{l \in R(t_i)} \exp(\beta' x_i)}$$

The schoenfeld residual for i^{th} individual on x_j is given $\gamma_{p_{ij}} = \delta_i \{x_{ij} - a_{ij}\}$. The schoenfeld residuals sum to zero.

3.6.3 Diagnostics for influential observations

Observations that have an undue effect on model-based inference are said to be influential. In the assessment of model adequacy, it is important to determine whether there are any influential observations. The most direct measure of influence is $\hat{\beta}_j - \hat{\beta}_{j(i)}$, where $\hat{\beta}_j$ is the j^{th} parameter is, $j = 1,2,...,p$ in a fitted Cox PH model and $\hat{\beta}_{j(i)}$ is obtained by fitting the model after omitting observation i. In this way, we have to fit the n + 1 Cox models, one with the complete data and n with each observation eliminated. This procedure involves significant amount of computation if the sample size is large. We would like to use an alternative approximate value that does not involve an iterative refitting of the model. To check the influence of observations on a parameter estimate, Cain and Lange (1984) showed that an approximation to $\hat{\beta}_j - \hat{\beta}_{j(i)}$ is the j^{th} component of the vector

$$\gamma_{S_i}' V\left(\hat{\beta}\right)$$

Where γ_{S_i} is the $p \times 1$ vector of score residuals for the i^{th} observation (Collette, 2003). Which are modifications of Schoenfeld residuals and are defined for all the observations, and $V\left(\hat{\beta}\right)$ is the variance-covariance matrix of the vector of parameter estimates in the fitted Cox PH model. The j^{th} element of this vector is called delta-beta statistic for the j^{th} explanatory variable, i.e., $\Delta i\ \hat{\beta}_j \approx \hat{\beta}_j - \hat{\beta}_{j(i)}$, which tells us how much each coefficient will change by removal of a single observation. Therefore, we can check whether there are influential observations for any particular explanatory variable. On the other hand, the statistic, $LD_i = 2l_p(\beta) - 2l_p(\beta_{(i)})$, which is called the likelihood displacement statistic, can be used as a measure of how the maximized partial log Likelihood changes if the i^{th} observation was deleted from the data set. Observations that influence a particular parameter estimate have a large absolute value of DFBETA than other observations in the data set. Observations that do influence the overall fit of the model are those which have large values of likelihood displacement statistics than the other observations in the data set (Collett, 2003).

3.7 Strategies for analysis of non-proportional data

Suppose that statistic tests or other diagnostic techniques give strong evidence of non-proportionality for one or more covariates. To deal with this we will describe two popular methods: stratified Cox model and Cox regression model with time-dependent variables which are particularly simple and can be done using available software. Another way to consider is to use a different model. A parametric model such as an AFT model might be more appropriate for the data.

3.7.1 Stratified Cox Model

One method that we can use is the stratified Cox model, which stratifies on the predictors not satisfying the PH assumption. The data are stratified into subgroups and the model is applied for each stratum. The model is given by

$$h_g(t) = h_{0g}(t)\exp\left(\beta' z_{ig}\right) \quad \text{.....(3.3)}$$

Where g represents the stratum. Note that the hazards are non-proportional because the baseline hazards may be different between strata. The coefficients β are assumed to be the same for each stratum g. The partial likelihood function is simply the product of the partial likelihoods in each stratum. A drawback of this approach is that we cannot identify the effect of this stratified predictor. This technique is most useful when the covariate with non-proportionality is categorical and not of direct interest.

3.7.2 Cox Regression Model with time-dependent variables

Until now, we have assumed that the values of all covariates did not change over the period of observation. However, the values of covariates may change over time t. Such a covariate is called a time-dependent covariate. The second method to consider is to model non-proportionality by time-dependent covariates. The violation of PH assumptions is equivalent to interactions between covariates and time. That is, the PH model assumes that the effect of each covariate is the same at all points in time. If the effect of a variable varies with time, the PH assumption is violated for that variable. To model a time-dependent effect, one can create time dependent covariates $z(t), \beta z(t) = \beta z \times g(t)$; where g (t) is a function of t such as t; logt or Heaviside functions, etc. The choice of time-dependent covariates may be based on theoretical considerations and strong clinical evidence.

The Cox regression with both time independent predictors Z_i and time-dependent covariates $Z_j(t)$ can be written

$$h\left(t / z(t)\right) = h_0(t)\exp\left[\sum_{i=1}^{p_1}\beta_i z_i + \sum_{j=1}^{p_2}\alpha_j z_j(t)\right] \quad \text{.....(3.4)}$$

The hazard ratio at time t for the two individuals with different covariates z and z* is given by

$$\hat{HR}(t) = \exp\left[\sum_{i=1}^{p_1}\beta_i\left(\hat{z_i^*} - z_i\right) + \sum_{j=1}^{p_2}\alpha_j\left(\hat{z_j^*}(t) - z_j(t)\right)\right] \quad \text{.....(3.5)}$$

Note that, in this hazard ratio formula, the coefficient $\hat{\alpha}_j$ is not time-dependent. $\hat{\alpha}_j$, represents overall effect of $Z_j(t)$ considering all times at which this variable has been measured in this study. But the hazard ratio depends on time t. This means that the hazards of event at time t is no longer proportional, and the model is no longer a PH model.

In addition to considering time-dependent variable for analyzing a time-independent variable not satisfying the PH assumption, there are variables that are inherently defined as time-dependent variables. One of the earliest applications of the use of time-dependent covariates is in the report by Crowder (1977) on the Stanford Heart Transplant study.

3.8 Parametric proportional hazards model

The parametric proportional hazards model is the parametric versions of the Cox proportional hazards model. It is given with the similar form to the Cox PH models. The hazard function at time t for the particular patient with a set of p covariates $(x_1, x_2, ..., x_p)$ is given as follows:

$$h(t/x) = h_0(t)\exp\left(\beta_1 x_1 + \beta_2 x_2 +,...,+\beta_p x_p\right) = h_0(t)\exp(\beta' x)$$

The key difference between the two kinds of models is that the baseline hazard function is assumed to follow a specific distribution when a fully parametric PH model is fitted to the data, whereas the Cox model has no such constraint. The coefficients are estimated by partial likelihood in Cox model but maximum likelihood in parametric PH model. Other than this, the two types of models are equivalent. Hazard ratios have the same interpretation and proportionality of hazards is still assumed. A number of different parametric PH models may be derived by choosing different hazard functions. The commonly applied models are exponential, Weibull, or Gompertz models.

3.8.1 Weibull PH Model

The exponential model is nice enough, but the restriction that the hazard be constant over time is often questioned in practice. In fact, we can imagine a number of processes where we might expect hazard rates to be changing over time. If the (conditional) hazard is increasing or decreasing steadily over time, the exponential model will miss this fact.

The Weibull can be thought of as a hazard rate model in which the hazard is:

$$h(t) = \lambda\gamma\left(t\right)^{\gamma-1}, \text{ With } \lambda,\ \gamma > 0.$$

Here, the parameter γ is sometimes called a "shape parameter," because it defines the shape of the Weibull distribution.

- γ = 1 corresponds to an exponential model (thus the Weibull "nests" the exponential model),
- γ > 1 means that the hazards are rising monotonically over time, and
- 0< γ < 1 means hazards are decreasing monotonically over time.

Recalling that the survival function can be expressed as the exponent of the negative integrated hazard, we can see that:

$$S(t) = \exp\left[-\int_0^t \lambda\gamma(t)^{\gamma-1} dt \right] = \exp\left(-\lambda t^{\gamma}\right)$$

Under the Weibull PH model, the hazard function of a particular patient with covariates $(x_1, x_2, ..., x_p)$ is given by

$$h(t/x) = \lambda\gamma(t)^{\gamma-1} \exp\left(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p\right) = \lambda\gamma(t)^{\gamma-1} \exp\left(\beta' x\right).$$

We can see that the survival time of this patient has the Weibull distribution with scale parameter $\lambda \exp\left(\beta' x\right)$ and shape parameter γ. Therefore the Weibull family with fixed γ possesses PH property. This shows that the effects of the explanatory variables in the model alter the scale parameter of the distribution, while the shape parameter remains constant. Then the corresponding survival function is given by

$$S(t/x) = \exp\left\{-\exp\left(\beta' x\right)\lambda t^{\gamma}\right\} \text{.....(3.6)}$$

After a transformation of the survival function for a Weibull distribution, we can obtain

$$\log\left\{-\log\left(S(t)\right)\right\} = \log\lambda + \gamma\log(t)$$

The $\log\left\{-\log S(t)\right\}$ versus $\log(t)$ should give approximately a straight line if the Weibull distribution assumption is reasonable. The intercept and slope of the line will be rough estimate of $\log\lambda$ and γ respectively. If the two lines for two groups in this plot are essentially parallel, this means that the proportional hazards model is valid. Furthermore, if the straight line has a slope nearly one, the simpler exponential distribution is reasonable.

In the other way, for an exponential distribution, there is $\log S(t) = -\lambda t$ .Thus we can consider the graph of $\log S(t)$ versus t. This should be a line that goes through the origin if exponential distribution is appropriate.

Another approach to assess the suitability of a parametric model is to estimate the hazard function using the non-parametric method. If the hazard function were reasonably constant over time, this would indicate that the exponential distribution might be appropriate. If the hazard function increased or decreased monotonically with increasing survival time, a Weibull distribution or Gompertz distribution might be considered.

3.8.2 Exponential PH Model

The exponential PH model is a special case of the Weibull model when $\gamma = 1$. The hazard function under this model is to assume that it is constant over time. The survival and hazard function are written as

$$S(t) = \exp(-\lambda t),\ h(t) = \lambda .$$

Under the exponential PH model, the hazard function of a particular patient is given by

$$h(t/x) = h_0(t)\exp\left(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p\right) = \lambda \exp\left(\beta' x\right)$$

3.8.3. Gompertz PH model

The survival and hazard function of the Gompertz distribution are given by

$$S(t) = \exp\left(\frac{\lambda}{\theta}(1 - e^{\theta t})\right),\ h(t) = \lambda \exp(\theta t)$$

Here $0 \le t < \infty$ and $\lambda > 0$. The parameter θ determines the shape of the hazard function. When $\theta = 0$, the survival time then have exponential distributions, i.e., the exponential

distribution is also a special case of the Gompertz distribution. Like the Weibull, hazard function, the Gompertz hazard increases or decreases monotonically. For the Gompertz distribution $\log(h(t))$ is linear with t.

Under the Gompertz PH model, the hazard function of a particular patient is given by

$$h(t/x) = \lambda \exp(\theta t) \exp(\beta_1 x_1 + \beta_2 x_2 +, ..., \beta_p x_p) = \lambda \exp(\beta' x) \exp(\theta t)$$

It is straightforward to see that the Gompertz distribution has the PH property. But the Gompertz PH model is rarely used in practice.

3.9 Accelerated failure time (AFT) models

Although parametric PH models are very applicable to analyze survival data, there are relatively few probability distributions for the survival time that can be used with these models. In these situations, the AFT is an alternative to the PH model for the analysis of survival time data. Under AFT models, we measure the direct effect of the explanatory variables on the survival time instead of hazard, as we do in the PH model. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean survival time. Currently, the AFT model is not commonly used for the analysis of clinical trial data, although it is common in the field of manufacturing. Similar to the PH model, the AFT model describes the relationship between survival probabilities and a set of covariates.

Definition: For a group of patients with covariate $(x_{1,}, x_2, ..., x_p)$ the model is written mathematically as $S(t/x) = S_0(t/\eta(x))$,where $s_0(t)$ is the baseline survival function and η is an 'acceleration factor' that is a ratio of survival times corresponding to any fixed value of $S(t)$.The acceleration factor is given according to the formula $\eta(x) = \exp(\alpha_1 x_1 + \alpha_2 x_2 +, ..., +\alpha_p x_p)$.

Under an accelerated failure time model, the covariate effects are assumed to be constant and multiplicative on the time scale, that is, the covariate impacts on survival by a constant factor (acceleration factor).

According to the relationship of survival function and hazard function, the hazard function for an individual with covariate $(x_{1,}, x_2, ..., x_p)$ is given by

$$h(t/x) = [1/\eta(x)]h_0[t/\eta(x)] \quad \text{.....(3.7)}$$

The corresponding log-linear form of the AFT model with respect to time is given by

$$\log T_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + ... + \alpha_p x_{pi} + \sigma\varepsilon_i$$

Where μ the intercept, σ is scale parameter and ε_i is a random variable, assumed to have a particular distribution. This form of the model is adopted by most software package for the AFT model.

For each distribution of ε_i, there is a corresponding distribution for T. The members of the AFT model class include the exponential AFT model, weibull AFT model, log-logistic AFT model, log-normal AFT model, and gamma AFT model. The AFT models are discussed in details in text books (Cox, 1984; lawless, 1982; Hosmer and Lemeshow, 1999). The AFT models are named for the distribution of T rather than the distribution of ε_i or log T.

Table 3.2: Distribution of ε

Distribution of ε	*Distribution of T*
Extreme value (1 parameter)	*Exponential*
Extreme value (2 parameters)	*Weibull*
Logistic	*Log-logistic*
Normal	*Log-normal*
Log-Gamma	*Gamma*

The survival function of T_i can be expressed by the survival function of ε_i:

$$\begin{aligned} S_i(t) &= P(T_i \geq t) \\ &= P(\log T_i \geq \log t) \\ &= P(\mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + ... + \alpha_p x_{pi} + \delta\varepsilon_i \geq \log t) \\ &= P\left(\varepsilon_i \geq \frac{\log t - \mu - \alpha' x}{\delta}\right) \\ &= S_{\varepsilon_i}\left(\frac{\log t - \mu - \alpha' x}{\delta}\right) \end{aligned} \quad \text{.....(3.8)}$$

The distributions of ε_i and the corresponding distributions of T_i are summarized in Table (3.2).

The effect size for the AFT model is the time ratio. The time ratio comparing two levels of covariate $x_i (x_i = 1 \, v_s \; x_i = 0)$; after controlling all the other covariates is $\exp(\alpha_i)$ which is interpreted as the estimated ratio of the expected survival times for two groups. A time ratio above 1 for the covariate implies that this covariate prolongs the time to event, while a time ratio below 1 indicates that an earlier event is more likely. Therefore, the AFT models can be interpreted in terms of the speed of progression of a disease. The effect of the covariates in an accelerated failure time model is to change the scale, and not the location of a baseline distribution of survival times.

3.9.1 Estimation of AFT model

AFT models were fitted using the maximum likelihood method. The likelihood of the n observed survival times $t_1, \; t_2 \; ,..., \; t_n$ is given by

$$L(\alpha, \mu, \sigma) = \prod_{i=1}^{n} \{ f_i(t_i) \}^{\delta_i} \{ S_i(t_i) \}^{1-\delta_i}$$

Where $f_i(t_i)$ and $S_i(t_i)$ are the density and survival functions for the i^{th} individual at t_i and δ_i is the event indicator for the i^{th} observation. Using equation (3.8), the log-likelihood function is then given by

$$\log L(\alpha,\mu,\sigma)=\sum_{i=1}^{n}\left\{-\delta_i\log\left(\sigma t_i+\delta_i\log f_{\varepsilon_i}(z_i)+(1-\delta_i)\log S_{\varepsilon_i}(z_i)\right)\right\}$$

where $z_i=\left(\log t_i-\mu-\alpha_1x_{1i}-\alpha_2x_{2i}-...-\alpha_px_{p_i}/\delta\right)$. The maximum likelihood estimates of the unknown parameters μ, σ, α_1, α_2,..., α_p, are found by maximizing this function using the Newton- Raphson procedure in STAT, which is the same method used to maximize the partial likelihood in the Cox regression model.

3.9.2 Weibull AFT model

Weibull AFT model $W(\lambda,\gamma)$ distribution with scale parameter λ and shape parameter γ under AFT model the hazard function for the i^{th} individual is

$$\begin{aligned}h_i(t)&=\left[1/\eta_i(x)\right]h_0\left(t/\eta_i(x)\right)\\&=\left[1/\eta_i(x)\right]\lambda\gamma\left(t/\eta_i(x)\right)^{\gamma-1}\\&=1/\left[\eta_i(x)\right]^{\gamma}\lambda\gamma(t)^{\gamma-1}\end{aligned}$$

where $\eta_i=\exp\left(\alpha_1x_{1i}+\alpha_2x_{2i}+...+\alpha_px_{pi}\right)$ for individual i with p explanatory variables. So the survival time for the i^{th} patient is $W\left(1/\left[\eta_i(x)\right]^{\gamma}\lambda,\gamma\right)$. The weibull distribution has the AFT property.

If T_i has a Weibull distribution, then ε_i has an extreme value distribution (Gumbel distribution). The survival function of Gumbel distribution is given by

$$S_{\varepsilon_i}(\varepsilon) = \exp(-\exp(\varepsilon))$$

The AFT representation of the survival function of the Weibull model is given by

$$S_i(t) = \exp\left[-\exp\left(\frac{\log t - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \ldots - \alpha_p x_{pi}}{\sigma}\right)\right]$$
$$= \exp\left[-\exp\left(\frac{-\mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \ldots - \alpha_p x_{pi}}{\sigma}\right) t^{1/\sigma}\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.9)$$

From equation (3.6), the PH representation of the survival function of the Weibull model is given by

$$S_i(t) = \exp\left\{-\exp\left(\beta_1 x_{1i} + \ldots + \beta_p x_{pi}\right)\lambda t^{\gamma}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.10)$$

Comparing the above two formulas (3.9) and (3.10), we can easily see that the parameter λ, σ, β_j in the PH model can be expressed by the parameters μ, σ, α_j in the AFT model:

$$\lambda = \exp\left(\frac{-\mu}{\sigma}\right),\ \gamma = 1/\sigma,\ \beta_j = {}^{-\alpha_j}\!\!\diagup_{\sigma} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.11)$$

The AFT representation of hazard function of the Weibull model is given by

$$h_i(t) = \frac{1}{\sigma} t^{1/\sigma - 1} \exp\left(\frac{-\mu - \alpha_p x_{1i} - \ldots - \alpha_p x_{pi}}{\sigma}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.12)$$

The median survival time is

$$t_i(50) = \exp\left[\sigma \log(\log 2) + \mu + \alpha' x_i\right]$$

3.9.3 Log-logistic AFT model

One limitation of the Weibull hazard function is that it is a monotonic function of time. However, the hazard function can change direction in some situations. We will describe the log-logistic model in this section. The log-logistic survival and hazard function are given by

$$S(t) = \frac{1}{1+e^{\theta}t^{k}}, \quad h(t) = \frac{e^{\theta}kt^{k-1}}{1+e^{\theta}t^{k}}$$

Where θ and k are unknown parameters and k > 0. When k $\leq$ 1, the hazard rate decreases monotonically and when k > 1, it increases from zero to a maximum and then decreases to zero.

Suppose that the survival times have a log-logistic distribution with parameter θ and k, and then from equation (3.7), under the AFT model, the hazard function for the i^{th} individual is

$$\begin{aligned} h_i(t) &= (1/\eta_i)\, h_0(t/\eta_i) \\ &= \frac{e^{\theta}k(t/\eta_i)^{k-1}}{\eta_i\left(1+e^{\theta}(t/\eta_i)^{k}\right)} \\ &= \frac{e^{\theta-k\log\eta_i}kt^{k-1}}{1+e^{\theta-k\log\eta_i}t^{k}} \end{aligned}$$

Where $\eta_i = \exp(\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_p x_p)$ for individual i^{th} with p explanatory variables.

Therefore, the survival time for the i^{th} individual has a log-logistic distribution with parameter $\theta - k\log\eta_i$ and k. Log-logistic distribution has AFT property.

If the baseline survival function is $S_0(t) = \left\{1 + e^{\theta} t^{k}\right\}^{-1}$, where θ and k are unknown parameters, then the baseline odds of surviving beyond time t are given by

$$\frac{S_0(t)}{1 - S_0(t)} = e^{-\theta} t^{-k}$$

The survival time for the i^{th} individual also has a log-logistic distribution, which is

$$S_i(t) = \frac{1}{1 + e^{\theta - k\log\eta_i}\, t^{k}} \quad \text{.....(3.13)}$$

Therefore, the odds of the i^{th} individual surviving beyond time t is given by

$$\frac{s_i(t)}{1 - s_i(t)} = e^{\log\eta_i - \theta} t^{-k} \quad \text{.....(3.14)}$$

We can see that the log-logistic distribution has the proportional odds (PO) property. So this model is also a proportional odds model, in which the odds of an individual surviving beyond time t are expressed as

$$\frac{S_i(t)}{1 - S_i(t)} = \exp\left(\beta_1 x_1 + ... + \beta_p x_{pi}\right) \frac{S_0(t)}{1 - S_0(t)}$$

In a two-group study, using (3.14), the log (odds) of the i^{th} individual surviving beyond time t is

$$\log\left[\frac{S_i(t)}{1 - S_i(t)}\right] = \beta x_i - \theta - k\log t$$

Where x_i is the value of a categorical variable, which takes the value one in one group and zero in the other group. A plot of $\log\left[\frac{(1-S(t))}{S(t)}\right]$ versus $\log t$ should be linear if log-logistic distribution is appropriate. Therefore, we can check the suitability of log-logistic distribution using the (PO) property.

If T_i has a log-logistic distribution, then ε_i has a logistic distribution. The survival function of logistic distribution is given by

$$S_{\varepsilon_i}(\varepsilon) = \frac{1}{1+\exp(\varepsilon)}$$

Using equation (3.8), the AFT representation of survival function of the log-logistic model is given by

$$S_i(t) = \left[1 + t^{1/\delta}\exp\left(\frac{-\mu-\alpha_1 x_{1i} - .... - \alpha_p x_{pi}}{\sigma}\right)\right]^{-1} .....(3.15)$$

Comparing the formula (3.13) and (3.15), we can easily find a $\theta = \frac{-\mu}{\sigma}$, k=$\sigma^{-1}$.

According to the relationship of survival and hazard function, the hazard function the i^{th} individual is given by

$$h_i(t) = 1/\sigma t\left\{1 + t^{-1/\sigma}\exp\left(\frac{\mu+\alpha_1 x_{1i} + ... + \alpha_p x_{pi}}{\sigma}\right)\right\}^{-1} .....(3.16)$$

The median survival time is

$$t_i(50) = \exp\left(\mu + \alpha' x_i\right).$$

3.9.4 Log-normal AFT model

If the survival times are assumed to have a log-normal distribution, the baseline survival function and hazard function are given by

$$S_0(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right),\ h_0(t) = \frac{\phi\left(\frac{\log t}{\sigma}\right)}{\left[1 - \Phi\left(\frac{\log t}{\sigma}\right)\right]\sigma t}$$

Where μ and σ are parameter, $\phi(x)$ is the probability density function and $\Phi(x)$ is the cumulative density function of the standard normal distribution. The survival function for the i^{th} individual is

$$s_i(t) = s_0\left(t/\eta_i\right) = 1 - \Phi\left(\frac{\log t - \alpha' x_i - \mu}{\sigma}\right)$$

where $\eta_i = \exp(\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_p x_p)$. Therefore the log survival time for the i^{th} individual has normal $(\mu + \alpha' x_i, \sigma)$. The log-normal distribution has the AFT property.

In a two-group study, we can easily get

$$\Phi^{-1}(1 - S(t)) = 1/\sigma(\log t - \alpha' x_i - \mu),$$

Where x_i is the value of a categorical variable, which takes the value one in one group and zero in the other group. This implies that a plot of $\Phi^{-1}[1 - S(t)]$ versus $\log t$ will be linear if the log-normal distribution is appropriate.

3.10 Model checking for AFT models

The graphical methods can be used to check if a parametric distribution fits the observed data. Specifically, if the survival time follows an exponential distribution, a plot of $\log[-\log S(t)]$ versus $\log t$ should yield a straight line with slope of 1. If the plots are parallel but not straight, then PH assumption holds but not the Weibull. If the lines for two groups are straight but not parallel, the Weibull assumption is supported but the PH assumption is violated. The log-logistic assumption can be graphically evaluated by plotting $\log\left[\frac{(1-S(t))}{S(t)}\right]$ versus $\log t$. If the distribution of survival functions is log-logistic, then the resulting plot should be a straight line. For the log-normal distribution, a plot of $\Phi^{-1}(1-S(t))$ versus $\log t$ should be linear. All these plots are based on the assumption that the sample is drawn from a homogeneous population, implying that no covariates are taken into account. So this graphical method is not very reliable in practice. There are other methods to check the fitness of the model.

3.10.1 Using statistical criteria

We can use statistical tests or statistical criteria to compare all these AFT models. Nested models can be compared using the likelihood ratio test. The exponential model, the Weibull model and lognormal model are nested within gamma model. For comparing models that are not nested, the Akaike information criterion (AIC) were used instead, which is defined as

$$AIC = -2l + 2(k + c)$$

Where l the log-likelihood, k is is the number of covariates in the model and c is the number of model-specific ancillary parameters. The addition of 2(k +c) can be thought of as a penalty if nonproductive parameters are added to the model. Lower values of the

AIC suggest a better model. But there is a difficulty in using the AIC in that there are no formal statistical tests to compare different AIC values.

3.10.2 Using Residual Plots

Residual plots can be used to check the goodness of fit of the model. Procedures based on residuals in the AFT model are particularly relevant with the Cox PH model. One of the most useful plots is based on comparing the distribution of the Cox-Snell residuals with the unit exponential distribution. The Cox-Snell residual for the i^{th} individual with observed time t_i is defined as

$$\gamma_{c_i} = \hat{H}(t_i / x_i) = -\log\left[\hat{S}(t_i / x_i)\right],$$

Where t_i is the observed survival time for individual i, x_i is the vector of covariate values for individual i, and $\hat{S}(t_i)$ is the estimated survival function on the fitted model. From equation (3.8), the estimated survival function for the i^{th} individual is given by

$$\hat{S}_i(t) = S_{\varepsilon_i}\left(\frac{\log t - \mu - \alpha\, x_i}{\hat{\sigma}}\right)$$

where $\hat{\mu}$, $\hat{\alpha}$ and $\hat{\sigma}$ are the maximum likelihood estimator of μ, α and σ respectively, $S_{\varepsilon_i}(\varepsilon)$ is the survival function of ε_i in the AFT model, and $\frac{\log t - \hat{\mu} - \hat{\alpha} x_i}{\sigma} = \gamma_{\varepsilon_i}$ is referred to as standardized residual.

The Cox-Snell residual can be applied to any parametric model. The corresponding form of residual based particular AFT model can be obtained. For example, under the Weibull AFT model, since $S_{\varepsilon_i}(\varepsilon) = \exp\left(-e^{\varepsilon}\right)$, the Cox-Snell residual is then

$$\gamma_{\varepsilon_i} = -\log\left\{\hat{S}(t_i)\right\} = -\log S_{\varepsilon_i}(\gamma_{s_i}) = \exp(\gamma_{s_i})$$

Under the log-logistic AFT model, since $S_{\varepsilon_i}(\varepsilon) = (1+e^{\varepsilon})^{-1}$, the Cox-Snell residual is then

$$\gamma_{c_i} = \log\left[1+\exp(\gamma_{s_i})\right].$$

If the fitted model is appropriate, the plot of $\log(-\log S(\gamma_{c_i}))$ versus γ_{c_i} is a straight line with unit slope through the origin. These residuals lead to the deviance residuals for the particular AFT model. A plot of deviance residuals against the survival time or explanatory variables will be used to check whether there are particular times, or particular values of explanatory variables, for which the model is not a good fit.

3.11 Ethical considerations

Permission to undertake this study was obtained from College of Natural Science through Ethical Review Board and official letter of co-operation will be written by the Department of Statistics to Jimma University Specialized Hospital.

CHAPTER FOUR

4. RESULTS AND ANALYSIS

4.1. Summary Statistics and Non-Parametric Analysis

A total of 218 HIV positive children (who are lessthan15 years of age) who started ART at Jimma University referral hospital from 2005 to 2013 were included in the study. Out of the total of 218 participants in the study 30 (13.76%) children died due to HIV/AIDS, 22 (10.1%) were transferred to other hospitals, 5 (1.83%) were lost to follow up, 12 (5.51%) were drop out while 150 (68.80%) remained alive during the time of data collection. The response variable was the length of time from start of treatment to death. The mean and median survival time of the entire observations was found to be 40.15 and 38 months, respectively, with standard deviation 26.187 months. The minimum survival time was 1 month and the maximum survival time was 107 months.

Exploratory analysis including descriptive statistics was used to get some information about the distributions of the variables based on baseline characteristics in 218 participants. Two continuous and ten categorical baseline covariates are included in the analysis. The average baseline age and weight of the patients are 6.31 years (with a standard deviation of 4.51 years) and 14.77 kilograms (with a standard deviation of 7.24 kilograms) respectively.

Table 4.1: Summary statistics for continuous variables included in the study of HIV patients under ART in JUSH, Jimma, 2005- 2013.

Status of Patients	*Continuous Variables*	*Mean*	*Std.dev*	*Minimum*	*Maximum*	*Median*	*Q1*	*Q3*
	Time	*40.15*	*26.186*	*1*	*107*	*38*	*19*	*57.5*
	Age	*6.31*	*4.5155*	*1*	*14*	*6*	*6*	*9*
	Weight	*14.77*	*7.243*	*4*	*50*	*13.85*	*9.5*	*18.62*

Out of the total 218 ART followers, 117 (53.7%) were male. More than half of the patients 155 (71.1%) were live in urban areas while 34 (15.6 %) were live in semi- urban areas and 29 (13.3%) live in rural areas. Only 5 (2.3%) of pediatrics families were free of HIV/AIDS. 114 (52.3%) were at clinical stage IV and III, and the rest 104 (47.7%) were at clinical stage II and I when they started HAART. About 146 (67 %) of patients had their CD4 count greater than 200 mm^3. There were 42 (19.3%) patients who were able to work, 141 (64.7%) were ambulatory and 35 (16.1 %) were bedridden. The TB screen tests show 172 (78.9%) have developed TB and 46 (21.1%) didn't. 196 (89.9%) have not take TB preventive treatment while the remaining 22 (10.1%) take TB preventive treatment. Only 13 (6%) of the patients were take AZT based regimen and the remaining 205 (94%) take 4dT based regimen out of which 81 (37.2%) develop opportunistic infections. All the detail results have been summarized in Table 4.2 below. For comparing the survival experiences between groups, the log-rank test is applied to all categorical variables. Table 4.2 below shows that the different groups of gender, place of residence, TB preventive treatment and ART regimen are statistically not equal in experiencing the death event, whereas levels of family history of HIV/AIDS, functional status, WHO clinical stage, TB screen and opportunistic infections are statistically the same in experiencing the event death. The log-Rank test results suggest that Family history of HIV/AIDS (p=0.000), functional status (p=0.001), WHO clinical stage (p=0.000); TB screen (p=0.001) and opportunistic infections (0.000) groups are significant covariates whose different levels have an impact in the survival longevity of

HIV patients; while the other do not have an impact. The K-M curves also show the same result as the log-rank test in (appendix A).

Table 4.2: Frequency distribution for Baseline characteristics of HIV infected children taking HAART at JUSH based on the variables under study and log-rank test results, 2005-2013.

Covariate	*Category*	*Death*	*Death (%)*	*Total*	*Percent*	*p-value*
Gender	*Female*	*14*	*13.9*	*101*	*46.3*	*0.945*
	Male	*16*	*13.7*	*117*	*53.7*	
Place of residence	*Urban*	*19*	*12.3*	*155*	*71.1*	*0.509*
	Semi-Urban	*6*	*17.6*	*34*	*15.6*	
	Rural	*5*	*17.2*	*29*	*13.3*	
Family history HIV	*Yes*	*26*	*22.2*	*213*	*97.7*	*0.000*
	No	*4*	*80.0*	*5*	*2.3*	
Functional status	*Working*	*13*	*31*	*42*	*19.3*	*0.001*
	Ambulatory	*12*	*8.5*	*141*	*64.7*	
	Bedridden	*5*	*24.3*	*35*	*16.1*	
WHO clinical stage	*Stage IV/III*	*26*	*22.8*	*114*	*52.3*	*0.000*
	Stage II/I	*4*	*3.8*	*104*	*47.7*	
CD4 count	*< 200mm^3*	*20*	*27.8*	*72*	*33*	*0.000*
	≥200mm^3	*10*	*6.8*	*146*	*67*	
TB screen	*Yes*	*18*	*36.0*	*50*	*22.9*	*0.001*
	No	*12*	*7.1*	*168*	*77.1*	
TB/Prophylaxis	*Yes*	*32*	*13.6*	*21*	*10.1*	*0.701*
	No	*27*	*13.8*	*196*	*89.9*	
ART Regimen	*AZT based*	*4*	*30.8*	*13*	*6*	*0.098*
	4dT based	*26*	*12.7*	*205*	*94*	
Opportunistic infections	*Yes*	*20*	*27.2*	*75*	*37.2*	*0.000*
	No	*8*	*5.8*	*143*	*62.8*	

4.2 Standard Cox PH model

We use univariate analysis to check all the risk factors before proceeding to more complicated models. We use a univariate Cox proportional hazards regression for every potential risk factor. The Wald test is considered in each univariate Cox PH model. Variables are identified as significant using a 0.1 significance level in the univariate model. We then fit the full multi-variable Cox PH model including all the potential risk factors. Consequently, in the univariate Cox proportional hazards models the model with a single covariate, Age, weight, CD4 count, functional status, WHO clinical stage, TB screen and opportunistic infections show a statistically significant association with the survival time. But other characteristics such as sex, place of residence, family history of HIV/AIDS, TB Prophylaxis/treatment, and ART regimen are not statistically significant, suggesting that these variables are not associated with the survival time revealing that these variables will not be included in the multivariate model. Therefore, we will consider the model that includes all the significant predictors.

The categorical predictor functionlal status has three levels and therefore we will include this predictor using two dummy variables (Ambulatory; Bedridden) with the Working as the reference group, WHO clincal stage having two categories and therefore we used one dummy variable (stage IV&III) with (stage I & III) as reference. By the same fashion TB screen has two levels with a single dummy variable (Yes) taking (No) as reference. Opportunistic infections has also two levels with a single dummy variable (Yes) taking (No) as referance.

Table 4.3: Uni-variable and multi-variable Cox PH model for the relative hazard of survival time for HIV (+) pediatrics under HARRT at JUSH from, 2005-2013.

Covariate	***Uni-variable Analysis***				***Multi-variable Analysis***			
	B	***HR***	***p-value***	***95% CI***	***B***	***HR***	***p-value***	***95% CI***
Age	*-0.271*	*0.763*	*0.001*	*(0.664, 0.876)*	*-0.1510*	*0. 860*	*0.12*	*(0.71,1.04)*
Weight	*-0.244*	*0.783*	*2.32e-06*	*(0.708, 0.867)*	*-0.121*	*0.886*	*0.054*	*(0.78,1.002)*
CD4 count								
<200 mm3	*1.543*	*4.68*	*7.29 e-06*	*(2.183,10.04)*	*1.691*	*5.425*	*1.95e-04*	*(2.23,13.21)*
Functional Ambulatory	*-1.387*	*0.249*	*5.95e-04*	*(0.113, 0.551)*	*-1.516*	*0.219*	*0.001*	*(0.08,0.56)*
Bedridden	*-0.711*	*0.491*	*0.181*	*(0.173,1.393)*	*-0.809*	*0.444*	*0.14*	*(0.15,1.32)*
WHO clinical Stage III/IV	*1.868*	*6.474*	*5.12e-04*	*(2.257, 18.57)*	*2.287*	*9.366*	*1.61e-04*	*(2.93,29.93)*
TB screen Yes	*-1.140*	*0.32*	*1.97 e-03*	*(0.155, 0.658)*	*1.347*	*3.849*	*1.92e-03*	*(1.64,9.01)*
Opportunistic infections Yes	*1.508*	*4.52*	*3.06 e-04*	*(1.993,10.25)*	*1.439*	*4.439*	*2.46e-03*	*(1.66,10.71)*
Log likelihood								*-103.78*
AIC								*223.57*

AIC Akaike Information Criterion; β: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%CI: 95% confidence interval for HR

From Table 4.3 above, Age and weight are significant at Uni-variable model at 5 % but not significant at multi-variable model. Then by removing age from Table 4.3, we get the model containing weight, CD4 counts, functional status, WHO clinical stage, TB screen and opportunistic infections variables are significant. But the model is not best as compared with the above model because the AIC value is relatively high.

Table 4.4: Uni-variable and multi-variable Cox PH model for the relative hazard of survival time for HIV (+) pediatrics under HARRT at JUSH from, 2005-2013.

Covariate	***Uni-variable Analysis***				***Multi-variable Analysis***			
	β	***HR***	***p-value***	***95% CI***	***β***	***HR***	***p-value***	***95 CI%***
Weight	*0.244*	*0.783*	*2.32e-06*	*(0.708,0.867)*	*-0.185*	*0.831*	*9.05e-05*	*(0.757,0.912)*
CD4 count <200 mm3	*0.543*	*4.68*	*7.29e-06*	*(2.183,10.04)*	*1.477*	*4.379*	*8.80e-04*	*(1.83,10.45)*
Functional Ambulatory	*-1.387*	*0.249*	*5.95e-04*	*(0.113,0.551)*	*-1.398*	*0.247*	*2.26e-03*	*(0.101,0.606)*
Bedridden	*-0.711*	*0.491*	*0.181*	*(0.173,1.393)*	*-0.805*	*0.447*	*0.149*	*(0.149,1.337)*
WHOclinicalStage III/IV	*1.868*	*6.474*	*5.12e-04*	*(2.257,18.57)*	*2.056*	*7.815*	*3.26e-03*	*(2.549,23.95)*
TBscreen Yes	*1.632*	*5.11*	*1.33e-05*	*(2.45,10.66)*	*1.292*	*3.641*	*2.56e-03*	*(1.572,8.432)*
Opportunistic infections Yes	*1.508*	*4.52*	*3.06e-04*	*(1.993,10.25)*	*1.635*	*5.134*	*5.31e-04*	*(2.035,12.95)*
Log likelihood								*-105.05*
AIC								*224.12*

AIC: Akaike Information Criterion; β: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%CI: 95% confidence interval for HR

After we built a multi-variable model of main effects of variables we then checked the adequacy of the fitted model, including the PH assumption and the goodness of fit. The PH assumption checking with graphical method and two statistical test methods (adding time-dependent covariates in the Cox model and tests based on the Schoenfeld residuals) have been used. The log (-log (survival)) versus survival time plot used to check the PH assumption for all the categorical variables. In (appndix B) the graphs for each of the categorical variables display lines that appeared to be parallel implying that the proportional-hazards assumption among categorical variables WHO clinical stage and opportunistic infections has not been violated but CD4 counts and TB screen has been seems violeted . We also create the time-dependent covariate in (Table 4.5) by creating

interactions of the predictors and survival time and include them in the model. The result indicates that the PH assumption for opportunistic infections is violated (p-value for ols*log(time) are less than 0.05 i.e. 0.006). The Schoenfeld residuals are also used to check the PH assumption in (Table 4.5). The p-value for testing whether the correlation between Schoenfeld residual for this covariate and ranked survival time is zero. The p-values for weight and TB screen are less than 0.05 and all the other covariates are greater than 0.05, suggesting that the PH assumption is violated for weight and TB screen, but reasonable for all the other covariates. Graphically we can see from appendix C.

Table 4.5: Statistical test for proportional hazards assumption (PH) of the covariates and their interaction with log of time and schonfeld residual

Covarites	***Covarite interaction with log(time)***				***Schoenfeld Residual***
	β	***HR***	***p-value***	***95% CI***	***Covarites***
weight:log(time)	*0.017*	*1.059*	*0.087*	*(0.992,1.132)*	*0.002*
CD4:log(time)	*0.127*	*1.136*	*0.738*	*(0.538,2.397)*	*0.225*
Ambulatory:log(time)	*-0.352*	*0.703*	*0.508*	*(0.247, 1.998)*	*0.385*
Bedridden :log(time)	*1.256*	*3.510*	*0.056*	*(0.965, 1.276)*	*0.985*
Stage III&IV:log(time)	*0.186*	*1.205*	*0.731*	*(0.416, 3.489)*	*0.669*
TB Yes:log(time)	*-0.408*	*0.6647*	*0.327*	*(0.294, 1.506)*	*0.036*
Ols yes:log(time)	*1.517*	*4.559*	*0.006*	*(1.514, 1.313)*	*0.777*
Global test					*0.003*

β: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%CI: 95% confidence interval for HR

We assess goodness of fit by residual plots (Section 3.6). A plot of the Cox-Snell residuals against the cumulative hazard of Cox-Snell residuals is presented in Figure 4.1. There is some evidence of a systematic deviation from the straight line, which gives us some concern about the adequacy of the fitted model. The plot of deviance residual against the risk score shows that the deviance residuals seem not to be symmetrically distributed about zero. There are very high or very low deviance residuals which suggest that these observations may be outliers (Figure 4.3). Therefore, we have some concern

about the adequacy of the fitted Cox PH model. We also use delta-beta statistic to measure the influential observations on the model as a whole. It shows that the coefficients do not change too much when the observations corresponding to the largest delta-beta statistics are removed. Therefore, we do not remove them from the dataset and conclude that there are no influential observations (Figure 4.2). Lastly, we can say that applying Cox proprtional hazards for pediatrics data is not suggested because the basic assumptions of Cox prortional hazards are violated. The residuals also support the violation of assumptions. So we don't need to talk about the hazards ratio and the relation of covariates with survival time of HIV positive pediatrics. It is better to apply stratified Cox PH , Cox with time varying and AFT models.

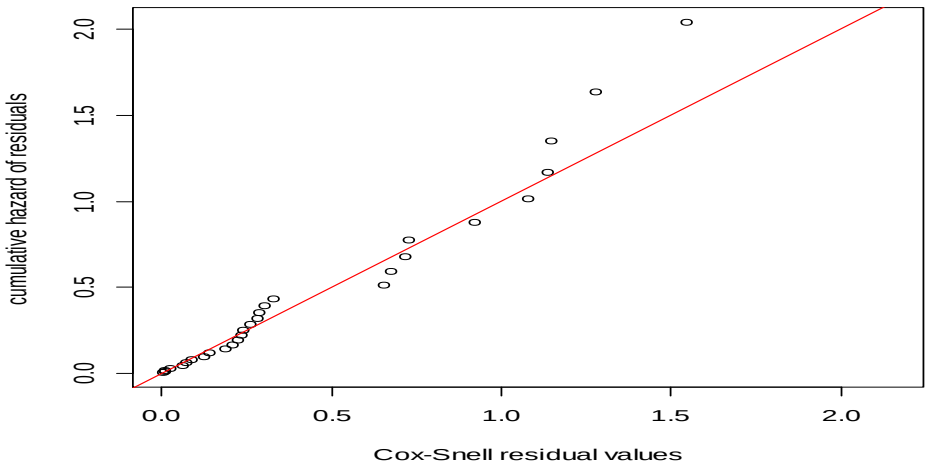


Figure 4.1: Cumulative hazard plot of the Cox-Snell residual for Cox PH model

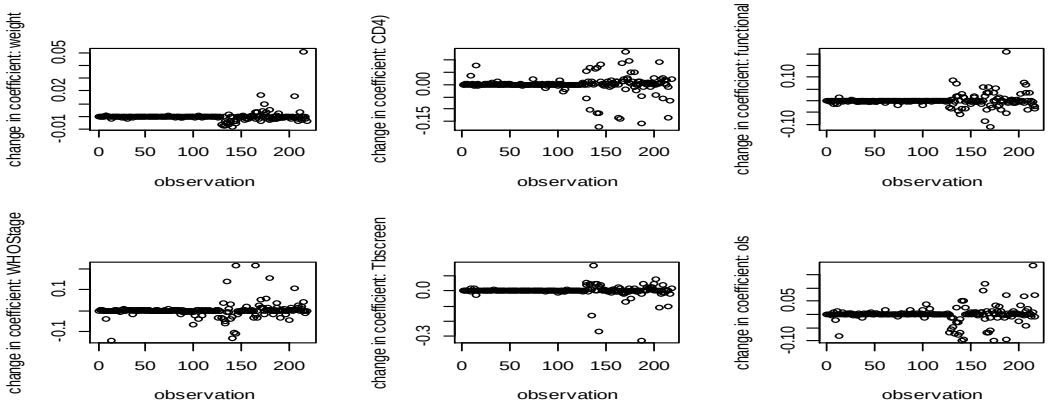


Figure 4.2 Index plots of dfbeta for the multivariate Cox regression moedel

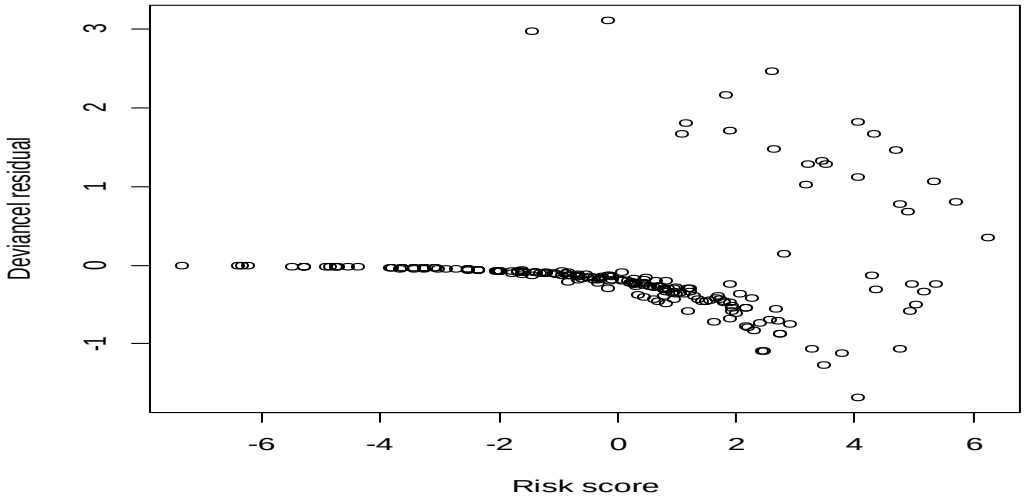


Figure 4.3: Deviance residuals plotted against the risk score for Cox PH model

4.3 Cox mdel with time-dependent variables

We have shown that the Cox model displayed nonproportionality for variables weight and TB/ HIV co-infected .We believe that there is an interaction between weight and time. It is not appropriate to use stratified Cox model because weight is a continuous variable. We then incorporate time-varying covariate in to the model. The result beleow shows that except the intaraction of TB/HIV co-infected with time all covariates have statisticaly singificant impact on the survival time of pediatrics. Based on log likilhood and AIC criteria Cox with time varying is considered as best as compared with standard Cox PH model.

Table 4.6: Multivariable Cox withtime-varying model for survival time of HIV positive pediatrics.

Covariate	*Coef*	*HR*	*Std. Err*	*P-value*	*95% CI*
Weight	*-0.329*	*0 .719*	*0.057*	*0.00*	*(0.614,0.841)*
CD4 count <200 mm3	*1.056*	*2.875*	*1.303*	*0.020*	*(1.182 ,6.990)*
Functional Ambulatory	*-1.076*	*0.340*	*0.170*	*0.032*	*(0.127, 0.910)*
Bedridden	*-0.762*	*0.466*	*0.257*	*0.167*	*(0.158, 1.374)*
WHO clinical Statge III/IV	*1.947*	*7.014*	*4.045*	*0.001*	*(2.265,21.719)*
TB screen Yes	*1.057*	*2. 87*	*1.487*	*0.041*	*(1.045 7.927)*
Ols Yes	*1.624*	*5.074*	*2.412*	*0.001*	*(1.998 12.884)*
Tvc					
TB screen Yes	*1.014*	*0.013*	*0.0194*	*0.466*	*(0.976 1.052)*
weight	*0.007*	*1.007*	*0.002*	*0.002*	*(1.002,1.011)*
Log Liklihood					*-98.189*
AIC					*216.32*

AIC: Akaike Information Criterion; Coef*: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%CI: 95% confidence interval for HR, Std.Err: standared error, tvc: time variying covariate.*

4.4 AFT and PH models

The accelerated failure time (AFT) model is the alternative of the Cox PH model because the PH assumption is violated. The AFT model can be used to express the magnitude of effect in a more accessible way in terms of difference between variables in survival time. We fitted the data set using Exponential, Weibull, Log-Logistic, Log-Normal and Gompertz model. For each kind of model, we fitted both the univariate and multi-variable

AFT model for those statistically significant at 10 % level. In multi-variable AFT models except age and functional status of bedriddn all variables show a statistically significant association with the survival time. However, weight is not statistically significantly associated with survival time for weibull model. The results from the different AFT models applied to the survival time of HIV patientes are presented in Table 4.7 and Table 4.8 . There is no big difference for the estimations in different models.

Table 4.7: Results from AFT models and Gompertz model for time-to-death uni-variable analysis.

Covariate	*Exponetial*		*Weibull*		*Log Normal*		*Log Logistic*		*Gompertz*	
	Coef(TR)	*AIC*	*Coef(TR)*	*AIC*	*Coef(TR)*	*AIC*	*Coef(TR)*	*AIC*	*HR*	*AIC*
Age	*0.28(1.32)*	*275.9*	*0.53(1.69)*	*258.7*	*0.48(1.61)*	*254.4*	*0.5(1.64)*	*257.8*	*0.75*	*264.5*
Weight	*0.29(1.33)*	*255.2*	*0.43(1.53)*	*243.5*	*0.31(1.36)*	*243.3*	*0.4(1.49)*	*243.2*	*0.77*	*249.0*
CD4count <200mm^3	*-1.63(.19)*	*277.0*	*-2.88(.05)*	*261.6*	*-2.83(.06)*	*261.2*	*-2.9(.054)*	*261.6*	*0.21*	*268.6*
Functional										
Ambulatoy	*1.28(3.59)*	*288.0*	*2.68(14.6)*	*269.5*	*3.19(24.3)*	*261.5*	*3.01(20.8)*	*266.8*	*0.22*	*275.0*
Bedridden	*0.44(1.55)*		*1.29(3.63)*		*1.51(4.48)*		*1.54(4.66)*		*0.46*	
WHO clinical										
StageIII/I	*-1.83(0.16)*	*278.9*	*-3.56(0.03)*	*261.3*	*-3.37(0.03)*	*258.4*	*-3.56(0.03)*	*260.7*	*6.61*	*267.5*
TB screen										
Yes	*-1.67(0.19)*	*276.1*	*-3.24(0.04)*	*258.4*	*-3.18(0.04)*	*257.5*	*-3.25(0.04)*	*258.4*	*5.71*	*264.1*
Ols Yes	*-1.35(0.25)*	*283.8*	*-3.00(0.05)*	*263.5*	*-2.94(0.05)*	*260.9*	*-3.03(0.05)*	*263.1*	*4.92*	*278.7*

AIC: Akaike Information Criterion, Coef: coefficient for covariate, HR: hazard ratio; TR: time ratio

Table 4.8: Results from AFT models and Gompertz model for time-to-death multi-variable analysis.

Covariates	***Exponetial***		***Weibull***		***Log Normal***		***Log Logistic***		***Gompertz***	
	Coef(TR)	*p-value*	*Coef(TR)*	*p-value*	*Coef(TR)*	*p-value*	*Coef(TR)*	*p-value*	*(HR)*	*p-value*
Age	*0.8(1.19)*	*0.06*	*0.22(1.24)*	*0.09*	*0.21(1.23)*	*0.12*	*0.24(1.27)*	*0.06*	*0.84*	*0.08*
Weight	*0.13(1.14)*	*0.03*	*0.15(1.16)*	*0.06*	*0.16(1.17)*	*0.04*	*0.16(1.17)*	*0.03*	*0.88*	*0.05*
CD4 count										
<200 mm^3	*-2.19(0.11)*	*0.00*	*-2.47(0.08)*	*0.00*	*-1.85(0.16)*	*0.003*	*-2.08(0.12)*	*0.00*	*0.15*	*0.00*
Functionl										
Ambulatory	*1.91(6.75)*	*0.00*	*2.13(8.41)*	*0.001*	*1.75(5.75)*	*0.01*	*1.88(6.55)*	*0.004*	*0.17*	*0.00*
Bedridden	*0.88(2.41)*	*0.12*	*1.1(3)*	*0.142*	*1.49(4.44)*	*0.06*	*1.29(3.63)*	*0.07*	*0.4*	*0.104*
WHOclinical										
Stage III/IV	*-2.57(0.07)*	*0.00*	*-3.03(0.05)*	*0.00*	*-2.40(0.09)*	*0.006*	*-2.9(0.06)*	*0.00*	*11.2*	*0.00*
TB screen										
Yes	*-1.65(0.19)*	*0.00*	*-2.03(0.13)*	*0.001*	*-2.20(0.11)*	*0.001*	*-1.99(0.14)*	*0.002*	*4.54*	*0.001*
Ols										
Yes	*-1.37(0.25)*	*0.006*	*-1.79(0.17)*	*0.008*	*-2.07(0.12)*	*0.003*	*-1.84(0.16)*	*0.006*	*4*	*0.0004*
Intercept	*4.41*	*0.00*	*7.40*	*0.00*	*4.42*	*0.00*	*4.23*	*0.00*		
Scale			*1.318*		*0.74*		*0.08*			
Shape			*0.758*		*2.11*		*1.09*		*-0.018*	*0.110*

AIC: Akaike Information Criterion, Coef: coefficient for covariate, HR: hazard ratio; TR: time ratio

Table 4.9: Akaike Information Criterion (AIC) in the AFT models and Gompertz model

Distribution	*N0 of parameters*	*Log-likelihood*	*K*	*C*	*AIC*
Exponential	*1*	*-86.52*	*8*	*1*	*191.03*
Weibull	*2*	*-84.60*	*8*	*2*	*189.20*
Log-Normal	*2*	*-87*	*8*	*2*	*195.44*
Log-logistic	*2*	*-85.628*	*8*	*2*	*191.24*
Gompertz	*2*	*-85.09*	*8*	*2*	*190.19*
Standared Cox-PH		*-103.788*	*8*		*223.57*
Cox PH with time variying coefficient		*-98.189*	*7*		*216.32*

AIC: Akaike Information Criterion; C: is the no of model-specific ancillary parameters;

k: is the no of covariates in the model.

We compared all these AFT models including Gompertz and Cox-PH model using statistical criteria (log likelihood and AIC). The nested AFT models can be compared using the likelihood ratio (LR) test. The exponential model, the Weibull model and the log-normal model are nested within the gamma model but here the gamma model has not converged . In this case, we used only AIC to compare the models (Table 4.10).

The Weibull AFT model appears to be an appropriate AFT model according to AIC compared with other AFT models in multi-variable analysis, although it is only slightly better than exponential and Gompertz model. However the Log -normal AFT model is best in the uni-variable analysis but has poor fit according to AIC in multi-variable analysis. Furthermore, we check the goodness of fit of the model using residual plots. Cumulative hazard plot of the Cox-Snell residuals in AFT models are presented in Figure 4.4. The plotted points lie on a line that has a unit slope and zero intercept for weibull model. So there is no reason to doubt the suitability of this fitted weibull model. At last,we conclude that the weibull model is the best fitting the AFT model based on AIC criteria and residuals plot containing the statistically significant covariates functional status of ambulatory, CD4 count, WHO clinical statge,TB screen and opportunistic infections.

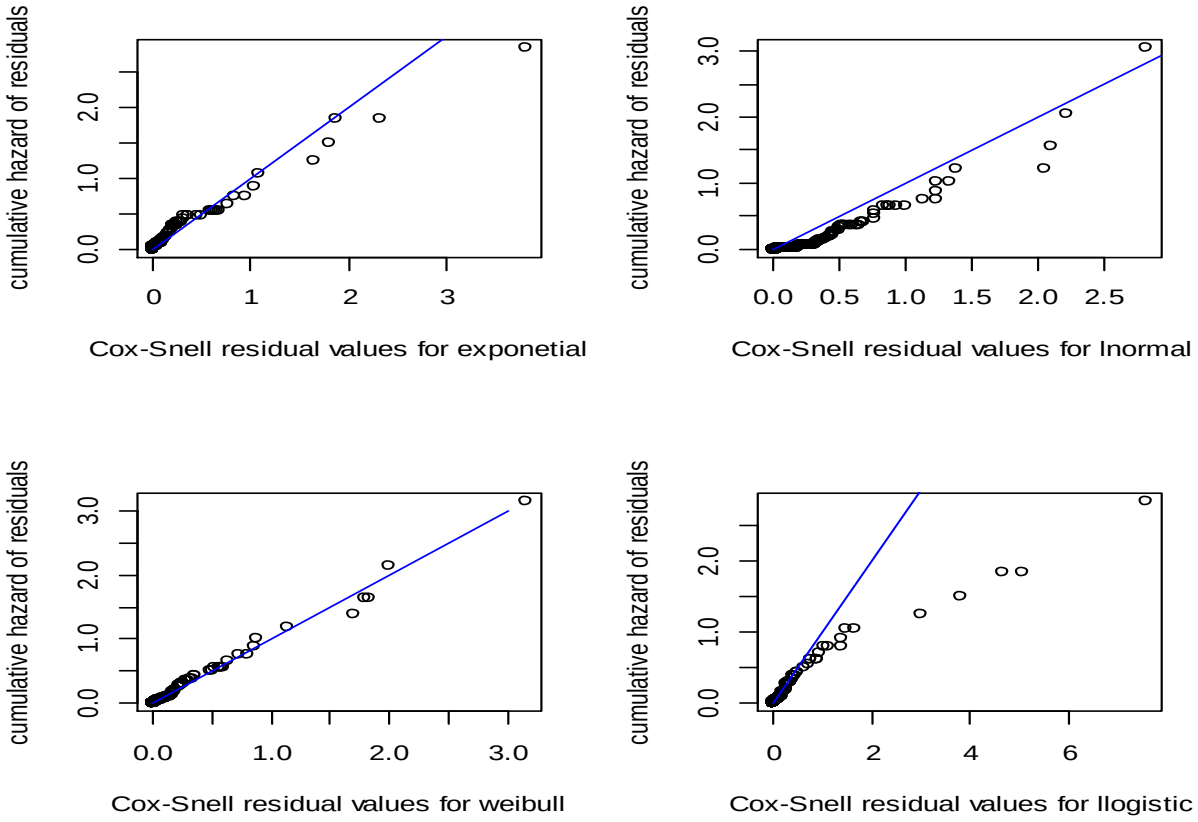


Figure 4.4: Cox-Snell residual plots of staging subgroup for exponential model; Weibull model; lognormal model and log-logistic model comparing these graphs, the straight line in the Weibull plot appears to provide the best fit to the pediatric data.

The estimated acceleration factor for an individual in ambulatory group relative to an individual in working group is 8.41. This indicates that the effect of prolongs the time to death. The CD4 count less than 200 mm^3 in a HIV-infected patient speeds up mortality than that higher than 200mm^3. The patient with WHO clinial statge III and IV has shorter survival time than patients with WHO clincal stage I and II. TB co-infected patient has shorter survival time than not co-infected patient. Finally the patient with extra opportunistic infections speed up mortality than not infected.

The exponential and Weibull AFT models are also PH models. The signs of the coefficients in the AFT model are opposite to the signs for the PH model. The estimate of shape parameter in Weibull model is 0.758 which is less than 1, and the 95% CI is (1.05,1.77) which does not cover the null value of 1. This suggests that the Weibull model may be better than the exponential model. Therefore we only compare Weibull AFT model and PH model here. Using equation (3.8), we can calculate the regression coefficients in Weibull PH model. The correspondence between the parameters of the Weibull PH and AFT models are presented in (Table 4.10). From equation 3.9

$$\hat{S}_i(t) = \exp\left\{-t^{1/\hat{\sigma}}\hat{\zeta}_i\right\}$$

Where $\hat{\zeta}_i = \dfrac{-\mu - \alpha x_i}{\hat{\sigma}}$

$$= \frac{1}{1.318}\begin{pmatrix} -7.4 - 0.22age - 0.15weight + 2.47cd4 - 2.13ambulatory - 1.1bedridden + \\ 3.03statgeIII/IV + 2.03tb + 1.79ols \end{pmatrix}$$

From equation (3.9), the estimated hazard function for the i^{th} indvidual is given by

$$\hat{h}_i(t) = \frac{1}{\hat{\sigma}} t^{\frac{1}{\hat{\sigma}}-1}\exp\left(\hat{\zeta}_i\right) = \frac{1}{1.318} t^{\frac{1}{1.312}-1}\exp\left(\hat{\zeta}_i\right) \quad \text{.....(4.1)}$$

We can also obtain this by the PH representation of the Weibull model. The estimated hazard function for the i^{th} individual is

$$\hat{h}_i(t) = \lambda\gamma t^{\gamma-1}\exp\left(\beta' x_i\right)$$

$$= (0.0036)(0.76)t^{-0.24}\exp\begin{pmatrix} -0.16age - 0.11weight + 1.87cd4 - 1.617ambulatory - 0.82bedridden \\ +2.30stageIII\ /\ IV + 1.53tb + 1.36ols \end{pmatrix}$$

This turns out to be the same as equation (4.1).

Table 4.10: Comparison of Weibull PH and AFT model

Covariate	***Weibull PH***				***Weibull AFT***			
	Coef	***se***	***HR***	***95% CI***	***Coef***	***se***	***TR***	***95% CI***
Age	*-0.16*	*0.08*	*0.84*	*(-0.69,1.002)*	*0.22*	*0.16*	*1.24*	*(0.96,1.61)*
Weight	*-0.11*	*0.05*	*0.89*	*(0.78,1.004)*	*0.15*	*0.09*	*1.16*	*(0.99,1.36)*
CD4 count	*1.87*	*3.01*	*6.51*	*(2.63,16.11)*	*-2.47*	*5.29*	*0.084*	*(2.47,28.83)*
Ambulatory	*-1.617*	*0.09*	*0.19*	*(0.07,0.511)*	*2.13*	*2.20*	*2.97*	*(0.69,12.74)*
Bedridden	*-0.82*	*0.24*	*0.43*	*(0.14,1.29)*	*1.1*	*0.04*	*3*	*(0.98,13.45)*
WHO clinical satage	*2.30*	*5.94*	*9.99*	*(3.11,32.06)*	*-3.03*	*0.08*	*0.05*	*(0.03,0.44)*
TB screen	*1.53*	*2.02*	*4.66*	*(1.99,10.91)*	*-2.03*	*0.05*	*0.13*	*(0.02,0.27)*
Ols	*1.36*	*1.84*	*3.9*	*(1.54,9.83)*	*-1.79*	*0.11*	*0.17*	*(0.04,0.62)*
Constant				*(-0.57,0.019)*	*7.4*	*0.15*		*(-0.57,0.019)*
Scale	*1.318*			*(0.56,0.947)*	*1.318*	*0.19*		*(0.56,0.947)*
Shape	*0.758*			*(1.05,1.77)*	*0.758*	*0.11*		*(1.05,1.77)*

CHAPTER FIVE

5. DISCUSSION AND CONCLUSION

5.1. Discussion

In medical science, researchers are more interested in Cox PH model than other parametric models to estimate the survival model, mainly due to the less assumption required in the model. We require some hypotheses to analyze the survival by means of Cox model. This model develops that any change at independent variables level in the hazard function is independent of time. The hypotheses required for modeling Cox hazard model may not work in many conditions, especially in biomedical fields (Cox, 1984). If these hypotheses do not work,the results obtained from Cox model may be invalid. One solution is to include the time-dependent variable for the predictors with non-proportional hazards. When this approach is used to account for a variable with non-proportionality, different results may be obtained from different choices of time-dependent variables. It is difficult to choose between models. Alternatively we can use a model where we stratify on the non-proportional predictors. The stratified Cox model is not appropriate when the covariate with non-proportionality is continuous or of direct interests. And both ways are still based on comparison of hazards. The AFT model is an alternative method for the analysis of survival data even when hazards are not proportional. Unfortunately, a study on the survival analysis dealing with cancer reporeted that only five percent of the researches that used Cox model had checked its hypotheses (Altman,1995). Moreover, various parametric models like Weibull, log-normal and log-logistic are widely used in the analysis of survival data. These models can interpret the survival time based on a specific distribution irrespective of proportional hazard hypothesis.

If the survival times use Weibull or exponential distribution, the analysis with parametric models will be stronger. This means that, under special conditions, parametric models

such as Weibull, log-logistic and lognormal may have more accurate results than Cox model (Pourhoseingholi, 2011). The population of survival times usually has an exponential or Weibull distribution; therefore, a parametric model is more efficient and similar than its corresponding non-parametric or semi-parametric models. Moreover, it has more flexibility in adding covariates to the model.

In this thesis, we have analyzed the pediatrics HIV dataset using these alternative methods. This study provides an example of a situation where the PH model is inappropriate and where the AFT model provides a better description of the survival data. Based on the results of model diagnostics, the PH assumption does not hold in this dataset. After fitting the Cox PH model, the goodness of fit of the model is assessed through residual plots. The PH model seems to display lack of fit. In contrast, the AFT model provides an adequate description of the data. The family of the AFT models containing the exponential AFT model, Weibull AFT model, log-logistic AFT model and log-normal AFT model are applied to this data set.We select the model that best describes the data. In addition, the example illustrates that the AFT model has a more realistic interpretation and provides more informative results as compared to PH model. Therefore, we suggest that using the Cox PH model may not be the optimum approach. The AFT model may provide an alternative method to fit survival data. By using AIC value, the models developed from exponential, Weibull, log-normal, log-logistic,Gompertz and Cox model were evaluated. Multi-variable Results indicated the general preference of parametric models over Cox semi-parametric model. Among AFT models, the log-normal AFT model had a worse condition than others did. From log-logistic AFT, Exponetial AFT and Weibull AFT models with close values, the weibull AFT model had the lowest AIC quantity.Thus, weibull AFT model was the best fit model over the data of pediatrics survival time in this research containing significant predictors of mortality CD4 counts, WHO clincal stage, functional stustus, TB co-infected and other opportunistic infections. Various Studies Have Been Conducted to Compare Cox Model with Parametric models, most of which reporting that parametric models are more

efficient. For instance, (Sayehmiri, 2008; Nardi, and Schemper, 2003) have compared Cox model and alternate parametric models in two clinical studies, showing that Weibull model is superior to others based on the changes in the estimated parameter which is similar to this study in multivariate analysis. While in univariate analysis generalized gamma model had the lowest AIC value, but in this study log-normal model is superior. One limitation of the present study is its high percentage of right-censored observations. To reach an appropriate fit for parametric models, it is better not to have right-censored observations more than 40 to 50 percent (Nardi and Schemper, 2003). Although right-censored observations in this study is about 86 percent, parametric models resulted in a better model fitness (Nasir, 2013).

5.2 Conclusion

This study is based on HIV/AIDS pediatrrics under the era of HAART from Jimma University Specialzied Hospital from 1st september 2005 to 1st spetember 2013 were used. A finding of the present study revealed that after initiation of the treatment, HIV-positive pediatrics lived on average 40.14 months (3.34 years) with median survival time estimated to be nearly 38 months (3.17 years). We used Cox-PH and accelerated faliure time models to modelling the survival time of HIV/AIDS pediatrics under the era of HAART. Here, applying Cox–PH model is not appropriate becouse the PH assumption does not hold for some covariates. To overcome this, time-variying covariates are incorporated into the Cox model. We also used four different AFT models to fit the data. We found that the weibull AFT model fit better for this data set.

In our study, CD4 counts, WHO clinical stage, functional status, TB co-infected and other opportunistic infections are significantly associated with mortality among HIV/AIDS pediatrics affter HAART. Peditrics who have CD4 counts morethan $200mm^3$,WHO clinical stage I and II, no TB co- infected and no opportunistic infections are accelerate survival time. While Peditrics who have CD4 counts lessthan $200mm^3$,WHO clinical stage III and IV, TB co- infected and opportunistic infections are decelerate survival time.

The Cox PH model is routinely applied to the analysis of survival data. The study considered here provides an example of a situation where Cox-PH model is inappropriate and where the AFT model provides a better description of the data. We have seen that the AFT model is a more valuable and realistic alternative to the PH model in some situations. Furthermore, the AFT model makes it possible for clinicians to interpret the treatment benefit in terms of an effect on expected duration of illness. To this content the AFT model may have explanatory advantage in that covariates have a direct effect on survival times rather on hazard functions as in the PH model.

5.3 Recommendation

Being HIV infection is the most serious disease in the world, modelling the survival time of this disease helps to identify the factors that affect the success of the therapy. Thus further studies should be done in the area using these newly developed and most flexible methodologies by including additional covariates like (anemia, adherence, , nutritional status, and viral load) that may affect the mortality rate of HIV-positive patients receiving ART service.

Despite most researchers' interest in the application of Cox model for the survival analysis, parametric models have the ability to present better results than Cox model in cases where there is fairly less censored observations, whether hazard proportion hypothesis exists or not. Therefore, it is suggested that, while considering the results of all different models for survival analysis, the best and most efficient model be selected. The choice of the appropriate model will certainly lead to identify more reliable and precise prognostic factors and thus help to have a more effective treatment program.

REFERENACE

Abrams E, El-Sadr W, Rabkin M. (2011).*The Pediatric Clinical Manual; the International Center for AIDS Programs*. Available at: http://www.medbox.org/preview/5255d0fa-3e6c-41db-af3f-7b560e695ecc/doc.pdf.

Alemu, A. W., & San Sebastian, M. (2010). Determinants of survival in adult HIV patients on antiretroviral therapy in Oromiyaa, Ethiopia. *Global health action*, *3*.

Altman, D., B. De Stavola, S. Love and K. Stepniewska, (1995). Review of survival analyses Scientific Research, British Journal of Cancer, 72(2): 511.

Assefa, T., & Wencheko, E. (2012). Survival analysis of patients under chronic HIV-care and antiretroviral treatment at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia. *Ethiopian Journal of Health Development*, *26*(1), 22-29.

Assefa, Y., Jerene, D., Lulseged, S., Ooms, G., & Van Damme, W. (2009). Rapid scale-up of antiretroviral treatment in Ethiopia: successes and system-wide effects. *PLoS medicine*, *6*(4), e1000056.

Atnafu, H. (2011). *GRADUATE STUDIES PROGRAMME FACULTY OF SCIENCE STATISTICS DEPARTMENT* (Doctoral dissertation, Addis Ababa University).

Barlow, W. E., and Prentice, R. L, (1988). Residuals for relative risk regression. *Biometrika* 75, 65-74.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis Part III: multivariate data analysis–choosing a model and assessing its adequacy and fit. *British Journal of Cancer*, *89*(4), 605.

Brady, M. T., Oleske, J. M., Williams, P. L., Elgie, C., Mofenson, L. M., Dankner, W. M., & Van Dyke, R. B. (2010). Declines in mortality rates and changes in causes of death in HIV-1-infected children during the HAART era. *Journal of acquired immune deficiency syndromes (1999)*, *53*(1), 86.

Buckley, I. V., and James, I. (1979). Linear regression with censored data. *Biometrika*, 429-43

Cain, K. C., & Lange, N. T. (1984). Approximate case influence for the proportional hazards regression model with censored data. *Biometrics*, 493-499.

Collett, D. (2003). *Modelling Survical Data in Medical Research* (Vol. 57). CRC press.

Conge C, Tsoikos CP. (2010). Statistical modeling of breast cancer relapse time with different treatments. *Journal of applied sciences*; 10:37-44. 55.

Cook, G. (2007). *Tropical medicine: an illustrated history of the pioneers*. Academic Press. Cox, D. R. Partial likelihood, (1975). *Biometrika* 62 269-276.

Cox, D. D. R., & Oakes, D. (1984). *Analysis of survival data* (Vol. 21). CRC Press.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 187-220.

Crowder, J., and Hu, M, (1977). Covariance analysis of heart transplant survival data. *Journal of American Statistical Association* 78 27.36.

DE BAETS, A. J., BULTERYS, M., ABRAMS, E. J., KANKASSA, C. & PAZVAKAVAMBWA, I. E., (2007). Care and treatment of HIV-infected children in Africa: issues and challenges at the district hospital level. *Pediatr Infect Dis J*, 26, 163-73.

Dunn, D., Woodburn, P., Duong, T., Peto, J., Phillips, A., Gibb, D., & Porter, K. (2008). Current CD4 cell count and the short-term risk of AIDS and death before the availability of effective antiretroviral therapy in HIV-infected children and adults. *Journal of Infectious Diseases*, *197*(3), 398-404.

Eley B, Nuttall J, Davies MA, Smith L, Cowburn C, H,et al: (2004). Initial experience of a public-sector antiretroviral treatment programme for HIV-infected children and their infected parents.*S Afr Med J*,94(8):643–646.

Ersoy B, Balkanb C, Gunayc T, et al. (2003). Effects of different socioeconomic conditions on menarche in Turkish female students. 76: 115-125.

Ferradini, L., Jeannin, A., Pinoges, L., Izopet, J., Odhiambo, D., Mankhambo, L., ... & Brasher, C. (2006). Scaling up of highly active antiretroviral therapy in a rural district of Malawi: an effectiveness assessment. *The Lancet*, *367*(9519), 1335-1342.

Gebremedhin, A., Haile, F., & 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*(1), 1047.

Getie M, Modelling Time-to-Menarche.(2011) . Cox Proportional Hazard versus Accelerated Failure Time Models.

Glucksberg H, Storb R, Fefer A, Buckner CD, Neiman PE, Clift RA,Lerner KG.(1974). Clinical manifestations of graft-versushost disease in human recipients of marrow from HL-Amatched sibling donors.*Transplantation*, 18:295-304.

Hosmer Jr, D. W., & Lemeshow, S. (1999). Applied Survival Analysis: Regression Modelling of Time to Event Data (1999).

Hussey GD, Reijnhart RM, Sebens AM, Burgess J, Schaaf S, Potgieter S ,(1998. 'Survival of children in Cape Town known to be vertically infected with HIV-1', *S Afr Med J* May; 88(5):554-8.

Jerene, Are Næss and BerntLindtjørn. (2006a). Antiretroviral therapy at a district hospital in Ethiopia prevents death and tuberculosis in a cohort of HIV patients. *AIDS Research and Therapy;* 3(10).

Jerene, D., Endale, A., & Lindtjørn, B. (2007). Acceptability of HIV counselling and testing among tuberculosis patients in south Ethiopia. *BMC international health and human rights*, *7*(1), 4.

Kabue, M. M., Kekitiinwa, A., Maganda, A., Risser, J. M., Chan, W., & Kline, M. W. (2008). Growth in HIV-infected children receiving antiretroviral therapy at a pediatric infectious diseases clinic in Uganda. *AIDS patient care and STDs*, *22*(3), 245-251.

Kalbfleisch, J. D., and Prentice, R. L, (1973). Marginal likelihoods based on Coexist regression and life model. Biometrika 60 267-278.

Kleinbaum, D. G. (1998). Survival Analysis, a Self-Learning Text. *Biometrical Journal*, *40*(1), 107-108.

Koye, D. N., Ayele, T. A., & Zeleke, B. M. (2012). Predictors of mortality among children on Antiretroviral Therapy at a referral hospital, Northwest Ethiopia: A retrospective follow up study. *BMC pediatrics*, *12*(1), 161.

Lutalo, T., Nalugoda, F., Sewankambo, N., Kiduggavu, M., Wawer,M. & Gray, R. (2006). Mortality in HIV-infected and uninfected children of HIV-infected and uninfected mothers in rural Uganda. *J Acquire Immune DeficSyndr*, 41, 5048.

Moran JL, Bersten AD, Solomon PJ, Edibam C, Hunt T. (2008). Modelling survival in acute severe illness: Cox versus accelerated failure time models 1. *J EvalClin Pract*, 14:83-93.

Nardi, A. and M. Schemper, (2003). Comparing Cox and parametric models in clinical studies. Statistics in medicine, 22(23): 3597-610.

Nasir, S. (2013). Microfinance in India: Contemporary Issues and Challenges. *Middle East Journal of Scientific Research*, *15*(2).

Office of AIDS Research Advisory Council (OARAC) .(2009).Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents.

Orbe, J., Ferreira, E., & Núñez-Antón, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in medicine*, *21*(22), 3493-3510.

Ponnuraja C, Venkatesan P. (2010). Survival models for exploring tuberculosis clinical trial data an empirical comparison. *Indian Journal of Science and Technology*; 3: 755-58. 52.

Pourhoseingholi MA, Pourhoseingholi A, and Vahedi M, et al..(2011). Alternative for the Cox regression model: using parametric models to analyze the Survival of Cancer Patients. *Iran J Cancer Prev*. 2011; 4:1- 9. 30.

Prendergast, A., Mphatswe, W., Tudor-Williams, G., Rakgotho, Rabaud C, Charreau I, Izard S, Raffi F, Meiffredy V, Leport C, Aboulker JP.(2000). Adverse reactions to Cotrimoxazole in HIV infected patients: Predictive factors and subsequent disease progression. Antimicrobial Agents and Chemotherapy. 40(6):33-35.

Sayehmiri, K., Eshraghian, M. R., Mohammad, K., Alimoghaddam, K., Foroushani, A. R., Zeraati, H. ... & Ghavamzadeh, A. (2008). Journal of Experimental & Clinical Cancer Research. *Journal of Experimental & Clinical Cancer Research*, *27*, 74.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model Biometrika 69, 239-241.

Subbaraman, R., Chaguturu, S. K., Mayer, K. H., Flanigan, T. P. & Kumarasamy, N. (2007). Adverse effects of highly active antiretroviral therapy in developing countries. *Clin Infect Dis*, 45, 1093-101.

Swindle WR,(2009). Accelerated failure time models provide a useful statistical framework for aging research. *Experimental Gerontology*; 44: 190–200. 54.

Tindyebwa, d., Kayita, J., Musoke, P., Eley, B., Nduati, R., Coovadia, H., Bobart, R., Mbori-ngacha, D. & KIEFFER, M. P.(2006). Handbook on Pediatric AIDS in Africa. Uganda, African Network.

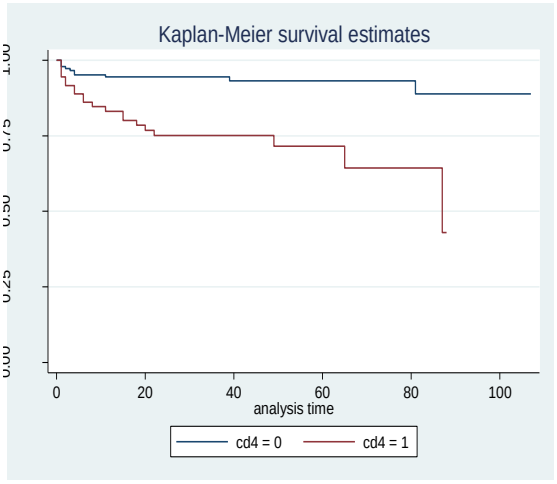
Tuner BJ, MarksonLE, Mckee L, et al. (1991). The aids-defining diagnosis and subsequent complications: a survival-based severity index. Journal of Acquired Immune Deficiency Syndromes 4 1059-1071.

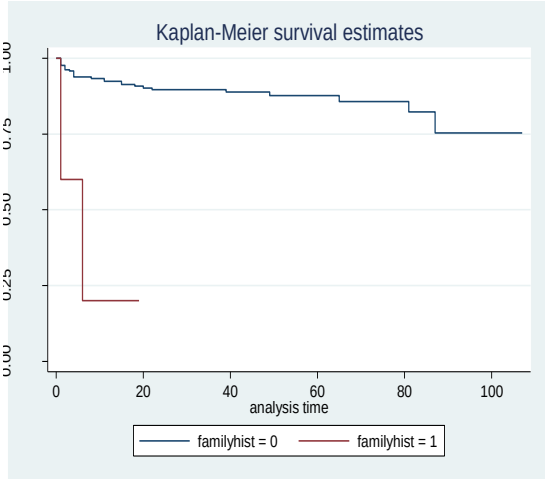
Violari, A., Cotton, M. F., Gibb, D. M., Babiker, A. G., Steyn, J., Madhi, S. A., ... & McIntyre, J. A. (2008). Early antiretroviral therapy and mortality among HIV-infected infants. *New England Journal of Medicine*, *359*(21), 2233-2244.

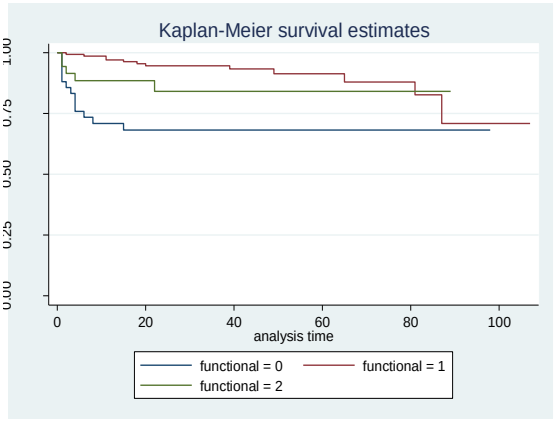
WHO/UNAIDS/UNICEF.(2011).*Towards universal access and Scaling up priority HIV/AIDS interventions in the health sector, Progress Report.* Available at: http://www.unicef.org/eapro/Towards_Universal_Access_on_HIVAIDS.pdf

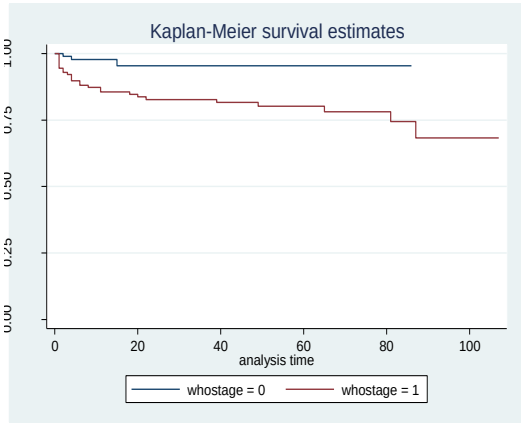
APPENDIX

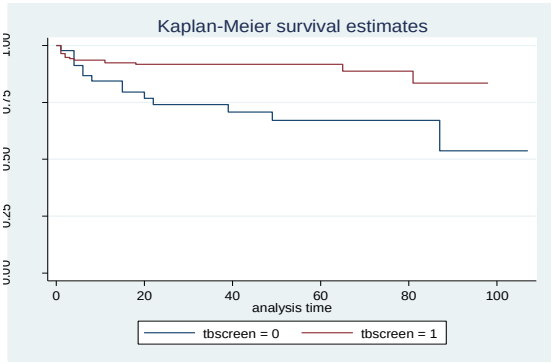
APPENDIX A: THE KAPLAN-MEIER SURVIVAL FUNCTION ESTIMATES FOR TIME TO DEATH

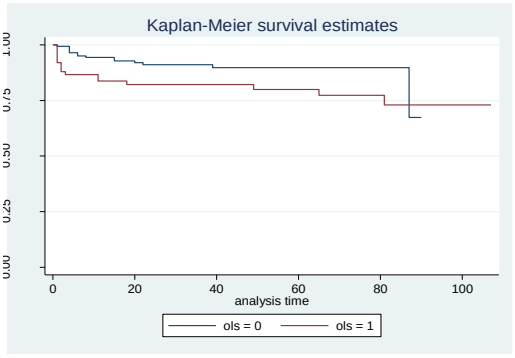






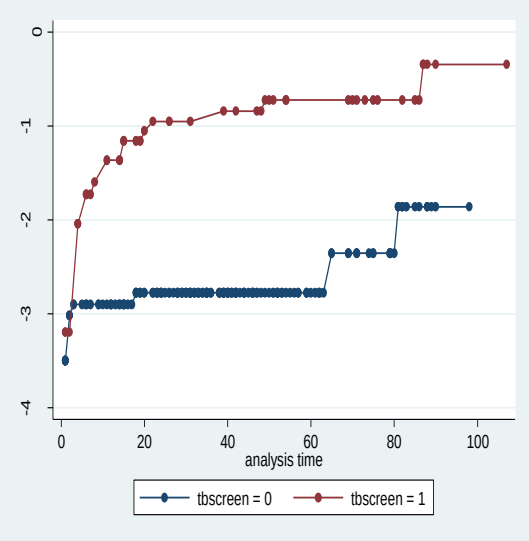


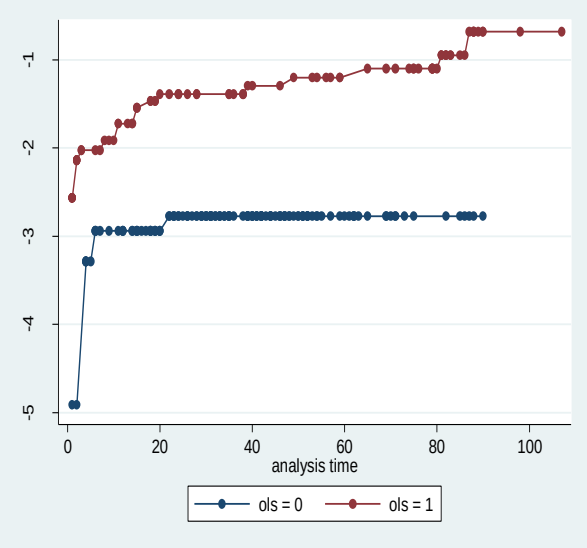


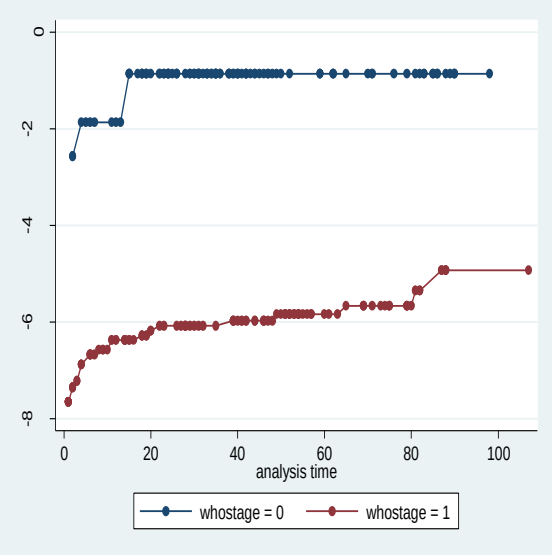


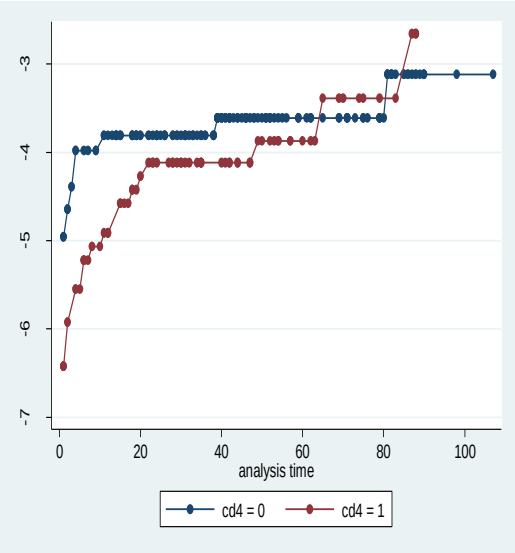
Survival function estimate of time to death based CD4 counts, family history of HIV/AIDS, functional status, WHO stages, TB screen and opportunistic infections.

APPENDIX B: PLOT OF LOG (-LOG (SURVIVAL) VERSUS SURVIVAL TIME FOR CATEGORICAL VARIABLES





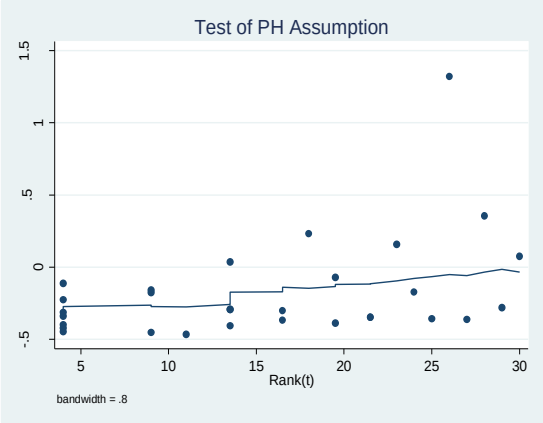


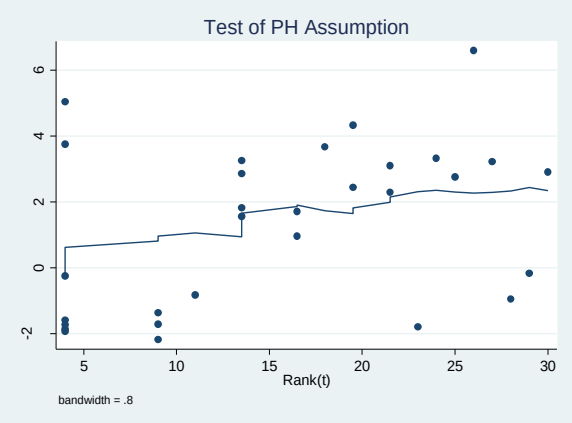


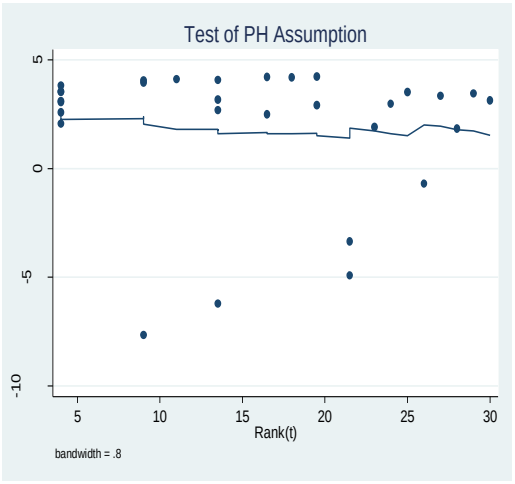
Plot of log (-log (survival)) versus follow up time to assess the proportionality assumption for

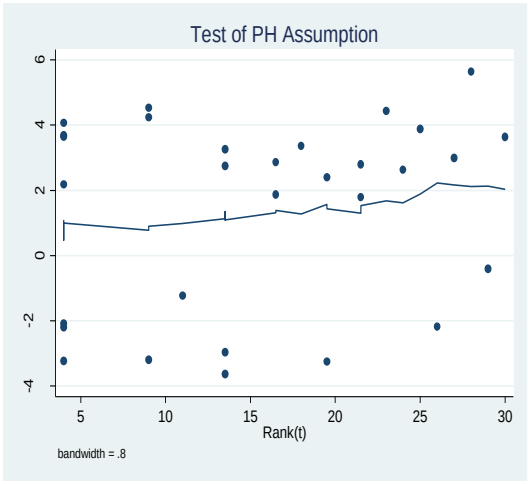
TB screen, opportunistic functions, WHO stages and CD4 counts.

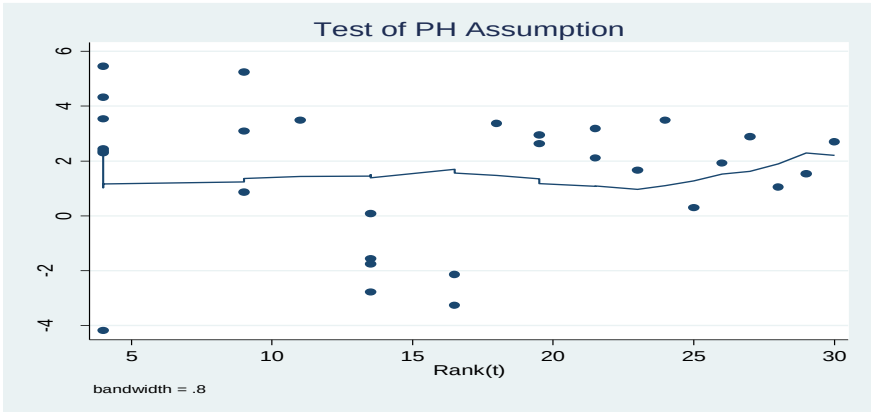
APPENDIX C: PLOT OF SCHONEFELD RESIDUALS VERSUS RANKED SURVIVAL TIME











Plot of schonefeld residuals versus ranked survival time to assess the proportionality assumption for TB screen, opportunistic functions, WHO stages and CD4 counts.

APPENDIX D: PLOT OF $\log[-\log S(t)]$ VERSUS $\log t$ FOR WEIBULL AFT