ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

PARAMETRIC MODELING OF SURVIVAL DATA BASED ON HIV INFECTED ADULT PATIENTS UNDER HAART: A CASE OF ZEWDITU REFERRAL HOSPITAL, ADDIS ABABA

HAFTU LEGESSE

A THESIS SUBMITTED TO

DEPARTMENT OF STATISTICS

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS (APPLIED STATISTICS)

ADDIS ABABA UNIVERSITY

ADDIS ABABA, ETHIOPIA

JUNE, 2015

Addis Ababa University

Office of Graduate Program

College of Natural Science

Department of Statistics

This is to certify that the thesis prepared by Haftu Legesse Gebreyesus, entitled: "Parametric modeling of survival data based on HIV infected adult patients under HAART: A case of Zewditu Referral Hospital, Addis Ababa" and submitted in partial fulfillment of the requirements for the degree of master of science in Statistics (Applied Statistics) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Advisor Prof. M.K Sharma Signature_____Date _____

Examiner Prof. Eshetu Wencheko Signature _____Date _____

Examiner Ato Mekonnen Tadesse Signature_____Date _____

_____

Chair of Department or Graduate Program Coordinator

Abstract

Parametric modeling of survival data based on HIV infected adult patients under HAART: A case of Zewditu Referral Hospital, Addis Ababa.

Haftu Legesse

Addis Ababa University, 2015

Acquired Immunodeficiency Syndrome (AIDS) is one of the most destructive epidemics in the history of mankind. HIV/AIDS has now become a worldwide issue in general and developing countries in particular. The study was aimed to compare the performance of the common parametric models namely; Exponential, Weibull, Gompertz, and Log-logistic using HIV infected adult patients' dataset. A second objective of the study was to determine the factors/variables that affect the survival time of HIV infected patients. A retrospective cohort study was conducted in Zewditu Referral Hospital located in Addis Ababa, Ethiopia. Records of patients enrolled between September 2010 and August 2014 were reviewed continuously using patients' ART unique identification numbers as reference. Kaplan-Meier survival curves and Log-Rank test were used to compare the survival experience of different category of patients and parametric survival models were employed to predict survival time of the patients. All fitted models were compared by using AIC and log likelihood. Of all 638 HIV infected adult patients, 64(10%) died during the follow up period. The log-logistic model gave a better description of the time-to-death of HIV infected adult patients than the other models considered. Based on log-logistic model, age, weight, functional status, TB screen, WHO clinical stage and educational level were found to be the most prognostic factors of time-to-death. Furthermore a high risk of death of patients was found to be associated with lower initial weight, WHO clinical stage IV, lower CD4 count, being ambulatory, bedridden, and TB screened and illiterate.

Acknowledgement

My heartfelt gratitude goes to Prof. M.K Sharma, my advisor and instructor, for his excellent scientific guidance and tireless efforts to make this work a reality.

Next, I want to thank Ato Getahun Worku, who have provided me help and moral support in one way or another to accomplish this study.

Finally, I would like to thank the staff members of Zewditu Referral Hospital for their constant insight, encouragement, and in providing necessary materials, references in implementing this research.

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

Acronyms

AA	Addis Ababa
AFT	Accelerated Failure Time
AIC	Akakie Information Criteria
AIDS	Acquired Immune Deficiency Syndrome
APHOD	Asia Pacific HIV Observational Database
ART	Antiretroviral Therapy
DF	Degrees of Freedom
EDHS	Ethiopian Demographic Health Survey
EFMOH	Ethiopian Federal Ministry of Health
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immune Virus
HR	Hazard Ratio
KM	Kaplan-Meier
NNRTI	Non-nucleoside Reverse Transcriptase Inhibitors
NRTI	Nucleoside Reverse Transcriptase Inhibitors
OARAC	Office of AIDS Research Advisory Council
OIs	Opportunistic Infections
PHM	Proportional Hazard Model
PI	Protease Inhibitors
PO	Proportional Odds
UARTCG	Ugandan Antiretroviral Treatment and Care Guidelines
UNAIDS	Joint United Nations Program on HIV/AIDS
TB	Tuberculosis
WHO	World Health Organization

Tables of Contents

CONTENTS **Pages**

List of Tables

List of Figures

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the Study

The rate of spread of the HIV/AIDS and the damages accompanying it has reached a level which shock economists, health workers, politicians etc. HIV/AIDS has now become a worldwide issue in general and developing countries in particular. The disease being one without any cure is still accountable for economic, social and health crises especially in developing countries. Its high prevalence and/or distribution among the youth made things even more complicated.

Human immunodeficiency Virus (HIV), the agent that causes acquired immune deficiency syndrome (AIDS), is classified as members of the lent virus subfamily of retroviruses. There are two main types of HIV: HIV type 1(HIV-1): the most prevalent throughout the world. HIV type 2 (HIV-2) is prevalent in West Africa. They both cause ADIS and the routes of transmission are the same. However, HIV-2 causes AIDS much more slowly than HIV-1(Seoane and Resino, 2008).

Acquired Immunodeficiency Syndrome (AIDS) is one of the most destructive epidemics in the history of mankind. Since its detection in 1981, HIV/AIDS has become one of the most challenging problems of our age. According to the UNAIDS 2007 report, an estimated 33.2 million people were living with HIV world wide and 2.5 million became newly infected with the virus while 2.1 million lost their lives to AIDS (UNAIDS, 2007).

The number of people living with HIV continues to rise - in 2006 there were an estimated 32.7 million living with HIV/AIDS and in 2007 the number rose to 33.2 million. Globally, this means that every day, 6,800 people have been infected and 5,700 die from HIV/AIDS. AIDS remains the single largest cause of death in Africa and the worst public health crisis worldwide (UNAIDS,2007).

Africa is the region most affected by the spread of HIV/AIDS and within Africa, Sub-Saharan Africa has remained to be the most devastated by the epidemic. In 2007, Sub-Saharan Africa accounted for more than two thirds (68%) of all persons infected with HIV, and 72% of global AIDS deaths (UNAIDS, 2007).

HIV/AIDS affects society and economies at various levels, from the family and community to the national and international levels - particularly by eroding the human capital. It is for example noted that particularly in Sub-Saharan Africa, HIV/AIDS continues to slow or even reverse improvements in life expectancy and distort the age-sex structure of the entire population. The 2005 Human Development Report identifies AIDS as the factor inflicting the single greatest reversal in human development history. Between 1990 and 2003 many of the country‟s most severely affected by AIDS dropped sharply in the global ranking of countries on the human development index (UNAIDS, 2006).

In Ethiopia the adult prevalence of HIV was estimated to be 1.5% in 2011. The total number of People Living with HIV/AIDS (PLHIV) in the same period was estimated to be 1,037,267 adults and 68,136 of them were children. Furthermore the number of deaths due to AIDS for the same period was estimated to be 58,290 for adults and 9,284 among children (UNAIDS, 2007).

The goals of treatment with antiretroviral drugs are to inhibit viral replication while minimizing toxicities and side effects associated with the available drugs. The inhibition of virus replication permits restoration of the immune system (suppression of HIV replication, as reflected in plasma HIV concentration, to as low as possible and for as long as possible, the preservation or enhancement of the immune function (CD4 restoration), thereby preventing or delaying the clinical progression of HIV disease. Viral eradication from the host genome is not achievable, thus a cure for HIV is not yet possible. By using HAART, it is possible to promote growth in children and prolong the survival of all HIV infected patients, reduce their morbidity and improve their quality of life (UARTCG, 2008).

In just 25 years, HIV has spread relentlessly from a few widely scattered "hot spots" to virtually every country in the world, infecting 65 million people and killing 25 million. An estimated 38.6 million [33.4 million–46.0 million] people worldwide were living with HIV in 2005. An estimated 4.1 million [3.4 million–6.2 million] became newly infected with HIV and an estimated 2.8 million [2.4 million–3.3 million] lost their lives due to AIDS. In sub-Saharan Africa, the region with the largest burden of the AIDS epidemic, data also indicate that the HIV incidence rate has peaked in most countries. However, the epidemics in this region are highly diverse and especially severe in southern Africa, where the epidemic is still expanding. New survey data underscore the disproportionate impact of the AIDS epidemic on women, especially in sub-Saharan Africa where, on average, three women are HIV-infected for every two men. Among young people (15–24 years), that ratio widens considerably, to three young women for every young man (UNAIDS, 2006).

Knowledge of HIV status helps HIV negative individuals make specific decisions to reduce risk and increase safer sex practices so that they can remain free of disease. For those who are infected with HIV, knowledge of their status allows them to take action to protect their sexual partners, to seek treatment, and to plan for the future. The predominant mode of HIV transmission is through sexual contact. Other modes of transmission are mother-to-child transmission (in which the mother passes HIV to her child during pregnancy, delivery, or breastfeeding), use of contaminated blood supplies for transfusions, and injections using contaminated needles or syringes (EDHS, 2011).

The majority of Ethiopian adults (63 percent of women and 78 percent of men) know that a healthy-looking person can have HIV. The most common misconception about HIV transmission is that it can be transmitted by mosquitoes. Only about half of women (52 percent) and six men of every ten (63 percent) know that HIV cannot be transmitted by mosquitoes. The second most common misconception is that HIV can be transmitted by supernatural means. About three-fourths of women and men age 15-49 (72 and 76 percent, respectively) correctly believe that HIV cannot be transmitted through supernatural means. Seventy-six percent of women and 85 percent of men know that a person cannot become infected with HIV by sharing food with a person who has HIV.

About three-quarters of women and more than four men of every five report that people can reduce their chance of getting HIV by abstaining from sexual intercourse (73 and 86 percent, respectively). More than nine of every ten women and men know that people can get HIV by sharing sharp materials, such as razors or other blades, if they should cut the skin or by injection with unsterilized needles (92 and 97 percent respectively) (EDHS , 2011).

Several cohort studies and clinical trials have shown that the CD4 count is the strongest predictor of subsequent disease progression and survival (Mellors et al, 1997).The use of the CD4 count as an independent and reliable marker for treatment outcome is attractive from various aspects. First, CD4 counts are already the most important factor in deciding whether to initiate antiretroviral therapy and opportunistic prophylaxis – all HIV-positive patients in high-income countries, and an increasing number of patients in low-income countries have a baseline CD4 count at entry into care (Panel ART Guidelines and Adolescents, 2008). Second, the CD4 count is a relatively objective and simple marker to follow. Finally, the cost of CD4 counts has become more affordable, including in developing countries (MacLennan et al, 2007).

An adequate CD4 response for most patients on therapy is defined as an increase in the range of 50–150 cells/mm3 per year with an accelerated response in the first 3 months of treatment (Panel ART Guidelines and Adolescents, 2008). In general, CD4 counts should be checked every 3–4 months to determine when to start anti-retroviral therapy, to assess immunologic response to therapy and to evaluate the need for initiation or discontinuation of prophylaxis for opportunistic infections (OIs). Patients with good virologic control average approximately 50–100 cells/mm3 per year until a steady-state level is reached (Panel ART Guidelines and Adolescents, 2008). A clinically significant change between CD4 counts approximates a 30% change in the absolute count or an increase or decrease in CD4 percentage by 3% (Panel ART Guidelines and Adolescents, 2008). For those patients who adhere to therapy with sustained viral suppression and are clinically stable for more than 2–3 years, the frequency of CD4 count monitoring may be extended to every 6 months. In cases of discordant CD4 and viral-load results, the clinician should first exclude a laboratory error and consider retesting the patient.

This study further evaluates the use of the CD4 count in assessing the clinical status of HIV-infected individuals, in making informed decisions regarding the initiation of antiretroviral therapy and in monitoring the success of such therapy by using statistical methodology.

1.2 Statement of the Problem

The basic research questions this study attempted to answer are:

1. Which parametric survival model is appropriate to model the HIV infected adult patient‟s treated with HAART dataset?
2. What are the determinant factors and/or covariates that affect the survival of HIV infected adult patients treated with HAART?

1.2 Objective of the Study

1.2.1 General Objective of the Study

The general objective of this research is to compare the performance of four parametric models and identify the prognostic factors for time-to-death of HIV infected adult patients treated on HAART using Zewditu Referral Hospital as case study area.

1.2.2 Specific Objectives

- ✓ To see the change of CD4 counts over time after starting HAART.
- ✓ To predict the survival time of HIV infected adult patients treated with HAART.
- ✓ To demonstrate the application of parametric models to the dataset.

1.3 Significance of the Study

The study will have the following significance

- It may be used for patients‟ management and to predict the disease progression and/or treatment outcomes.
- It is hoped that this research will be useful because it is assumed to serve as an input for policy makers and concerned health specialists who work on providing care, support and treatment aspect of the HIV/AIDS programs of the country.

1.3 Limitation of the Study

- The study was conducted based on secondary data which might have incomplete and biased information.
- The study presumed that all deaths are caused by HIV/AIDS.
- The study was merely based on adults due to insignificant number of infants and children so that the results do not cover this portion of population.
- High percentage of censored observation which might be due to lost to follow up.
- Moreover, the study is based on baseline values of the variables of interest such as weight, WHO clinical stage, functional status.
- Information might have been missed in case of many censored observations.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Theoretical Literature

2.1.1 Global Epidemiology of HIV/AIDS

In 2008, an estimated 2.7 million HIV infections occurred worldwide; this was 30% lower than the 3.5 million new infections at the peak of the epidemic in 1996 (UNAIDS, 2009). Sub-Saharan Africa remains the most heavily affected region, accounting for about 71% of all new HIV infections in 2008. There are two related but distinct types of HIV: HIV-1 and HIV-2 (Kakuda et al., 2002). HIV-1 is the most pathogenic and causes over 99 % of HIV infections. HIV-2 is also known to cause AIDS but is much less prevalent, being present in fewer and isolated geographic locations such as West Africa. Therefore most research is done on HIV-1 (Klos et al., 2009). AIDS related disease remains one of the leading causes of death globally. According to UNAIDS, the number of people living with HIV/AIDS worldwide was estimated at 33.4 million in 2008; more than 20% higher than the number in 2002. It was estimated that 2 million deaths due to AIDS-related illness occurred worldwide in 2008; this was ~10% lower than in 2004. The declines in new infections and AIDS-deaths may be attributed to the scaled-up of ART programmes, especially in the developing world. As of December 2008, approximately 4 million people in low-and middle-income countries were on ART, representing a 10-fold increase over five years. In eastern and Southern Africa, ART coverage rose from7% in 2003 to 48% in 2008 (UNAIDS, 2009).

2.1.2 Epidemiology of HIV/AIDS in Ethiopia

The HIV pandemic created unprecedented burden on the economies and health care systems of affected countries, particularly in sub-Saharan Africa, where prevalence is highest. In Ethiopia, the total number of people who have died due to HIV/AIDS in 2006 alone was 88,997; and in 2007, it is estimated that, 71,902 people will die.

In 2007, an estimated 898,350 children have lost one or both parents to the epidemic (AIDS orphans). According to the calibrated single point estimate (from 2005 sentinel

surveillance and EDHS data), prevalence of adult infection is 2.1% (urban 7.7%, rural 0.9%). In 2007, the estimated number of people living with HIV is 977,394, including 64,813 children. The current estimates of people requiring ART is 258,264 and of these 6 % (UNAIDS, 2009) are children.

2.1.3 Recovery of CD4 cells after Initiation of ART

ART usually results in a biphasic increase in CD4 cell count. The initial increase in CD4 cell count is very rapid and is usually observed in the first 3-6 months and a second phase of slower increase follows. More than 95% of successfully treated individuals with well – controlled HIV-1 viraemia reach a CD4 cell count of more than 200 cells per μL. However, one-third of successfully treated patients appear not to reach normal CD4 cell count within five years. This observation raises concerns for the long term prognosis of patients and suggests that ART should be initiated before CD4 cell counts fall below a certain threshold. With adequate monitoring, CD4 cell counts can be maintained above pre-specified levels during scheduled treatment interruption. However, a lower CD4 cell count may expose patients to an increased risk of clinical events. This risk needs to be carefully balanced against the potential benefits associated with decreased exposure to antiretroviral drugs (Battegay et al., 2006).

2.2 Empirical Literature

In one study in Ethiopia, a total of 887 HIV positive patients were involved; Out of these 472 (53.2%) were female and 415 (46.8%) male patients. None of them have any opportunistic infection during the time of follow up. The mean age of the study group was 36.76 (17-76). The mean baseline CD4+ count was 81.40; the mean CD4 count at the 6-th, 9-th and 12-th month was 191.65, 284 and 331 respectively. There was a good immune recovery at the 6-th month of therapy from the baseline mean CD4+T cell count of 81 cells/mm3 to 191.65 cells/mm3, which has statistically significant ($p<0.0001$). Most of the HIV infected patients enrolled in the study were young age between 20 and 40 years old who were sexually more active and thus have a higher risk of infection compared to the other age groups (Derbe et al., 2013).

Another study in Ethiopia by Seid et al (2014) found that gender, age, clinical stage, functional status and education level to be significantly associated with defaulting. In

addition, the results show that the patient‟s survival in the HAART treatment is associated with patient-specific CD4 fluctuations such that a patient with higher CD4 trend is less likely to default from the treatment. An individual with higher CD4 variability is more likely to default than an individual with smaller CD4 variability. .

Numerous studies have demonstrated that the baseline CD4 count serves as a significant prognostic indicator for treatment outcome. In one study, patients starting therapy with a CD4 count below 200 cells/mm3 were almost twice as likely (HR: 1.90) to fail treatment, compared with those starting with a CD4 count higher than 200 cells/mm3 (Robbins et al., 2007). Another study showed an inverse relationship between the CD4 count at baseline and a risk of progression to AIDS or death (Egger and Chene, 2002). This effect was quite dramatic: the adjusted HR for progression to AIDS or death was 0.24 (95% CI: 0.20–0.30) for patients starting HAART with a baseline CD4 count of 200–350 cells/mm^3, compared with patients having CD4 count below 50 cells/mm^3 Recent data support the prognostic value at higher CD4 cell-count levels. In a large cohort study, patients initiating HAART with CD4 counts of 350–500 cells /mm^3 had a 94% increased risk of death, relative to those with baseline CD4 counts above 500 cells/mm^3. An increased risk of death when HAART was deferred until the CD4 count fell below 350 cells/mm^3 (Kitahata et al., 2009).

A study was undertaken about 1,638 patients who started ART (defined as three or more antiretroviral drugs) after January 1, 1997, and who had a baseline CD4 cell count and viral load measurement within six months before and up to one month after starting ART. Changes in CD4 cell count response after starting ART were studied for up to six years of follow-up time. The result showed that the long-term CD4 cell count response was determined by a relationship between baseline CD4 cell count, elevated viral load, and time. Greater increases in average CD4 cell counts were seen among patients who had complete or partial viral load suppression during the follow-up period (Sam et al., 2009).

Seage et al. (1997) based on data collected from Boston hospital through a survival analysis showed that functional status and recent opportunistic diseases as the major predictors of survival time.

The World Health Organization (WHO) reported in 1999 that of a total 53.9 million deaths, 1.5 million deaths was caused by TB. In addition, it was claimed as TB co-infection is the leading cause of mortality among those infected with HIV worldwide. A finding of a cross-sectional study based on 241 cases reported from nine domestic hospitals throughout mainland China was in agreement with the stated claim. The patients in the study were followed from January 2003 to December 2005. In spite of the fact that treatments for TB and HIV were provided to the patients, mortality attributable to co-infection was reported for 15.8% of the cases. As a result, the study concluded that HIV/TB co-infection was related to high mortality even when HAART and/or drug therapy for TB was provided (Xueyan et al., 2008).

A retrospective survival time study of 790 HIV-infected patients was conducted between 16 May 1985 and 31 December 2001 Singapore to determine independent predictors of HIV disease progression. The results showed that patients of younger age and higher baseline CD4 cell count associated with a lower risk of progression to AIDS (Chow et al., 2005).

A survival study was conducted in South Africa based on 18 published cohort studies on 39,536 HIV/AIDS patients to assess the proportion of survival time and random-effects model to find hazard ratio of prognostic variables (Lawn et al., 2008). The study suggested advanced WHO clinical stage and low CD4 cell count as indicators of high mortality. A similar study was done in Malawi based on 1,308 patients to assess survival and predictors of death. The study found low body-mass index, WHO clinical stage IV, male gender, and baseline CD4 count lower than50 cells/ml as determinants of death (Ferradini et al., 2006).

To determine the relationship between mortality risk and the CD4 cell response to ART, a cohort of 2,423 patients on ART and who had a median baseline CD4 105 cells/µl were observed for up to 5 years of ART in South Africa. Kaplan-Meier analyses were used to estimate cumulative mortality, univariate and multivariate mixed effect Poisson regression models were used to estimate the association between baseline risk factors, updated CD4 cell counts and viral load and the incidence of mortality. In addition Wilcoxon rank-sum test has been used to compare medians. Older age, WHO clinical

stage 4, updated CD4 cell count and detectable updated viral load measurements were significantly associated with mortality risk in both the crude analyses and the multivariate model with the exception of male sex which is significant in the crude analysis only. However, updated CD4 cell counts were the variable most strongly associated with death (Lawn et al, 2009).

A study conducted by Nakhaee and Law (2011), showed that, survival following a diagnosis of HIV infection was modeled by applying parametric survival models on people who were only diagnosed with HIV or HIV and AIDS registered in the Australian surveillance system from 1997 to 2003. Likelihood based criteria for model selection indicated that the Weibull model was the best fitting parametric model for predicting survival following both HIV and AIDS diagnoses.

Baghestani et al., (2010) indicated that the early detection of a cancer at a young patient age and in primary stages is important to increase survival from gastric cancer. According to statistical criteria, a parametric model can be a useful statistical model to find prognostic factors in the presence of interval censoring. Deviance supported the log-logistic model as the best option.

Jiezhi Qi (2009) found that, after comparison of parametric models and assessment of goodness of fit, the log-logistic accelerated failure time (AFT) model fits better for randomized placebo-controlled trial to prevent Tuberculosis in Uganda adults infected with HIV.

Ponnuraja and Venkatesan (2010), using applied likelihood-based criteria for model selection, showed that the Gamma model was the best fitting parametric model for tuberculosis clinical trial data. Hayat et al., (2010), showed that the Gompertz model was more suitable, for breast cancer registry data from ege university cancer research center.

CHAPTER THREE

3. DATA AND METHODOLOGY

3.1 Data and Source

The data for this study is a longitudinal cohort follows up retrospective cohort design of HIV infected adult patients data obtained from Zewditu Referral Hospital, Addis Ababa, Ethiopia. The ART clinic Zewditu Referral Hospital provides HIV/AIDS interventions including free diagnosis, treatment and monitoring. The center diagnoses new cases and monitors those on therapy. This study is based on a review of the patients‟ intake forms and follow-up cards of HIV patients on HAART. The patient‟s forms have been designed by FMOH for uniformity of use in the country so that those forms can be used to document almost all relevant clinical and laboratory variables.

We had taken all patients older than 15 years (i.e., both adolescents and adults) who received HAART in 2010. A total of 653 patients in the clinic who started HAART between September, 2010 and August, 2011 were included in the study. Patients were eligible for ART on the basis of the 2010 WHO guidelines (WHO clinical stage I/II disease with CD4 cell count below 350 cells/μL and WHO stage III/ IV disease with CD4 cell count above 350 cells/μL). The patients were followed up until August 2014. However, the study used data on included 638 HIV infected adult patients for whom data for variables of interest are complete.

3.2 Study Variables

3.2.1 The Dependent Variable

The response (dependent) variable is the survival time of HIV infected adult patients, the length of time from HAART start date until the date of death (or censor) measured in months.

HIV infected adult patients, who stayed alive during the study time, transferred to other hospitals, lost and dropped before death, are considered as censored. This means that the type of the survival data is random right censored.

3.2.2 Predictor (Independent) Variables

Explanatory variables which are assumed to influence the survival of HIV infected patients and are given below.

Table3. 1**:** Description of independent variables used in the analysis

Variables	Description	Values/Codes
Gender	Gender	(0) Female (1)Male
Age	Age in years	(0) =less than 30 (1)= [30-39] (2) = [40-49] (3)=50 and above
Weight	Baseline body weight in kg	Continuous variable
CD4 count	CD4 cells count	Continuous variable
Functional status	Functional status	(0) Working (1) Ambulatory (2) Bedridden
TB screen	TB screen	(0) No (1)Yes
OIs	Past Opportunistic infection	(0) No (1)Yes
WHO stage	WHO clinical stage	(0) WHO clinical stage I (1)WHO clinical stage II

		(2) WHO clinical stage III (3) WHO clinical stage IV
Educational level	Level of education	(0) No education (1)Primary (2) Secondary and above
Marital status	Marital status	(0) Never married (1) Married (2) Other
Religion	Religion	(0) Muslim (1) Coptic Orthodox (2) Other

Note: All predictor variables except CD4 count are taken as baseline values. CD4 count is recorded at every six month of visit time. In the analysis all variable categories which are coded by "0" are considered as reference categories.

3.3 Statistical Methodology

3.3.1 Survival Analysis

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. By time, we mean years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs. Survival analysis is an important statistical technique used to describe and model time-to-event data.

The use of survival analysis, as opposed to the use of other statistical methods, is most important when some subjects are lost to follow up or when the period of observation is finite and certain patients may not experience the event of interest over the study period.

In this latter case one cannot have complete information for such individuals. These incomplete observations are referred to as being censored.

Most survival analyses consider a key analytical problem of censoring. In essence, censoring occurs when we have some information about individual survival time, but we do not know the survival time exactly. There are generally three reasons why censoring may occur

a) a person does not experience the event before the study ends;

b) a person is lost to follow-up during the study period;

c) a person withdraws from the study because of death (if death is not the event of interest) or some other reason (e.g., adverse drug reaction or other competing risk)

There are three categories of censoring, (Klein, 1992),

i) Right censoring: Survival time is said to be right censored when it is recorded from its beginning to a defined time before its end time. This type of censoring is commonly recognized survival analysis and also considered in this study.

ii) Left censoring: Survival time is said to be left censored if an individual develops an event of interest prior to the beginning of the study; this is not common in survival studies.

iii) Interval censoring: Survival time is said to be interval censored when it is only known that the event of interest occurs within an interval of time but the exact time of its occurrence is not known.

3.4 Descriptive Methods for Survival Data

An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group. In summarizing survival data, the two common functions applied are the survivor function and the hazard function (Hosmer and Lemeshow, 1999).

3.4.1 Survival Function

The basic quantity employed to describe time-to-event phenomena is the survival function, the probability of an individual surviving or being event-free beyond time t (experiencing the event after time t). Moreover, the distribution of survival time is characterized by three functions: (a) the survivorship function, (b) the probability density function, and (c) the hazard function.

Let T be a random variable associated with the survival times, t be the realization of the random variable T and $f(t)$ be the underlying probability density function of the survival time t. The cumulative hazard function $\Lambda(t)$, which represents the probability that a subject selected at random will have a survival time less than some stated value t, is given by:

$$_T(t)=P_T(T\leq t)=\int_0^t \lambda(u)du \text{ , } t>0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

The survival function is defined as the probability that the survival time is greater or equal to t.

$$S(t) = P(T > t) \text{ , } t>0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

When T is a continuous random variable, the survival function is the complement of the cumulative distribution function, that is $S(t) = 1 - \quad (t)$ and density function is

$$f(t) = \frac{-dS(t)}{dt} \text{ , } t \geq 0.$$

Theoretically, as t ranges from 0 to infinity, the survivor function can be graphed as a smooth curve. Survivor functions have the characteristics that:

a) they are non-increasing
b) at time t = 0, S(t) = S(0) = 1; that is, at the start of the study, since no one has experienced the event yet, the probability of surviving past time 0 is one and
c) as time $t\rightarrow\infty$, $S(t)\rightarrow 0$; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually converge to zero.

3.4.2 Median Survival Time

In analysis of the survival data we use the median survival time than the mean because of the existence of censored and positively skewed nature of survival time. This is the time beyond which 50% of the individuals in the population under study are expected to survive and is given by that value $t(50)$which is such that $S\{\,\mathrm{t}(50)\}= 0.5$.

Median survival time $t_{0.5}$ is defined as that value for which $S(t_{0.5}) = 0.5$. If S(t) is not strictly decreasing, $t_{0.5}$ is the smallest number such that

$S(t_{0.5}) \leq 0.5$, or $t_{0.5}=S^{-1}(0.5)$ …….……..…… (3)

3.4.3 Hazard Function

The term has different meanings in different field of studies: it is known as the conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic processes, the age-specific failure rate in epidemiology and the inverse of the Mill"s ratio (the hazard rate) in economics.

The hazard function is a measure of the probability of failure during a very small interval, assuming that the individual has survived at the beginning of the interval. The hazard function describes the concept of the risk of an outcome (e.g., death, failure, hospitalization) in an interval after time t, conditional on the subject having survived to time t. It is the probability that an individual dies somewhere between t and $t + \Delta t$, divided by the probability that the individual survived beyond time t. The hazard function λ (t) can be formulated as:

$$\lambda(t) = \lim_{\Delta t\to 0} \frac{P(t\leq T\leq t+\Delta t/T\geq t)}{\Delta t}$$

$$= \lim_{\Delta t\to 0} \frac{1}{\Delta t}\frac{P([t\leq T\leq t+t\Delta]\cap[T\geq t])}{P(T\geq t)} = \lim_{\Delta t\to 0} \frac{P([t\leq T\leq t+t\Delta]\cap[T\geq t])}{\Delta t} \cdot \frac{1}{P(T\geq t)}$$

$$=\lim_{\Delta t\to 0} \frac{P([t\leq T\leq t+t\Delta]}{\Delta t} \cdot \frac{1}{P(T\geq t)}$$

$$=f(t) .\frac{1}{\mathrm{S(t)}} \qquad \text{…….....……..……..……(4)}$$

The survival and cumulative hazard functions can be given in terms of the hazard function as:

$$\Lambda(t) = \int_0^t \lambda(u)du \text{ and } S(t)=\exp\{-\Lambda(t)\}=\exp\{-\int_0^t \lambda(u)du\} \text{ respectively. .....(5)}$$

Using the above expressions the hazard function λ (t) can also be given as:

$$\lambda(t)=\frac{-dlogS(t)}{dt}=\frac{d\Lambda(t)}{dt} \quad \text{.....(6)}$$

3.5 Estimation of survivorship function

In survival analysis, it is always a good idea to present numerical or graphical summaries of the survival times for the individuals. In general, survival data are conveniently summarized through estimates of the survival function and hazard function. This method is non-parametric or distribution-free, since they require no specific assumptions to be made about the underlying distribution of the survival times (Hosmer and Lemeshow, 1999).

Among the other estimators of the survivor function the Kaplan-Meier estimator is the most common one. The Kaplan-Meier estimator of the survivorship function [Kaplan and Meier (1958)] also called product limit estimator, is the estimator used by most software packages. This estimator incorporates information from all of the observations available, both uncensored and censored, by considering survival to any point in time as a series of steps defined by the observed survival and censored times.

Suppose we have a sample of n independent observations, their survival times denoted by $t_1, t_2, \dots, t_n$ and indicators of censoring denoting by $\delta_1, \delta_2, \dots, \delta_n$ where

$$\delta_i = \begin{cases} 1, \text{if an event of death occur} \\ 0, \quad \text{Otherwise} \end{cases} \quad \text{.....(7)}$$

Thus, the survival data are denoted by (t_i,δ_i), $i = 1,2,\dots,n$. The first step to obtain the Kaplan-Meier estimator of the survival function is to order the survival times as $t_1, t_2, \dots, t_n$. Assume that among the n observations $m \leq n$ death occurred at distinct m times. The main quantity of interest is the probability that an event will not occur by time

t: $S(t)= P(T \geq t)$. Kaplan and Meier (1958) develop an estimator for the survival function.

$$\hat{S}(t)_{KM} = \prod_{t_{(i)} \leq t} (\frac{n_i - d_i}{n_i})^{\delta_i} = \prod_{t_{(i)} \leq t} (1 - \frac{d_i}{n_i})^{\delta_i}$$

Where

- d_i= number of patients died at $t_{(i)}$
- n_i= number of patients at risk before $t_{(i)}$

The variance of the Kaplan-Meier estimators which is referred to as Greenwood"s formula is given as:

$$\hat{V}ar[\hat{S}_{KM}(t)] = [\hat{S}_{KM}(t)]^2 \sum_{t_{(i)} \leq t} {}^{d_i}/_{n_i(n_i - d_i)} \quad \text{.....(8)}$$

An alternative estimator for the Kaplan-Merier estimator is the Nelson-Aalan estimator. Hereby we used the link between the survival functon and the cumulative hazard function, $S(t)$=exp(− (t)). Estimating first the cumulative hazard function , we then get another estmator for the survival function.

$$\hat{\Lambda}(t) = \sum_{t_{(i)} \leq t} \frac{d_i}{n_i} \Rightarrow \tilde{S}_{NA}(t) = \exp[-\hat{\Lambda}(t)] = \prod_{t_{(i)} \leq t} \exp\{-\frac{d_i}{n_i}\} \text{.....(9)}$$

It is merely in the case of small samples that the Nelson-Aalen estimate of the survivor function prevails over the KM estimate (Hosmer and Lemeshow, 1999). Moreover , the Kaplan-Meier estimate of the survivor function can be regarded as an approximate to the Nelsen-Aalan estimate.

$$\hat{S}_{KM}(t) \approx \tilde{S}(t) = \prod_{t_{(i)} \leq t} \exp\{-\frac{d_i}{n_i}\} \quad \text{.....(10)}$$

Log-rank test

The log rank test, developed by Mantel and Haenszel, is a non-parametric test for comparing two or more independent survival curves. Since it is a non-parametric test, no assumption about the distributional form of the data is required. This test is however most powerful when used for non-overlapping survival curves. This test can be generalized to accommodate other tests that are equally used sometime in practice such as Generalized Wilcoxon test, Tarone-Ware test, and Peto-Peto-Prentice test. Each of these tests uses different weights to adjust for censoring that is often encountered in survival data. For instance, the Wilcoxon test weights the j^{th}failure time by n_j(the number still at risk), the Tarone–Ware test weights the j^{th} failure time by $\sqrt{n_j}$and the Peto- Peto-Prentice test weights the j^{th}failure time by the survival estimate $\tilde{S}(t_j)$calculated over all groups combined (Kleinbaum and Klein, 2005 and Hosmer and Lemeshow, 1999). The log rank test statistic for comparing two groups is given by:

$$Q = \frac{[\sum_{i=1}^{m} w_i(d_{1i}-\hat{e}_{1i})]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \quad \ldots\ldots (11)$$

Where

- m is the number of rank ordered event (death) times.
- d_{1i} is the observed number of events (death in group1at event time $t_{(i)}$).
- $\hat{e}_{1i} = \frac{n_{1i}-d_i}{n_i}$ is the expected no of events (death) corresponding to d_{1i}.
- n_{1i} is the number of individuals at risk in group1 just prior to event (death) time $t_{(i)}$.
- $\hat{v}_{1i} = \frac{n_{1i}n_{2i}d_i(n_i-d_i)}{n_i^2(n_i-1)}$ is the variance of the number of events d_{1i}at time $t_{(i)}$.
- $n_{2i}$is the number of individuals at risk in group 2 just prior to event (death) time $t_{(i)}$.
- n_i and d_i are the number of individuals at risk and number of death in both groups (i.e., group 1 and group 2) just prior to event time $t_{(i)}$respectively.

Under the null hypothesis that two survival functions are equal, the log rank test statistic Q has an approximation of chi-square distribution with one degree of freedom $X^2(1)$for

large samples. The null hypothesis of equality of survival functions will be rejected for large values of Q. The most frequently used test is based on weights equal to one $w_i = 1$. Note that the log-rank test can be extended for comparing three or more groups of survival experience.

3.6 Parametric Survival Models

Linear regression, logistic regression, and Poisson regression are examples of parametric models that are commonly used in the health sciences. With these models, the outcome is assumed to follow some distribution such as the normal, binomial, or Poisson distribution. Typically, what is actually meant is that the outcome follows some family of distributions with unknown parameters. It is only when the value of the parameter(s) is known that the exact distribution is fully specified. A parametric survival model is one in which survival time (the outcome) is assumed to follow a known distribution. Examples of distributions that are commonly used for survival time are: the Weibull, exponential (a special case of the Weibull), log-logistic, lognormal, Gompertz, and generalized gamma (David G. and Mitchel K., 1996).

Parametric models make assumptions about the distribution of failure times and the relationship between covariates and survival experience. Parametric models fully specify the distribution of the baseline hazard/survival function according to some (defined) probability distribution. Parametric models are useful when we want to predict survival rather than identify factors that influence survival. Parametric models can be expressed in: (1) proportional hazard form, where a one unit change in an explanatory variable causes a proportional change in hazard; and (2) accelerated failure time (AFT) form, where a one unit change in an explanatory variable causes a proportional change in survival time.

3.6.1 Parametric Proportional Hazards model

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

$\lambda(t/\boldsymbol{x})=\lambda_0(t)\exp\{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\}=\lambda_0(t)\exp\{\boldsymbol{\beta}'\boldsymbol{x}\}$, where $\lambda_0(t)$= baseline hazard function and $\boldsymbol{\beta}' = (\beta_1, \beta_p, \dots, \beta_p)$ is a vector of regression coefficients. The

baseline hazard function is assumed to follow a specific distribution when a fully parametric PH model is fitted to the data. The hazard ratio is hence given by $HR = \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$.

3.6.2 Accelerated Failure Time Model

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 accelerated failure time model (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.

For a group of patients with covariate $(x_1, x_2, \ldots, x_p)$, the model is written mathematically as $S(t/\boldsymbol{x}) = S_0(t/\phi(x))$, where $S_0(t)$ is the baseline survival function and ϕ 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 $\phi(\boldsymbol{x}) = \exp(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p)$.

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

$\log(T_i) = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_p x_{pi} + \sigma\varepsilon_i$, where μ is intercept, σ is scale parameter and ε_i is random variable, assumed to have a particular distribution. For each distribution of ε_i, there is a corresponding distribution for T.

The advantage of the accelerated failure time approach is that the effect of covariates on survival can be described in absolute terms (e.g. numbers of years) rather than relative terms (a hazard ratio).

a) **Exponential Distribution**

The exponential distribution is the only distribution with a constant hazard i.e (t,) = , > 0. This implies that the conditional „probability" of an event is constant over time. In other words, the risk of an event occurring is flat with respect to time. The survivor function is $S(t,)$ =exp {− (t)} and the density is $f(t,) =$ exp(− t). It can be shown that $E(T) = {}^{1}/_{\lambda}$ and $Var(T) = {}^{1}/_{\lambda^2}$.

b) **Weibull Distribution**

T is Weibull with parameter $\lambda > 0$ and $\rho > 0$, denoted $T \sim W(\lambda, \rho)$. The cumulative hazard is $\Lambda(t, \lambda, \rho) = \lambda t^{\rho}$, the survivor function is $S(t, \lambda, \rho) = \exp\{-\lambda t^{\rho}\}$, and hazard is $\lambda(t, \lambda, \rho) = \rho\lambda t^{\rho-1}$.

The Weibull model is more general and flexible than the exponential model and allows for hazard rates that are non-constant but monotonic. It is a two-parameter model (λ and ρ), where, λ is the location parameter and ρ is the shape parameter ρ determines whether the hazard is increasing, decreasing, or constant over time. The shape parameter works in the following way:

- If $0 < \rho < 1$, then the hazard is monotonically decreasing with time.
- If $\rho = 1$, then the hazard is flat and we have the exponential model i.e. the Weibull model nests the exponential model. This means that we can use the Weibull model to test to see if the exponential model is appropriate.
- if $\rho > 1$, then the hazard is monotonically increasing with time.

The exponential and Weibull distributions can accommodate both the PH and AFT assumptions. The interpretation of parameters differs for AFT and PH models. The AFT assumption is applicable for a comparison of survival times whereas the PH assumption is applicable for a comparison of hazards.

c) **Gompertz Distribution**

A random variable T has the Gompertz distribution with the following hazard, density and survivorship functions $\lambda(t,\lambda,\rho) = \lambda\exp(\rho t)$, $S(t,\lambda,\rho) = exp\left(\frac{\lambda}{\rho}(1-\exp(\rho t))\right)$, $f(t,\lambda,\rho) = \lambda\exp(\rho t)\ exp\left(\frac{\lambda}{\rho}(1-\exp(\rho t))\right)$, where the scale parameter $\lambda > 0$, and shape parameter $\rho\epsilon(-\infty,\infty)$.

The Gompertz model is a parametric proportional hazards model but not an AFT model.

If > 0 , then the hazard exponentially increases over time. If $\rho < 0$ then the hazard decreases exponentially over time. If $\rho = 0$ then the hazard is constant and reduces to the exponential model.

d) **Log-logistic Distribution**

A random variable T has the log-logistic distribution with the following hazard, density and survivorship function $\lambda(t,\lambda,\rho) = \frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}}$, $S(t,\lambda,\rho) = \frac{1}{(1+\lambda t^{\rho})}$, and $f(t,\lambda,\rho) = \frac{\lambda\rho t^{\rho-1}}{(1+\lambda t^{\rho})^2}$, where scale parameter $\lambda > 0$, shape parameter $\rho > 0$.

- If $\rho < 1$, the hazard decreases monotonically from ∞ over time.
- If $\rho = 1$, the hazard decreases monotonically from λ.
- If $\rho > 1$, however, the hazard increases to a maximum point and then decreases over time. In this case ($\rho > 1$), the hazard function is said to be unimodal.

Unlike the Weibull model, a log-logistic AFT model is not a PH model. However, the log-logistic AFT model is a proportional odds (PO) model. A proportional odds survival model is a model in which the odds ratio is assumed to remain constant over time. This is analogous to a proportional hazard model where the hazard ratio is assumed constant over time.

The survival odds are the odds of surviving beyond time t (i.e. $S(t)/(1-S(t))$. This is the probability of not getting the event by time t divided by the probability of getting the event by time t. The failure odds is the odds of getting the event by time t (i.e., (1 −

$S(t))/S(t)$), which is the reciprocal of the survival odds $(1 - S(t)/S(t)$. The failure odds simplifies in a log-logistic model to λt^{ρ} as follows:

$$\frac{1-S(t)}{S(t)} = \frac{P(T \leq t)}{P(T>t)} = \frac{1-\frac{1}{(1+\lambda t^{\rho})}}{\frac{1}{(1+\lambda t^{\rho})}} = \frac{\frac{\lambda t^{\rho}}{(1+\lambda t^{\rho})}}{\frac{1}{(1+\lambda t^{\rho})}} = \lambda t^{\rho}$$

The underlying assumption for AFT models is that the effect of covariates is multiplicative (proportional) with respect to survival time, whereas for PH models the underlying assumption is that the effect of covariates is multiplicative with respect to the hazard.

3.7 Parameterization

When we say proportional hazards (PH) it means that the hazard function of a group is proportional to the hazard function of the other group, i.e., the hazard ratio is constant over time (Klein, 1992). The hazard ratio is hence given by $HR = e^{\beta_j}$, where $\beta' = (\beta_1, \beta_p, \dots, \beta_p)$ is a vector of regression coefficients. On the other hand, the acceleration failure time (AFT) model describes stretching out or contraction of survival time as a function of predictor variables. The acceleration factor which is usually denoted by $\phi = exp(\alpha_i)$, where $\alpha' = (\alpha_1, \dots, \alpha_p)$ is a vector of regression coefficients in case of AFT model. For the exponential, Weibull, and log-logistic survival model, the relationship between α and β is given below the following three. (Note that since Gompertz model is not an AFT model, parameterization is not needed.)

a) For exponential, $\beta_j = -\alpha_j$, the exponential PH and AFT are in fact the same model, except that the parameterization is different, and hence $HR = expe(-\alpha_j)$ is the hazard ratio of the j^{th} covariate with the reference group.
b) For Weibull, $\beta_j = -\alpha_j \rho$, where ρ is the shape parameter and hence, $HR = exp(-\alpha_j \rho)$ is the hazard ratio of the j^{th} covariate with the reference group.
c) For log-logistic, $\beta_j = -\alpha_j \rho$, where the ρ is the shape parameter and $OR = exp(-\alpha_j \rho)$ indicates the failure odds ratio of the j^{th} covariate with the reference group.

3.8 Method of Parameter Estimation

All parametric models may be fit by maximazing the appropraite likilihood function. In each distribution there are several parametr(s) which determined the shape of this distribution. In survival analysis, some observations are censored. Hence, estimation method has to be adopted to censoring. In parametric modeling, maximum likelihood estimation is commonly used.

No censoring

Let $T_1, T_2, \dots, T_n$ be a sample from a population $T \sim F(t, \theta)$. $F(t, \theta)$ is a continous distribution with density function $f(t, \theta)$. The likelihood function is defiend as $L(\theta) = \prod_{i=1}^{n} f(t_i, \theta)$. We estimate θ by maximizing this expression ($\hat{\theta}$).

With censoring

Suppose we have a censored sample (X_1, δ_1) , $((X_2, \delta_2), \dots$, (X_n, δ_n) where

X_i=$min(T_i, C_i)$ and δ_i=$I(T_i \leq C_i)$, $i = 1,2,\dots,n$

With

- a sample $T_1, T_2, \dots, T_n \sim f(t, \theta)$ life times. We denote the survival function by $S(t, \theta)$.
- a sample $C_1, C_2, \dots, C_n \sim g(c)$ censoring times with survival function $G(t)$.
- T_i and C_i are inependent.

For an uncensored Observation ($\delta = 1$), we calculate the contribution to the likelihood by $P(Y \leq y, \delta = 1) = P(\min(T, C) \leq y, T \leq C)$

$$= P(T \leq y, C \geq T)$$

$$= \int_0^y C(t) f(t, \theta)\, dt \text{ , Implies that } f_{Y,\delta=1}(y, \theta) = f(y, \theta) G(y)$$

For a censored observation ($\delta = 0$), we get similarly $f_{Y,\delta=0}(y, \theta) = g(y, \theta) S(y)$

Hence ,we get the likelihood function

$L(\theta) = \prod_{i,\delta_i=1} f(y_i, \theta) G(y_i) \quad \prod_{i,\delta_i=0} g(y_i) S(y_i, \theta)$

To the maximum we check that $\frac{\partial^2 l(\theta)}{\partial^2 \theta} \leq 0$, where $l(\theta) = ln[L(\theta)]$ and in general must be maximized numerically using a procedure such as Newton-Raphson.

Selection of covariates

The methods available to select a subset of covariates to include in a parametric survival model are essentially the same as those used in any other regression model. There are three methods of selection of influential covariates. These are purposeful selection, stepwise selection (forward selection and backward elimination) and best subset selection. Survival analysis using parametric regression method begins with a thorough univariable analysis of the association between survival time and all important covariates (Hosmer and Lemeshow, 1999).

3.9 Comparison of Models

Model comparison and selection are among the most common problems of statistical practice, with numerous procedures for choosing among a set of models (Kadane and Lazar, 2001) and (Rao and Wu, 2001). There are several methods of model selection. The most commonly used methods include Akaike information and likelihood based criteria. A data-driven model selection method such as an adapted version of Akaike's information criterion AIC (Akaike, 1974) is used to find the truncation point of a series of models. In some circumstances, it might be useful to easily obtain AIC value for a series of candidate models (Munda et al., 2012). In this study, we used the AIC criterion and log likelihood to compare four of parametric models. AIC is defined as

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

where l is the log-likelihood, k 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 non predictive parameters are added to the model. Small values of AIC suggest a better model.

3.10 Model Diagnostics (Checking)

The use of diagnostic procedures for model checking diagnostics is an essential part of the modeling process. There are different commonly used model diagnostics to evaluate

whether the appropriate functional form for a covariate is used in the model to assess the fitted model.

3.10.1 Cox-Snell Residuals

The Cox-Snell residuals method can be applied to any parametric model and the residual plots can be used to check the goodness of fit of the model. For the parametric regression problem, analogs of the semi-parametric residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates (Klein and Moeschberger, 2003).

The Cox-Snell residual for the i^{th} individual with observed survival time t_i is given by $r_i = \widehat{H}(T_i / X_i) = -\log(\hat{\mathrm{S}}(\mathrm{T_i} / \mathrm{X_i}))$, where $\widehat{H}$ and $\hat{\mathrm{S}}$ are the estimated values of the cumulative hazard and survivor function of the the i^{th} subject at time t_i. If the model fits the data , then $r_i's$ should have a standard ($\lambda = 1$) exponential distribution , so that a a hazard plot of r_i versus the Nelson-Alan estimator of the cumulative hazard of the $r_i's$ shoud be a stright line with slope unity and zero intercept. If yes, the fitted model is aduquate. In general ,Cox-Snell residual provides a check of the overall fits of the model (Cox and Snell, 1968).

3.10.2 Quantile - Quantile Plot

An initial method for assessing the potential for an AFT model is to produce a quantile-quantile plot. The plot is based on the fact that, for the accelerated failure-time model, $S_1(t) = S_0(\phi t)$, Where S_0 and S_1 are the survival functions in the two groups and ϕ is the acceleration factor. Let t_{0p} and t_{1p} be the p^{th} percentiles of groups 0 and 1, respectively, that is $t_{lp} = {S_l}^{-1}(1-p), l = 0,1$.

Using the relation $S_1(t) = S_0(t\phi)$, we must have $S_0(t_{0p}) = 1 - p = S_1(t_{1p}) = S_0(\phi t_{1p})$ for all t if the accelerated failure time model holds,$t_{0p} = \phi t_{1p}$.To check this assumption we compute the Kaplan–Meier estimators of the two groups and estimate the percentiles t_{0p}, t_{1p} for varous values of p. If we plot the estimated percentile in group 0 versus the estimated percentile in group 1(i.e., plot the points t_{1p}, t_{0p} for various values of p), the graph should be a straight line through the origin, if the accelerated

failure time model holds. If the curve is linear, a crude estimate of the acceleration factor q is given by the slope of the line (Klein, 1992).

3.11 Ethical Consideration

The data for the analysis were obtained from Zewditu Referral Hospital and an Ethical clearance for the study was provided by the Research Ethics Review Board of AAU.

CHAPTER FOUR

4. STATISTICAL DATA ANALYSIS AND DISCUSSION

4.1 Results of Descriptive Statistics

We used descriptive statistics to get some information about the distribution of the variables. The response variable in this study is survival time measured in months from HAART start to death/censor which is continuous. The censoring indicator is 0 for censored observations and 1 for event occurred; in our case death. Therefore, the outcome (response) variable is time-to-death.

The patients were followed up for a median period of 51 months. The minimum and maximum follow-up time was 3 and 59 months, respectively. Of all 638 HIV infected adult patients, 64(10%) died during the follow up period. The overall mean estimated survival time of patients under the study was 44 months. The survival (censor) time of the patients is skewed to the right, distribution with 25^{th}, 50^{th},75^{th} percentiles of the adult patients was 34,51,and 55 months, respectively. The majority of the patients were females 370(58%). Regarding educational attainment, about 71(11.1%) of the adult patients had no education (were illiterate) while 201(31.5%) of the adult patients had attended primary education and the remaining 366(57.4%) of the adult patients had attended secondary and above educations. A total of 482(75.5%) of HIV infected adult patients were able to work. The remaining 96(15%) and 60(9.4%) were ambulatory and bedridden respectively. And also 376(58.9%) of the HIV infected adult patients had TB. Among the 64(10%) dead adult patients, 24(3.8%) and 33(5.2%) were under WHO clinical stage III and IV respectively, indicating that about 9% of the dead patients were in WHO clinical stages III and IV. Regarding the marital status, 170(12.5%) of the subjects were single while 223(35%) of the subjects were married, and the marital status of remaining 335(52.5) was not known (see Tables 4.1 and 4.2).

Table 4. 1**:** Descriptive summary of baseline demographic and clinical characteristics of HIV infected adult patients under HAART in Zewditu Referral Hospital, AA, 2010.

Covariate/factor Variable	Category	Status of censoring and event		
		Censored (%)	Dead (%)	Total (%)
Age	<30	121(19)	5(0.8)	126(19.7)
	[30-39]	269(42.2)	19(3)	288(45.1)
	[40-49]	119(18.7)	22(3.4)	141(22.1)
	>=50	65(10.2)	18(2.8)	83(13)
Functional status	Working	457(71.6)	25(3.9)	482(75.5)
	Ambulatory	77(12.1)	19(3)	96(15)
	Bed ridden	40(6.3)	20(3.1)	60(9.4)
TB screen	No	254(39.8)	8(1.3)	262(41.1)
	Yes	320(50.2)	56(8.8)	376(58.9)
OIs	No	199(31.2)	12(1.9)	211(33.1)
	Yes	375(58.8)	52(8.2)	427(66.9)
WHO stage	WHO clinical stage I	121(19)	3(0.5)	124(19.4)
	WHO clinical stage II	170(26.6)	4(0.5)	174(27.3)
	WHO clinical stage III	195(30.6)	24(3.8)	219(34.3)
	WHO clinical stage IV	88(13.8)	33(5.2)	121(19)
Sex	Female	330(51.7)	40(6.3)	370(58)
	Male	244(38.2)	24(3.8)	268(42)
Marital status	Never married	70(11)	10(1.6)	170(12.5)
	Married	207(32.4)	16 (2.5)	223(35)
	Other	297(46.6)	38(6)	335(52.5)
Educational level	No education	55(8.6)	16(2.5)	71(11.1)
	Primary	181(28.4)	20(3.1)	201(31.5)
	Secondary and above	338(53)	28(4.4)	366(57.4)
Religion	Muslim	127(19.9)	16(2.5)	143(22.4)
	Coptic Orthodox	241(37.8)	27(4.2)	268(42)
	Other	206(32.3)	21(3.3)	227(35.6)

Table 4. 2: Summary statistics of baseline continuous variables (considered in this study) of HIV infected adult patients under HAART in Zewditu Referral Hospital, AA, 2010.

Patient Status	Continuous Variables	Mean	Standard deviation	Min.	Max.	Median	Q_1	Q_3
Censored	Time	46.8885	13.36446	3	59	52	48	55
	Age	36.46516	9.822225	16	99	35	30	41
	Weight	57.57143	11.06955	17	104	56	50	65
	CD4	160.1951	110.1105	2	978	147	85.25	212

Event/Death	Time	18.1875	13.11594	3	55	14.5	8.75	24
	Age	43.03125	10.24148	22	70	45	36	50
	Weight	50.40625	10.37965	32	85	50	44	54.25
	CD4	105.75	81.14713	2	420	91	44	141.5
Overall	Time	44.0094	15.87886	3	59	51	34	55
	Age	37.12382	10.05255	16	99	36	30	42
	Weight	56.85266	11.20356	17	104	55	50	64
	CD4	154.7335	108.7446	2	978	142	80.25	208

Table 4. 3: The mean and standard deviation of HIV infected adult patients longitudinally measured CD4 count at each visit time in Zewditu Referral Hospital, AA, 2010.

Time(month)	0	6	12	18	24	30	36	42	48	54	60
Mean (CD4)	154.7	160.0	190.0	255.3	310.3	351.5	381.3	416.5	444.8	452.7	465.4
Std.Dev(CD4)	108.8	110.3	110.1	114.0	127.5	124.3	133.3	130.1	131.4	118.8	123.6

As shown in Table 4.3, the actual mean of the CD4 count was increasing over time. This shows that after patients initiated to HAART the average CD4 count increased due to the positive effect of the therapy.

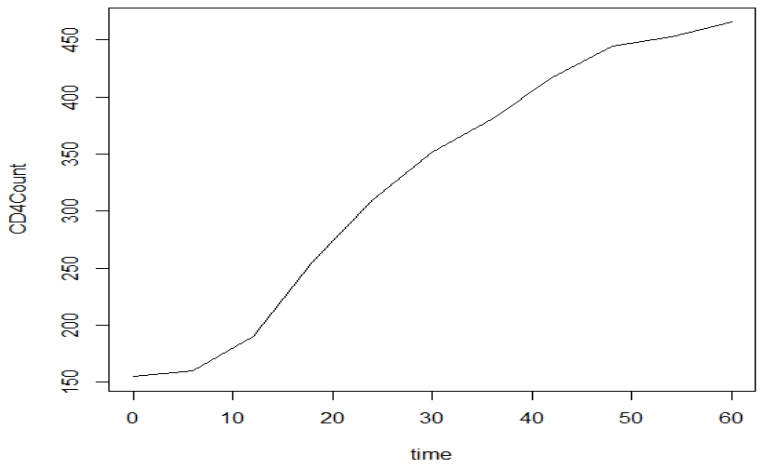


Figure 1**:** The average progression of actual CD4 count of HIV infected adult patients under HAART in Zewditu Referral Hospital, AA, 2010.

Figure 1 depicts that the mean CD4 count evolution shows an increase the patient‟s immune system or the progression of the disease declines over time (i.e. because CD4 count and HIV infection are negatively correlated).

4.1.1 Survival Function of Different Categorical Group of Covariates

The main interest was to compare the distributions of time-to-death by estimating survival function of different categorical covariates. Descriptive graphs of survivor function would be used for the purpose of comparing the event experiencing time of two or more groups and the survival quantities of covariates to describe survival experience. In order to get a closer look at estimate of the survival time we use the Kaplan-Meier and Nelson-Aalen estimation techniques. The estimated hazard function depicted in Figure 2 below shows that an increase in the hazard rate has direct relation with the increase in time.

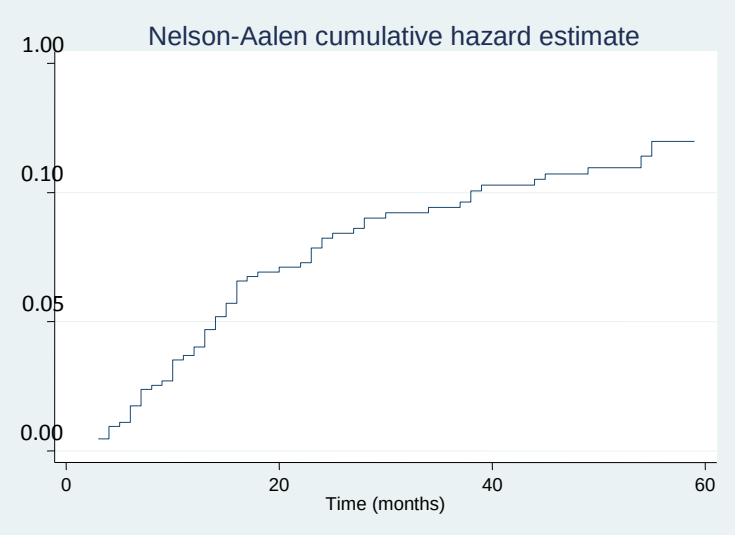


Figure 2: The Nelson-Aalen estimated cumulative hazard function of HIV infected adult patients under HAART in Zewditu Referral Hospital, AA, 2010.

Figure 3 is the estimate for overall Kaplan-Meier survivor function. It depicts that, relatively, a large number of the deaths occurred at the earlier months of HAART treatment, and a decrease over the follow up period.

Separate Kaplan-Meier survivor functions are constructed for different covariates to see for possible existence of differences in survival experience between the indicated categories. In general, the pattern of one survivorship function lying above another means the group defined by the upper curve had a better survival than the group defined by the lower curve. The differences are not clear among categories of gender; marital status and religion (see Appendix 3).

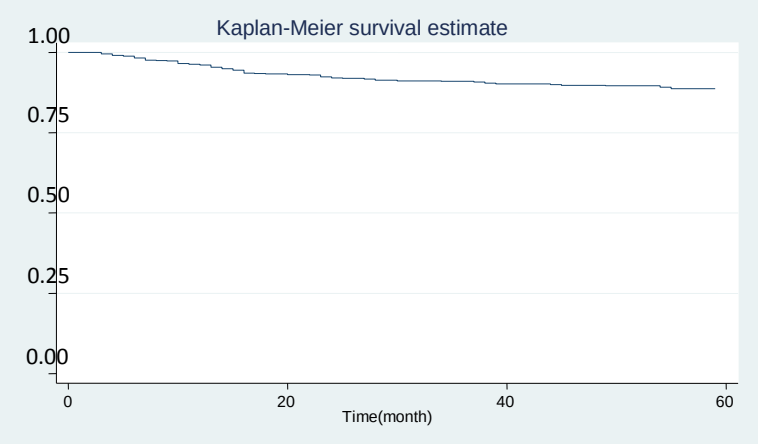


Figure 3: The plot of the overall estimate of Kaplan-Meier survivor function of HIV infected adult patients under HAART in Zewditu Referral Hospital, AA, 2010.

Age, functional status, TB screen, WHO clinical stage, and educational level manifest relatively larger gaps. For instance patients in working functional status have longer experience of survival time than those in ambulatory status and those who are bedridden. Patients in WHO clinical stages I and II have longer survival time than those in stages III and IV.

To check for significance differences among categories of factors that are shown using the Kaplan-Meier estimates of the survivor functions, we employ the log-rank statistical test. Based on the log-rank test, there were no significant differences in survival experience between the various categories of gender, marital status, and religion. However, the log-rank test showed that the survival experience of HIV infected adult patients in different categories of age, functional status, TB screen, OIs, WHO clinical stage and educational level differ significantly (Table 4.4).

Table 4. 4: Results of the log-rank test for each categorical variables of HIV infected adult patients under HAART in Zewditu Referral Hospital, AA, 2010.

Covariate/Factor	DF	Chi-square	p-value
Age	3	26.06	0.0000
Functional status	2	82.05	0.0000
TB screen	1	24.44	0.0000
OIs	1	6.96	0.0083
WHO clinical stage	3	68.92	0.0000
Gender	1	0.49	0.4855
Marital status	2	3.72	0.1558
Educational level	2	13.30	0.0013
Religion	2	0.47	0.7919

4.2 Results of the Univariate Parametric Survival Models

The aim of model development is to obtain a model that satisfactorily describes the data at hand. For this purpose, the first step is to select covariates which are important in a study at some relaxed level of significance. We used univariate analysis in order to see the effect of each covariate on the time-to-death before proceeding to the multivariable analysis. The univariate analyses were fitted for every covariate that had a p-value of less than 0.20 by different parametric survival models (Appendix 1).

In the univariate analysis we found that age, weight, functional status, TB screen, OIs, CD4 count, WHO clinical stage and educational level are significant at 20% level. In addition, confidence intervals of the acceleration factors $\phi_j = e^{\beta_j}$ for all covariates of significant categories do not include 1 in exponential, Weibull, and log-logistic models at 20% level of significance. Confidence intervals of the coefficients for all covariates of significant categories do not include 0 in the Gompertz model. This indicates that they are important prognostic factors of time-to-death. Therefore, based on these results, we ignore the factors gender, marital status, and religion and proceed the multivariable analysis using the above eight significant covariates. Hence, the effects of the remaining eight significant covariates on the time-to-death of HIV infected adult patients shall be interpreted using multivariable analysis.

4.3 Multivariable Analysis and Model Comparison

Multivariable analysis of exponential, Weibull, Gompertz, and log-logistic parametric models is done by using all significant covariates in univariate analysis at 5% level of significance (Appendix 2). We used backward elimination method to select the significant covariates/factors. In all used models of multivariable analysis age, weight, functional status, TB screen, CD4 count, WHO clinical stage and educational level were significant at 5% level and model comparison was done using those covariates.

From Table 4.5, we can see that the values of AIC and log likelihood of the four parametric models. In this case, we used AIC and log likelihood to compare the models. The lowest value of AIC in combination with the largest value of log likelihood is a criterion to select a model. The AIC value of the log-logistic model i.e. 448.9452 is the smallest. The largest log likelihood value of the log logistic model is -209.4726. This indicates that the log-logistic model is the most efficient model to describe the HIV infected adult patients dataset among the candidates parametric model.

Table 4. 5: AIC and log likelihood of the candidate parametric models

Model	AIC	Log likelihood
Exponential	452.2243	-212.1122
Weibull	454.2235	-212.1118
Gompertz	450.9601	-210.4801
Log-logistic	448.9452	-209.4726

Multivariable analysis based on log-logistic model shows that, all covariates were significant except some category of age and WHO clinical stage (Table 4.6).

The 95% confidence intervals of the acceleration factor for all significant categories of the covariates do not include 1 at 5% level of significance. This shows that they were prognostic covariates for determining the time-to-death of HIV infected adults patients. The estimated coefficient of the parameters for patients who had TB screen was -1.125203. The sign of the coefficient is negative which implies that decreasing logged survival time. Hence, their death time will decrease by a factor $\hat{\phi}$=0.3245866 than the reference category (patients who had not TB screen) at 5 % level of significance.

The acceleration factor for functional status of HIV infected adult patients was 0.4531and 0.3308 for group of ambulatory and bedridden respectively using working groups as a reference category. This indicates that for ambulatory and bedridden groups survival is reduced by a factor $\hat{\phi}$=0.4531 and $\hat{\phi}$ =0.3308, respectively, than the reference group at 5% level of significance. The coefficients of categorical variable age, shows the survival of age group [40-49] and >=50 years were reduced by a factor of ($\hat{\phi}$=0.220243) and ($\hat{\phi}$=0.1598542), respectively, by using age group younger than 30 years as a reference category.

The acceleration factors for those adult patients who had attended the primary and secondary and above educations were 2.7798 and 2.9006 respectively. This indicates that the two groups of primary and secondary and above were significantly prognostic factors for timing of time-to-death by using illiterate category as a reference. An acceleration factor of greater than 1 indicates prolonging the survival. Therefore, for patients who attended primary education death time was longer by a factor of $\hat{\phi}$=2.7798 than the reference group. For patients who attended secondary and above education was the factor 2.9006 relative to the reference group.

The acceleration factor and 95% CI of acceleration factor for WHO clinical stage of HIV infected adult patients who were in stage IV was 0.274140 and (0.08412, 0.89337), respectively, compared with patients in stage I as a reference category. This indicates patients with stage IV their survival was shrank by a factor of $\hat{\phi}$=0.274140 compared with patients who were in stage I.

For a 10 Kg change in weight the log of time is increased by 0.377, holding the remaining covariates constant. Similarly, for 100 cells/mm^3 change in CD4 count log of time is increased by 0.39, holding the remaining covariates constant.

The value of the shape parameter in the log-logistic model is ρ= 1.617352. Since this value is greater than unity the hazard function is unimodal.

Table 4. 6: Results of the multivariable analysis of log-logistic model

Covariate/factor Variable	$\hat{\beta}_j$	s.e	$\hat{\phi}_j$	p-value	95 % CI for ϕ	
					LCL	UCL
Age in years						
<30	Ref.					
[30-39]	-0.7687	0.4992	0.4636379	0.124	0.17427	1.23346
[40-49]	-1.5130	0.5131	0.220243	0.003	0.08056	0.60212
>=50	-1.8335	0.5443	0.1598542	0.001	0.05501	0.46453
Weight	0.0377	0.0136	1.03842	0.006	1.01106	1.06647
Functional status						
Working	Ref.					
Ambulatory	-0.7917	0.3333	0.4530713	0.018	0.23576	0.87069
Bedridden	-1.1063	0.3635	0.3307922	0.002	0.16222	0.67455
TB screen						
No	Ref.					
Yes	-1.1252	0.3788	0.3245866	0.003	0.15446	0.68203
CD4 count	0.0039	0.0016	1.00386	0.019	1.00065	1.0071
WHO clinical stage						
WHO stage1	Ref.					
WHO tage2	0.2011	0.6886	1.222776	0.770	0.31706	4.71572
WHO tage3	-0.6246	0.5763	0.5354494	0.278	0.17303	1.65698
WHO tage4	-1.2941	0.6027	0.2741401	0.032	0.08412	0.89337
Educational level						
No education	Ref.					
Primary	1.0224	0.3768	2.779878	0.007	1.32813	5.81849
Secondary and above	1.0649	0.3552	2.900569	0.003	1.44599	5.81833
Intercept	5.1358	1.0687	170.0018	0.000	20.9286	1380.91
Log likelihood = -209.47258, AIC=448.9452 , ρ= 1.617352						

$\hat{\beta}_j$= coefficient estimate, s.e= standard error, $\hat{\phi}_j$ = acceleration factor estimate, 95% CI=Confidence Interval for acceleration factor, LCL=lower class limit, UCL= upper class limit, Ref=Reference, ρ = shape parameter, AIC= Akaike Information Criterion.

4.4 Model Diagnostics

4.4.1 The Cox Snell Residual Plots

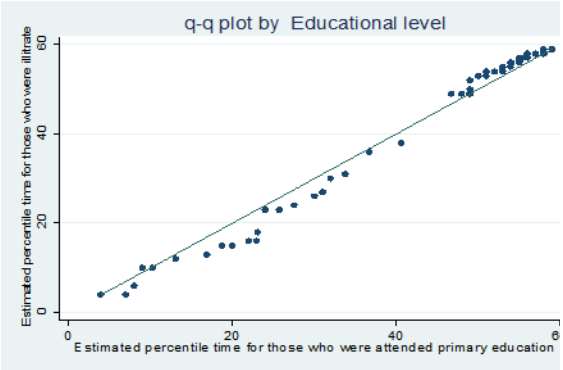
The Cox- Snell residuals (together with their cumulative hazard function) had been obtained from fitting using the exponential, Weibull , Gompertz and log-logistic models to our data via maximum likelihood estimation. It can be seen that the plot of the cumulative hazard function against Cox-Snell residuals (Figure 6 of Appendix 3) is

closest to the 45^0 straight lines through the origin for log-logistic model when compared to exponential, Weibull and Gompertz models. This suggests that log-logistic model provided the best fit for the HIV infected adult patients under HAART dataset.

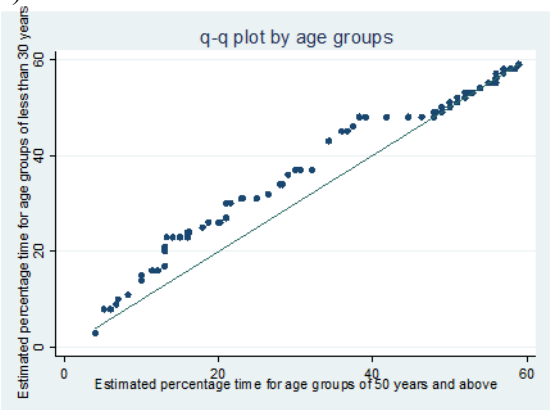
4.4.2 Adequacy of Accelerated Failure Time (q-q plots)

A quantile-quantile or q-q plot is used to check if the accelerated failure time provided an adequate fit to the data from two different groups of the population. We shall graphically check the adequacy of the accelerated failure-time model by comparing the significantly different age groups (HIV infected patients in the age group less than 30 years and 50 years and above), patients who were illiterate and patients who attended secondary and above groups. The plots in Figure 4 below appear to be approximately linear for both covariates (age group and educational level) with slopes equivalent to the acceleration factors 0.1598542 and 2.900569, respectively. The q-q plot approximates a 45^0 straight line through the origin indicating that the AFT model is appropriate model.

a)



b)



c)

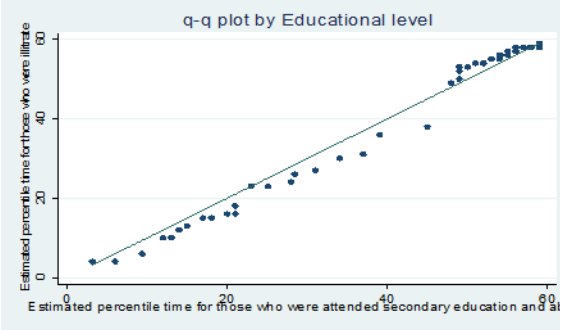


Figure 4: q-q plot to check the adequacy of the accelerated failure time model

4.5 Discussion

The main aim of the study was to fit parametric survival models to time-to-death data. We considered four parametric survival models (exponential, Weibull, Gompertz, and log-logistic). The study also attempted to determine the factors/variables for time-to-death of HIV infected adult patients taken from Zewditu Referral Hospital.

Univariate and multivariable survival models were employed to examine the factors that determine time-to-death. Factors/variables considered in the study were age, weight, functional status, OIs, TB screen, CD4 count, WHO clinical stage, educational level, gender, marital status, and religion. In the univariate results given in Appendix 1 all variables except gender, marital status and religion were significantly associated with time-to-death in all models at 20% level of significance. The significant variables in the univariate analysis were included in multivariable analysis. In all four models age, weight, functional status, WHO clinical stage, CD4 count and educational level were significant. Hence, these covariates were used in model comparisons.

The comparison of the four parametric models was done by using AIC criterion and log likelihood, where a model with smallest AIC and largest log likelihood will be taken to be the most appropriate. Accordingly, log-logistic model which had AIC value of 448.9452 and log likelihood -209.4726 was the most appropriate model to describe the HIV infected patient"s dataset.

Multivariable analysis using the log-logistic model showed that age, weight, functional status, TB screen, CD4 count, WHO clinical stage and educational level were prognostic factors/variables for the time-to-death.

A study conducted by Ang et al (2005) suggested that both univariable and multivariable analyses showed that patients of younger age and higher baseline CD4 cell count were associated with a lower risk of progression to AIDS. Our finding showed that the age groups 40-49 and 50 years and above are associated with low chance of survival. HIV infected patients with higher baseline CD4 count had a better chance of survival. This finding agrees with studies by Getie et al (2014), Egger and Chene (2002), Kitahata et al (2009).

The findings of this study revealed that the TB screen had a significant effect on the time-to-death. It shortened time-to-death by a factor $\phi = 0.3245866$ compared to those with no TB screen the reference category. A study conducted by Xueyan et al (2008) concluded that HIV/TB co-infection was related to high mortality even when HAART and/or drug therapy for TB was provided. World Health Organization reported in 1999 that TB co-infection is the leading cause of mortality among those infected with HIV worldwide.

The result of this study suggested that WHO clinical stage, age and CD4 count were significant predictive factors of time-to-death i.e. HIV infected patients of older age, in WHO Clinical stage IV, and with low CD4 count had lower chance survive. This finding is in agreement with studies by Lawn et al (2009).

CHAPTER FIVE

5. CONCLUSION AND RECOMMONDATIONS

5.1 Conclusions

This study was based on a dataset of HIV infected adult patients under HAART from Zewditu Referral Hospital with the objective to determine the prognostic factors of survival and applying parametric modeling of the time-to-death. Four parametric survival models (exponential, Weibull, Gompertz, and log-logistic) have been compared using HAART dataset. The result of our study showed that the log-logistic model was the most suitable one.

The result of multivariable log-logistic model showed that age, functional status, TB screen, weight, CD4 count, WHO clinical stage and educational level were significant prognostic factors/variables for time-to-death of HIV infected adult patients under HAART. The categories of [40-49] and >=50 years and WHO stage IV were significant. High educational level of HIV infected patients is associated with lower survival. It was also observed that patients who had TB screen, ambulatory and bedridden HIV infected patients have short survival.

HIV infected adult patients who were in WHO stage IV had very low chance of living than in stage I, II, and III. The average CD4 count showed a quadratic pattern of increase after patients initiated to HAART program.

5.2 Recommendations

These recommendations are based on the findings of this analysis of the HIV infected patients dataset taken from Zewditu Referral Hospital. Based on the study prognostic factors were identified for time-to-death. We recommend the following:

- Clinicians are expected to give training especially to illiterate patients. Currently the emphasis is on giving training for prevention purpose. But more emphasis should also give to those who infected by HIV.
- Further studies should be conducted in the regions of Ethiopia and identify other prognostic factors that are not identified in this study.

REFERENCES

- ✓ Akaike H. (1974). A new look at the statistical model identification. IEEE Trans Automatic Control.
- ✓ Ang, L.W., Chow, K.Y., Verghesse I., Chew, S.K., and Leo, Y.S. (2005). Measurable predictive factors for progression to AIDS among HIV-infected patients in Singapore. Ann Acad Med Singapore, 34(**1**):84-89.
- ✓ Baghestani, A. R., Hagizadeh, E., and Fatemi , S. R. (2010). Parametric model to analyse the survival of gastric cancer in the presence of interval censoring. Tumor, 69, 433-437.
- ✓ Battegay M., Nuesch R., Hirschel B., Kaufmann GR. (2006). Immunological recovery and antiretroviral therapy in HIV-1 infection: Division of infectious disease and hospital epidemiology the lancet infect. Dis, 6, 280-287.
- ✓ Cox, D. R. and Snell, E. J. (1968). A general definition of residuals with discussion. Journal of the Royal Statistical Society. Series B, 30, 248-275.
- ✓ David G. and Mitchel K. (1996). Statistics for Biology and Health: Survival Analysis a Self-Learning Text, (Second edition), 260-261.
- ✓ Derbe M., Monga DP., and Daka D. (2013). Immunological response among HIV/AIDS patients before and after ART therapy at Zewuditu Hospital, Addis Ababa, Ethiopia. American Journal of Research and Communication, (**1**):107-108.
- ✓ Egger M. and Chene G. (2002). Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy. A collaborative analysis of prospective studies. Lancet 360, 9327, 119–129.
- ✓ Ethiopian Demographic Health Survey, 2011.
- ✓ Ferradini L., Jeannin A., Pinoges L., Izopet J., Odhiambo D., Mankhambo L., Karungi G.,Szumilin E., Balandine S., Fedida G., Carrieri M., Spire B., Ford N., Tassie J.M., Guerin P.J. and Brasher C. (2006). Scaling up of highly active antiretroviral therapy in a rural district of Malawi: an effectiveness assessment. Lancet, 367:1335–13342.
- ✓ Hayat A., Suner A., Uyar B. (2010). Comparison of five survival models: breast cancer registry data from Ege University cancer research center. Turkey, J Med Sci, 30(**5**):1665-1674.

- ✓ Hosmer D.W. and Lemeshow S. (1999). Applied Survival Analysis. John Wiley and Sons, Inc., New York.
- ✓ Jiezhi Qi. (2009). Comparison of Proportional Hazards and Accelerated Failure Time models. Thesis, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
- ✓ Kadane J. and Lazar N. (2001). Methods and Criteria for Model Selection. Carnegie Mellon University, Technical Report, 759.
- ✓ Kakuda TN., Fletcher CV., Collier AC. Bone and joint infections. In: JT Dipro., RL Talbert, GC Yee., GR Matzke., BG Wells., LM Posey (2002). Pharmacotherapy a pathophysiologic approach. Mcgraw Hill Medical Publishing Division, United States of America, 5th edition, 2151-2174.
- ✓ Kaplan EL, Meier P. (1958). Nonparametric estimation from incomplete observations. J Am Stat Assoc, 53:457-481.
- ✓ Kitahata MM., Gange SJ., Abraham AG. (2009). Effect of early versus deferred antiretroviral therapy for HIV on survival. N. Engl. J. Med., 360 (**18**):1815–1826.
- ✓ Klein J. (1992). Survival analysis: techniques for censored and truncated data. Medical College of Wisconsin.
- ✓ Klein P. and Moeschberger L. (2003). Survival Analysis Techniques for Censored and Truncated Data - Google Books.htm.
- ✓ Klos M., Venter M, Milne PJ., Traore HN., Meyer D., Oosthuizen V. (2009). In vitro anti-HIV activity of five selected South Africa medicinal plant extracts. J. Ethnopharmaco, 124: 182-188.
- ✓ Lawn, S.D., Little F., Bekker L., Kaplan R., Campbel E., Orrell C. and Wood R. (2009). Changing mortality risk associated with CD4 cell response to antiretroviral therapy in South Africa. AIDS, 23:335–342.
- ✓ MacLennan CA., Liu MK., White SA. (2007). Diagnostic accuracy and clinical utility of a simplified low cost method of counting CD4 cells with flow cytometry in Malawi: diagnostic accuracy study. BMJ335, (7612), 190.
- ✓ Mellors JW., Munoz A.,Giorgi JV. (1997). Plasma viral load and CD4+lymphocytes as prognostic markers of HIV-1 infection. Ann. Intern. Med, 126 (**12**):946–954.
- ✓ Munda M., Rotolo F., and Legrand C. (2012). Parametric Frailty Models in R. Journal of American Statistical Association, 55:1-21.

- ✓ Nakhaee F. and Law M. (2011). Parametric modeling of survival following HIV and AIDS in the era of highly active anti-retroviral therapy: data from Australia, Eastern Mediterranean Health Journal, 17(**3**): 233.
- ✓ Panel on Antiretroviral Guidelines for Adults and Adolescents, guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. (2008).Department of Health and Human Services. OARAC, 1–139.
- ✓ Ponnuraja C. and Venkatesan P. (2010). Survival models for exploring tuberculosis clinical trial data an empirical comparison. Indian Journal of Science and Technology, 2(**7**):755-758.
- ✓ Rao C. and Wu Y. (2001). On Model Selection (with discussion). In P. Lahiri (Ed.), Model Selection, Volume **38** of IMS Lecture Notes - Monograph Series. Institute of Mathematical Statistics.
- ✓ Robbins G., Daniels B., Zheng H., Chueh H., Meigs J., Freedberg K. (2007). Predictors of antiretroviral treatment failure in an urban HIV clinic. J. Acquir. Immune Defic. Syndr. , 44 (**1**):30–37.
- ✓ Sam E., Kathy P., Adeeba K., Jennifer H., Somnuek S., John C., Kathleen F., Jialun Z., Matthew G . (2009). Long-term patterns in CD4 response is determined by an interaction between baseline CD4 cell count, viral load and time. The APHOD. J Acquir Immune Defic Syndr, 50:513–520.
- ✓ Seage GR., Gatsonis C., Weissman JS., Haas JS., Cleary PD., Fowler FJ., Massagli M P., Stone VE., Craven DE., Goldberg J., Coltin K. (1997). The Boston AIDS Survival Score (BASS): A Multidimensional AIDS severity instrument. American Journal of Public Health, 87(**4**).
- ✓ Seid A., Getie M., Birlie B., Getachew Y. (2014). Joint modeling of longitudinal CD4 cell counts and time-to-default from HAART treatment: a comparison of separate and joint models. Electronic Journal of Applied Statistical Analysis, 7(**2**): 292-314.
- ✓ Seoane E. and Resino S. (2008). Lipid and apoprotein profile in HIV-1-infected patients after CD4-guided treatment interruption. J Acquir Immune DeficSyndr, 48(**4**):455-459.
- ✓ Ugandan Antiretroviral Treatment and Care Guidelines for Adults, Adolescents, and Children.2nd Edition – January, 2008.

- ✓ UNAIDS, December 2007 Report.
- ✓ UNAIDS. (2006). Report on the Global HIV/AIDS Epidemic. Geneva.
- ✓ UNAIDS. (2009). The regional work shop on adopting minimum standards of practice for theta evaluation team. Participatory evaluation report. Innovation or reawakening? Roles of traditional healers in the management and prevention of HIV/AIDS in Uganda.
- ✓ Xueyan J., Hongzhou L., Yuexin Z., Zengquan Z., Hanhui Y., Qingxia Z., Hui W., Fuxiang W., Wangqian J., Hongqing S., Limin D., Yan S., Yun H., Wubuer M., Aibaidoula Y., Xiaofeng S., Yadong H., Wei T., Boping Z. and Cao Y. (2008). A cross-sectional study of HIV and tuberculosis co infection cases in mainland china. Southern Medical Association, 101(**9**):914-917.

APPENDICES

Appendix 1: Results of univariate analysis using exponential, Weibull, Gompertz, and log-logistic models of HIV infected adult patients under HAART dataset in Zewditu Referral Hospital, AA, 2010.

a) Exponential Univariate Analysis

Covariate/factor Variable	$\widehat{\beta_j}$	s.e	$\widehat{\phi_j}$	Z	p-value	80 % CI for ϕ_j	
						LCL	UCL
Age in years							
<30	Ref.						
[30-39]	-0.4715	0.5026	0.6240466	-0.94	0.348	0.327696	1.188402
[40-49]	1.4300	0.4954	4.178833	-2.89	0.004	0.126823	0.4515325
>=50	-1.7426	0.5055	0.1750674	-3.45	0.001	0.091590	0.3346306
Intercept	7.0130	0.4472	1111	15.68	0.000	626.3373	1970.698
Weight	0.0628	0.0120	1.064815	5.24	0.000	1.048572	1.081309
Intercept	2.7035	0.6174	14.93183	4.38	0.000	6.768544	32.94054
Functional status							
Working	Ref.						
Ambulatory	-1.6091	0.3044	0.200074	-5.29	0.000	0.135455	0.295519
Bed ridden	-2.3841	0.3	0.0921701	-7.95	0.000	0.062751	0.1353823
Intercept	6.82028	0.2	916.2397	34.1	0.000	709.0797	1183.923
TB screen							
No	Ref.						
Yes	-1.7076	0.3780	0.1813024	-4.52	0.000	0.111697	0.2942843
Intercept	7.3439	0.3535	1546.745	20.77	0.000	983.1995	2433.301
OIs							
No	Ref.						
Yes	-0.8496	0.3203	0.4275867	-2.65	0.008	0.283648	0.644569
Intercept	6.7095	0.2887	820.167	23.24	0.000	566.5447	1187.327
CD4 count	0.00785	0.0018	1.007877	4.44	0.000	1.005599	1.01016
Intercept	5.0647	0.2247	158.3367	22.54	0.000	118.7239	211.1665
WHO clinical stage							
WHO stage I	Ref.						
WHO stage II	0.00844	0.7638	1.00848	0.01	0.991	0.37895	2.683832
WHO stage III	-1.637	0.6124	0.1945084	-2.67	0.008	0.08874	0.4263508
WHO stage IV	-2.7226	0.6030	0.0657006	-4.51	0.000	0.03033	0.1422966
Intercept	7.6129	0.5774	2024.334	13.19	0.000	965.931	4242.466
Gender							

Female	Ref.						
Male	0.1610	0.2582	1.174661	0.62	0.533	0.84374	1.635379
Intercept	6.0204	0.1581	411.7498	38.08	0.000	336.227	504.2373
Marital status							
Never married	Ref.						
Married	0.57809	0.4031	1.782628	1.43	0.152	1.06341	2.988283
Others	-0.0026	0.3554	0.9974071	-0.01	0.994	0.63250	1.572834
Intercept	5.9065	0.3162	367.3999	18.68	0.000	244.983	550.988
Educational level							
No education	Ref.						
Primary	0.8123	0.3354	2.253155	2.42	0.015	1.46593	3.463145
Secondary and above	1.0757	0.3134	2.932168	3.43	0.001	1.96230	4.381407
Intercept	5.2788	0.25	196.125	21.12	0.000	142.361	270.1939
Religion							
Muslim	Ref.						
Orthodox	0.1822	0.3155	1.199905	0.58	0.564	0.80085	1.797804
Others	0.2515	0.3318	1.285954	0.76	0.449	0.84049	1.967519
Intercept	5.9199	0.25	372.3748	23.68	0.000	270.295	513.0067

β =coefficient, s.e= standard error, ϕ = acceleration factor, 80% CI=Confidence Interval for acceleration factor, LCL=lower class limit, UCL= upper class limit, Ref=Reference, Z= Score statistic

b) Weibull Univariate Analysis

Covariate/factor Variable	$\widehat{\beta_j}$	s.e	$\widehat{\phi_j}$	Z	p-value	80 % CI for ϕ_j	
						LCL	UCL
Age in years							
<30	Ref.						
[30-39]	-0.5695	0.6034	0.5657878	-0.94	0.345	0.261094	1.226058
[40-49]	-1.6988	0.6229	0.1829115	-2.73	0.006	0.082328	0.4063804
>=50	-2.0735	0.6502	0.1257394	-3.19	0.001	0.054647	0.2893206
Intercept	7.6204	0.6979	2039.31	10.92	0.000	833.8304	4987.57
Weight	0.0738	0.0165	1.076583	4.47	0.000	1.054049	1.099598
Intercept	2.5230	0.7446	12.46636	3.39	0.001	4.801037	32.37016
Functional status							
Working	Ref.						
Ambulatory	-1.7362	0.3768	0.1761969	-4.61	0.000	0.108715	0.2855666
Bed ridden	-2.5624	0.4121	0.0771191	-6.22	0.000	0.045477	0.1307768
Intercept	7.0786	0.4262	1186.285	16.61	0.000	687.052	2048.274
TB screen							
No	Ref.						

Yes	-2.0022	0.5010	0.1350329	-4.00	0.000	0.071053	0.2566248
Intercept	7.9665	0.6385	2882.837	12.48	0.000	1271.811	6534.584
OIs							
No	Ref.						
Yes	-1.0056	0.3994	0.3658075	-2.52	0.012	0.219258	0.6103093
Intercept	7.2625	0.5306	1425.797	13.69	0.000	722.366	2814.219
CD4 count	0.0091	0.0023	1.009097	3.97	0.000	1.006153	1.012051
Intercept	5.2936	0.3222	199.0685	16.43	0.000	131.7209	300.8504
WHO stage							
WHO stage I	Ref.						
WHO stage II	0.01302	0.8746	1.013102	0.01	0.988	0.33027	3.107663
WHO stage III	-1.8634	0.7315	0.1551387	-2.55	0.011	0.06076	0.3961225
WHO stage IV	-3.0918	0.7693	0.0454217	-4.02	0.000	0.01695	0.1217442
Intercept	8.1487	0.8257	3459.049	9.87	0.000	1200.57	9966.094
Gender							
Female	Ref.						
Male	0.1996	0.3126	1.220923	0.64	0.523	0.81790	1.822536
Intercept	6.4642	0.3657	641.7212	17.67	0.000	401.599	1025.417
Marital status							
Never married	Ref.						
Married	0.69012	0.4894	1.993959	1.41	0.159	1.06495	3.733395
Others	0.01122	0.4256	1.011284	0.03	0.979	0.58610	1.744916
Intercept	6.3042	0.4765	546.8847	13.23	0.000	296.948	1007.191
Educational level							
No education	Ref.						
Primary	0.9858	0.4226	2.679865	2.33	0.020	1.55926	4.605835
Secondary and above	1.3038	0.4103	3.683421	3.18	0.001	2.17723	6.231593
Intercept	5.5699	0.3644	262.4323	15.39	0.000	164.521	418.6143
Religion							
Muslim	Ref.						
Orthodox	0.2073	0.3800	1.230346	0.55	0.585	0.75600	2.002303
Others	0.2914	0.4003	1.338308	0.73	0.467	0.80127	2.235284
Intercept	6.3483	0.4284	571.5152	14.82	0.000	330.055	989.6229

β =coefficient, s.e= standard error, ϕ = acceleration factor, 80% CI=Confidence Interval for acceleration factor, LCL=lower class limit, UCL= upper class limit, Ref=Reference, Z= Score statistic.

c) Log-logistic Univariate Analysis

Covariate/factor Variable	$\widehat{\beta_j}$	s.e	$\widehat{\phi_j}$	Z	p-value	80 % CI for ϕ_j	
						LCL	UCL
Age in years							
<30	Ref.						
[30-39]	-0.5667	0.5896	0.5673908	-0.96	0.336	0.266510	1.207954
[40-49]	-1.7156	0.6122	0.1798462	-2.80	0.005	0.082066	0.3941294
>=50	-2.0928	0.6428	0.1233465	-3.26	0.001	0.054114	0.2811517
Intercept	7.4186	0.6682	1666.628	11.10	0.000	707.7982	3924.358
Weight	0.0786	0.0177	1.081781	4.45	0.000	1.057544	1.106573
Intercept	2.0407	0.8346	7.696179	2.44	0.014	2.640715	22.42997
Functional status							
Working	Ref.						
Ambulatory	-1.724	0.3692	0.1783702	-4.67	0.000	0.111131	0.2862909
Bed ridden	-2.6186	0.4068	0.0729047	-6.44	0.000	0.043288	0.1227849
Intercept	6.7992	0.3899	897.123	17.44	0.000	544.2986	1478.655
TB screen							
No	Ref.						
Yes	-2.0110	0.4862	0.1338483	-4.14	0.000	0.071780	0.2495869
Intercept	7.7580	0.6051	2340.294	12.82	0.000	1077.646	5082.36
OIs							
No	Ref.						
Yes	-1.0219	0.3975	0.3598763	-2.57	0.01	0.216221	0.598975
Intercept	7.0952	0.5111	1206.155	13.88	0.000	626.5308	2322.007
CD4 count	0 .0092	0.0023	1.009219	4.03	0.000	1.006281	1.012166
Intercept	5.0721	0.3230	159.5131	15.70	0.000	105.4431	241.3097
WHO clinical stage							
WHO stage I	Ref.						
WHO stage II	0 .0110	0.8239	1.011096	0.01	0.989	0.35176	2.906312
WHO stage III	-1.8061	0.6912	0.1642843	-2.61	0.009	0.06775	0.3983624
WHO stage IV	-3.1024	0.7287	0.0449428	-4.26	0.000	0.01766	0.1143526
Intercept	7.84268	0.7669	2547.032	10.23	0.000	953.222	6805.722
Gender							
Female	Ref.						
Male	0.2152	0.3181	1.240121	0.68	0.499	0.82489	1.864372
Intercept	6.2909	0.3528	539.6399	17.84	0.000	343.413	847.991
Marital status							
Never married	Ref.						
Married	0.7012	0.4988	2.016089	1.14	0.160	1.06380	3.820843

Others	0.0163	0.4388	1.016472	0.04	0.970	0.57928	1.78362
Intercept	6.1288	0.4753	458.893	12.90	0.000	249.572	843.7768
Educational level							
No education	Ref.						
Primary	1.03891	0.4408	2.826141	2.36	0.018	1.60642	4.971989
Secondary and above	1.3587	0.4258	3.890941	3.19	0.001	2.25468	6.714672
Intercept	5.3325	0.3769	206.952	14.5	0.000	127.671	335.4638
Religion							
Muslim	Ref.						
Orthodox	0.2241	0.3892	1.251174	0.58	0.565	0.75979	2.060365
Others	0.2953	0.4091	1.343592	0.72	0.470	0.79536	2.269714
Intercept	6.1735	0.4221	479.879	14.63	0.000	279.403	824.1998

β= coefficient, s.e= standard error, ϕ = acceleration factor, 80% CI=Confidence Interval for acceleration factor, LCL=lower class limit, UCL= upper class limit, Ref=Reference , Z= Score statistic

d) Gompertz Univariate Analysis

Covariate/factor variable	$\widehat{\beta_J}$	s.e	Z	p-value	80 % CI for β	
					LCL	UCL
Age in years						
<30	Ref.					
[30-39]	0.49216	0.5026	0.98	0.328	-0.15201	1.136327
[40-49]	1.4209	0.4954	2.87	0.004	0.786005	2.05588
>=50	1.7379	0.5055	3.44	0.001	1.090064	2.385784
Intercept	-6.3980	0.4755	-13.5	0.000	-7.00741	-5.78867
Weight	-0.0616	0.0121	-5.11	0.000	-0.07711	-0.046185
Intercept	-2.1579	0.6371	-3.39	0.001	-2.97441	-1.341464
Functional status						
Working	Ref.					
Ambulatory	1.55016	0.3052	5.08	0.000	1.159076	1.941243
Bed ridden	2.2661	0.3032	7.47	0.000	1.877525	2.654771
Intercept	-6.3187	0.2707	-23.3	0.000	-6.66567	-5.971788
TB screen						
No	Ref.					
Yes	1.6748	0.3781	4.43	0.000	1.190222	2.159295
Intercept	-6.7184	0.3922	-17.1	0.000	-7.22100	-6.2158
OIs						
No	Ref.					
Yes	0.8272	0.3203	2.58	0.010	0.416652	1.23768

Intercept	-6.0682	0.3334	-18.2	0.000	-6.49546	-5.641032
CD4 count	-0.0075	0.0018	-4.28	0.000	-0.00975	-0.005257
Intercept	-4.5178	0.2698	-16.8	0.000	-4.86351	-4.172031
WHO clinical stage						
WHO stage I	Ref.					
WHO stage II	-0.0165	0.7638	-0.02	0.983	-0.99528	0.9623404
WHO stage III	1.6065	0.6125	2.62	0.009	0.82162	2.391397
WHO stage IV	2.6570	0.6034	4.40	0.000	1.88371	3.430243
Intercept	-7.0251	0.6035	-11.6	0.000	-7.79857	-6.251577
Gender						
Female	Ref.					
Male	-0.1742	0.2582	-0.67	0.500	-0.50515	0.1566906
Intercept	-5.3759	0.2269	-23.7	0.000	-5.66672	-5.085216
Marital status						
Never married	Ref.					
Married	-0.5732	0.4031	-1.42	0.155	-1.08978	-0.056561
Others	-0.0291	0.3556	-0.08	0.935	-0.48471	0.426552
Intercept	-5.2641	03576	-14.7	0.000	-5.72241	-4.805784
Educational level						
No education	Ref.					
Primary	-0.8276	0.3354	-2.47	0.014	-1.2575	-0.3977825
Secondary and above	-1.0908	0.3134	-3.48	0.001	-1.49248	-0.6892008
Intercept	-4.6239	0.2984	-15.5	0.000	-5.00637	-4.241462
Religion						
Muslim	Ref.					
Orthodox	-0.1553	0.3156	-0.49	0.623	-0.55969	0.2491566
Others	-0.2249	0.3319	-0.68	0.498	-0.65026	0.2004508
Intercept	-5.3057	0.2949	-1.80	0.000	-5.68364	-4.927675

β= coefficient, s.e=standard error, 80% CI=Confidence Interval for coefficient, LCL=lower class limit, UCL= upper class limit, Ref=Reference, Z= Score statistic

Appendix 2

Table 1 : Results of the multivariable exponential model using the covariates which are significant at 20% level in the univariate analysis.

Covariates	$\widehat{\beta_J}$	s.e	Z	P>|z|	[95% CI for β_j]
Age in years	0.5396028	0.1312368	4.11	0.000	(0.2823834,0.7968221)
Weight	-0.0378824	0.0117782	-3.22	0.001	(-0.0609673,-0.0147974)
Functional Status	0.4887767	0.1657452	2.95	0.003	(0.163922,0.8136314)
TB screen	1.080774	0.4018695	2.69	0.007	(0.2931241,1.868423)
OIs	0.35586	.351778	1.01	0.312	(-0.3336123,1.045332)
CD4 count	-.0039531	.0017311	-2.28	0.022	(-0.0073461,-0.0005602)
WHO stage	0.6014473	0.1878023	3.20	0.001	(0.2333615,0.969533)
Educational level	-0.4758221	0.171222	-2.78	0.005	(-0.8114111,-0.140233)
Intercept	-6.204191	0.8481785	-7.31	0.000	(-7.86659,-4.541792)

Table 2 : Results of the multivariable exponential regression model after eliminating the variable OIs from the multivariable exponential regression in Table 1.

Covariates	$\widehat{\beta_J}$	s.e	Z	P>|z|	[95% CI β_j]
Age in years	0.5230311	0.1285762	4.07	0.000	(0.2710264,0.7750358)
Weight	-0.0347931	0.011068	-3.14	0.002	(-0.0564858,-0.0131003)
Functional level	0.5075166	0.1635054	3.10	0.002	(0.1870518,0.8279814)
TB screen	1.181743	0.3909354	3.02	0.003	(0.4155241,1.947963)
CD4 count	-0.0036457	0.0016662	-2.19	0.029	(-0.0069113,-0.00038)
WHO stage	0.6241244	0.1844989	3.38	0.001	(0.2625133,0.9857355)
Educational Level	-0.4691963	0.1699104	-2.76	0.006	(-0.8022146,-0.1361779)
Intercept	-6.248464	0.8930747	-7.45	0.000	(-7.89302,-4.603908)

Table 3: Multivariable analysis using exponential, Weibull, Gompertz , and log-logistic models of HIV infected adult patients under HAART dataset in Zewditu Referral Hospital, AA, 2010**.**

a) Results of the multivariable analysis of exponential model using all significance variables in Table 2.

Covariate/factor Variable	$\widehat{\beta_j}$	s.e	$\widehat{\phi_j}$	Z	p-value	95 % CI for ϕ_j	
						LCL	UCL
Age in years							
<30	Ref.						
[30-39]	-0.8458	0.5204	0.429207	-1.63	0.104	0.15477	1.1903
[40-49]	-1.6717	0.5121	0.1879354	-3.26	0.001	0.06888	0.5127
>=50	-1.7965	0.5233	0.1658735	-3.43	0.001	0.05947	0.4626
Weight	0.03509	0.0117	1.035716	3.01	0.003	1.0123	1.0597
Functional status							
Working	Ref.						
Ambulatory	-0.8578	0.3329	0.4241032	-2.58	0.010	0.2208	0.8144
Bed ridden	-1.0534	0.3479	0.3487538	-3.03	0.002	0.1764	0.6897
TB screen							
No	Ref.						
Yes	-1.1198	0.3958	0.3263542	-2.83	0.005	0.15023	0.70895
CD4 count	0.00359	0.0017	1.003601	2.16	0.031	1.00033	1.0069
WHO clinical stage							
WHO stage I	Ref.						
WHO stage II	0.1284	0.7661	1.137032	0.17	0.867	0.25331	5.10369
WHO stage III	-0.7638	0.6325	0.4658699	-1.21	0.227	0.13487	1.60926
WHO stage IV	-1.3631	0.6410	0.2558597	-2.13	0.033	0.07284	0.89872
Educational level							
No education	Ref.						
Primary	0.9737	0.3426	2.647715	2.84	0.004	1.3528	5.1822
Secondary and above	1.0722	0.3237	2.921868	3.31	0.001	1.54938	5.51014
Intercept	5.90910	1.0343	368.373	5.71	0.000	48.5129	2797.17
Log likelihood = -212.1122, AIC= 452.2243							

β= coefficient, s.e= standard error, ϕ= acceleration factor, 95% CI=Confidence Interval for acceleration factor, LCL=lower Class Limit, UCL=Upper Class Limit, Ref=Reference, Z= Score statistic, AIC= Akaike Information Criteria.

b) Results of the multivariable analysis of Weibull model using all significance variables in Table 2.

Covariate/factor Variable	$\widehat{\beta_j}$	s.e	$\widehat{\phi_j}$	Z	p-value	95 % CI for ϕ_j	
						LCL	UCL
Age in years							
<30	Ref.						
[30-39]	-0.8481	0.5282	0.4282428	-1.61	0.108	0.1521	1.2058
[40-49]	-1.6758	0.5343	0.1871657	-3.14	0.002	0.0657	0.5334
>=50	-1.8016	0.5549	0.1650417	-3.25	0.001	0.0556	0.4898
Weight	0.0352	0.0123	1.035821	2.87	0.004	1.0112	1.0610
Functional status							
Working	Ref.						
Ambulatory	0.8598	0.3415	0.4232592	-2.52	0.012	0.2167	0.8266
Bed ridden	-1.0556	0.3586	0.3479512	-2.94	0.003	0.1723	0.7026
TB screen							
No	Ref.						
Yes	-1.1228	0.4123	0.3253434	-2.72	0.006	0.1449	0.729
CD4 count	0.0036	0.0017	1.003611	2.12	0.034	1.0003	1.0070
WHO stage							
WHO stage I	Ref.						
WHO stage II	0.1289	0.7687	1.137547	0.17	0.867	0.2523	5.1316
WHO stage III	-0.7661	0.6395	0.4648246	-1.20	0.231	0.1327	1.6280
WHO stage IV	-1.3671	0.6583	0.2548546	-2.08	0.038	0.0701	0.9261
Education level							
No education	Ref.						
Primary	0.9762	0.3547	2.654221	2.75	0.006	1.3242	5.3198
Secondary and above	1.0752	0.3414	2.9305	3.15	0.002	1.5007	5.7223
Intercept	5.9154	1.0618	370.697	5.57	0.000	46.2613	2970.44
Log likelihood = -212.1118, AIC=454.2235 , ρ=9.997							

β = coefficient, s.e= standard error, ϕ = acceleration factor, 95% CI=Confidence Interval for acceleration factor, LCL=lower class limit, UCL= upper class limit, Ref=Reference, ρ = shape parameter, Z= Score statistic, AIC= Akaike Information Criteria.

c) Results of the multivariable analysis of log-logistic model using all significance variables in Table 2.

Covariate/factor Variable	$\widehat{\beta_j}$	s.e	$\widehat{\phi_j}$	Z	p-value	95 % CI for ϕ_j	
						LCL	UCL
Age in years							
<30	Ref.						
[30-39]	-0.7687	0.4992	0.4636379	-1.54	0.124	0.17427	1.23346
[40-49]	-1.5130	0.5131	0.220243	-2.95	0.003	0.08056	0.60212
>=50	-1.8335	0.5443	0.1598542	-3.37	0.001	0.05501	0.46453
Weight	0.0377	0.0136	1.03842	2.77	0.006	1.01106	1.06647
Functional status							
Working	Ref.						
Ambulatory	-0.7917	0.3333	0.4530713	-2.38	0.018	0.23576	0.87069
Bed ridden	-1.1063	0.3635	0.3307922	-3.04	0.002	0.16222	0.67455
TB screen							
No	Ref.						
Yes	-1.1252	0.3788	0.3245866	-2.97	0.003	0.15446	0.68203
CD4 count	0.0039	0.0016	1.00386	2.36	0.019	1.00065	1.0071
WHO clinical stage							
WHO stage I	Ref.						
WHO stage II	0.2011	0.6886	1.222776	0.29	0.770	0.31706	4.71572
WHO stage III	-0.6246	0.5763	0.5354494	-1.08	0.278	0.17303	1.65698
WHO stage IV	-1.2941	0.6027	0.2741401	-2.15	0.032	0.08412	0.89337
Educational level							
No education	Ref.						
Primary	1.0224	0.3768	2.779878	2.71	0.007	1.32813	5.81849
Secondary and above	1.0649	0.3552	2.900569	3.00	0.003	1.44599	5.81833
Intercept	5.1358	1.0687	170.0018	4.81	0.000	20.9286	1380.91
Log likelihood = -209.47258, AIC=448.9452 , ρ= 1.617352							

β= coefficient, s.e= standard error, ϕ = acceleration factor, 95% CI=Confidence Interval for acceleration factor, LCL=Lower Class Limit, UCL=Upper Class Limit, Ref=Reference, Z=Score statistic, ρ= shape parameter, AIC=Akaike Information Criteria.

d) Results of the multivariable analysis of Gompertz model using all significance variables in Table 2.

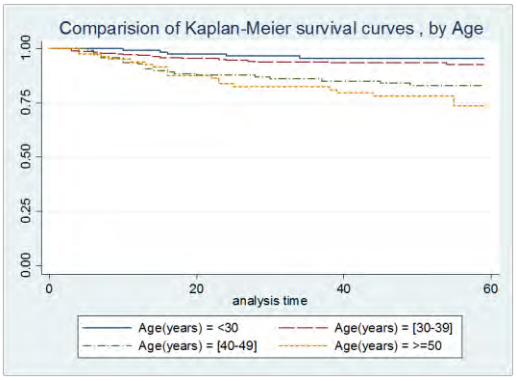
Covariate/factor Variable	$\widehat{\beta_J}$	s.e	Z	95% CI for $\widehat{\beta_J}$		p-value
				LCL	UCL	
Age in years						
<30	Ref.					
[30-39]	0.8291	0.5182045	1.60	-0.186601	1.84472	0.110
[40-49]	1.6057	0.5118372	3.14	0.602539	2.60890	0.002
>=50	1.7593	0.5215815	3.37	0.736969	2.78153	0.001
Weight	-0.0347	0.0117159	-2.96	-0.057637	-0.01171	0.003
Functional status						
Working	Ref.					
Ambulatory	0.8074	0.3333826	2.42	0.153963	1.46080	0.015
Bed ridden	0.9833	0.3480686	2.82	0.301067	1.66547	0.005
TB screen						
No	Ref.					
Yes	1.0918	0.3959365	2.76	0.315807	1.86785	0.006
CD4 count	-0.0035	0.0016576	-2.09	-0.006719	-0.00022	0.036
WHO clinical stage						
WHO stage I	Ref.					
WHO stage II	-0.1344	0.7656386	-0.18	-1.635024	1.36622	0.861
WHO stage III	0.7515	0.6331413	1.19	-0.48945	1.99242	0.235
WHO stage IV	1.3386	0.6420775	2.08	0.080126	2.59702	0.037
Educational level						
No education	Ref.					
Primary	-0.9396	0.3421579	-2.75	-1.61022	-0.26898	0.006
Secondary and above	-1.0532	0.3228747	-3.26	-1.686069	-0.42042	0.001
Intercept	-5.5352	1.054257	-5.25	-7.601514	-3.46890	0.000
Log likelihood = -210.4801 , AIC = 450.9601						

β= coefficient, s.e= standard error, 95% CI=Confidence Interval for β, LCL=Lower Class limit, UCL= Upper Class Limit, Ref=Reference, Z= Score statistic, AIC= Akaike Information Criteria.

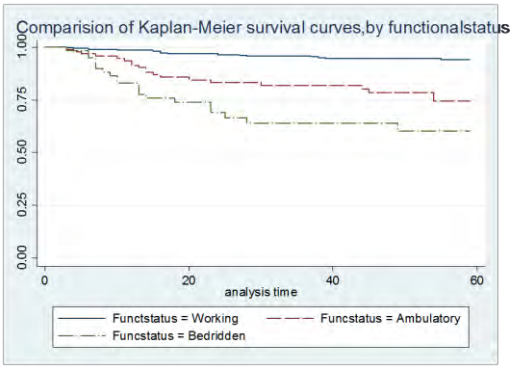
Appendix 3

Figure 5 : Plots of Kaplan-Meier survivor functions based on different factors, of HIV infected adult patients under HAART dataset in Zewditu Referral Hospital, AA, 2010.

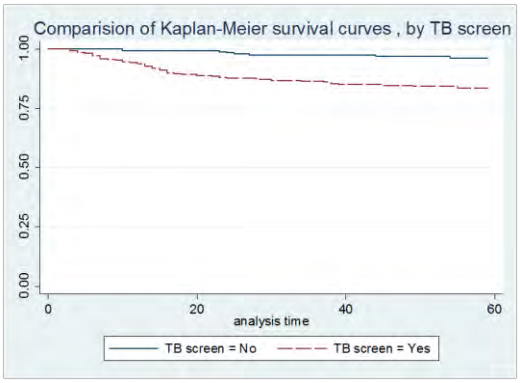
a)



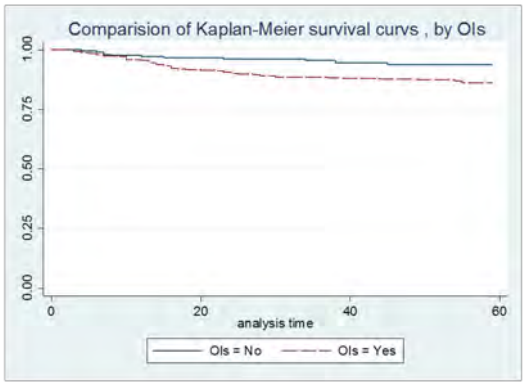
b)



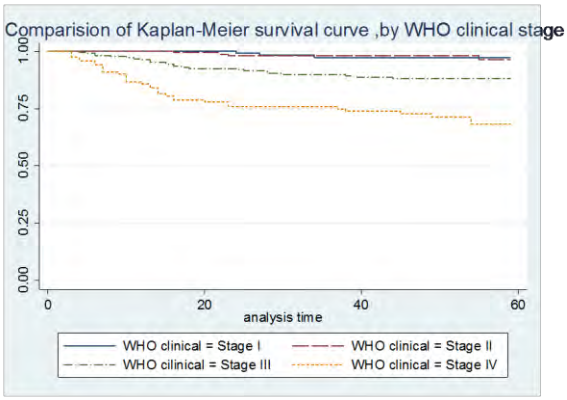
c)



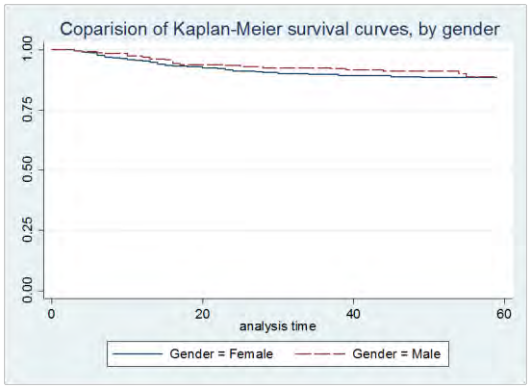
d)



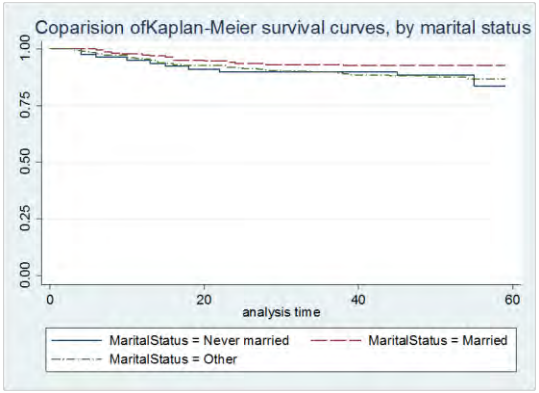
e)


Comparision of Kaplan-Meier survival curve ,by WHO clinical stage
0.00
0.25
0.50
0.75
1.00
0
20
40
60
analysis time
WHO clinical = Stage I
WHO clinical = Stage II
WHO cilinical = Stage III
WHO clinical = Stage IV

f)


Coparision of Kaplan-Meier survival curves, by gender
0.00
0.25
0.50
0.75
1.00
0
20
40
60
analysis time
Gender = Female
Gender = Male

g)


Coparision ofKaplan-Meier survival curves, by marital status
0.00
0.25
0.50
0.75
1.00
0
20
40
60
analysis time
MaritalStatus = Never married
MaritalStatus = Married
MaritalStatus = Other

h)

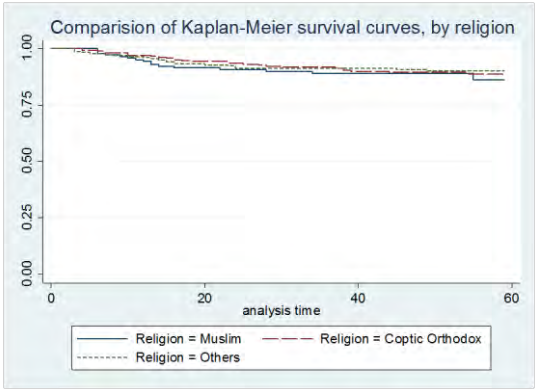
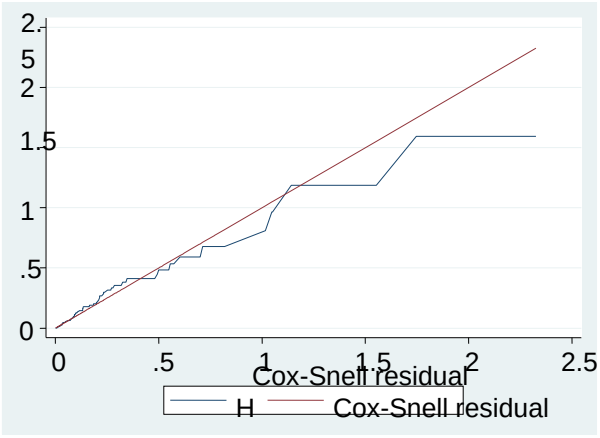
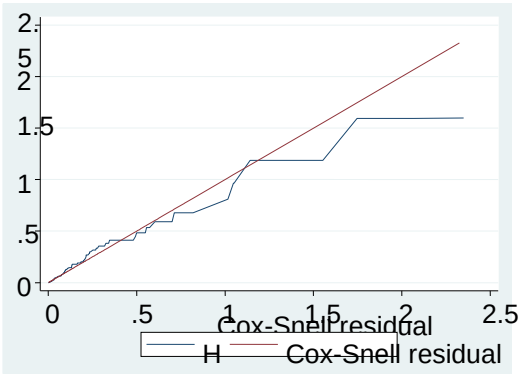

Comparision of Kaplan-Meier survival curves, by religion
0.00
0.25
0.50
0.75
1.00
0
20
40
60
analysis time
Religion = Muslim
Religion = Coptic Orthodox
Religion = Others

Figure 6 : Cox-Snell residuals obtained by fitting exponential, Weibull, Gompertz and log-logistic models for HIV infected adult patients under HAART dataset in Zewditu Referral hospital, AA, 2010.

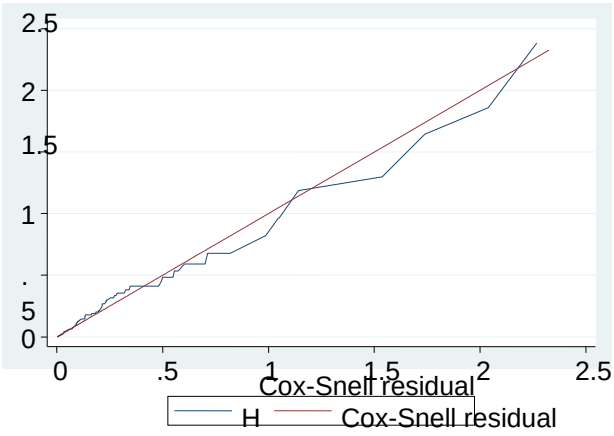
a) Cox- Snell for exponential



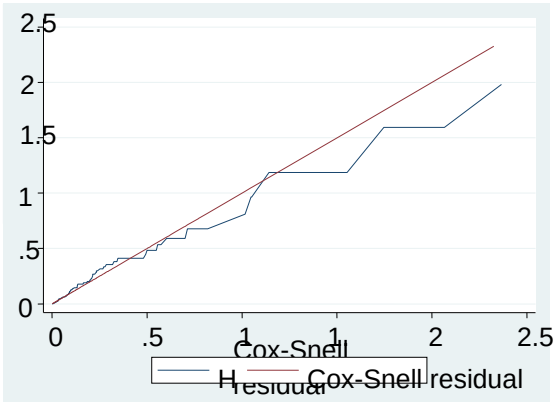
b) Cox- Snell for Weibull



c) Cox- Snell for log-logistic



d) Cox- Snell for Gompertz



GLOSSARY

Acquired Immunodeficiency Syndrome (AIDS)	The most severe manifestation of infection with HIV. There are many opportunistic infections and Cancers that constitute an AIDS diagnosis in the presence of HIV infection.
Antiretroviral therapy (ART)	A class of drugs which inhibit the activity of retroviruses such as HIV. It consists of the combination of at least three antiretroviral (ARV) drugs to maximally suppress the HIV virus and stop the progression of HIV disease.
Human immunodeficiency virus (HIV)	The retrovirus isolated and recognized as the etiologic (i.e., causing or contributing to the cause of the disease) agent of AIDS.
Highly Active Antiretroviral Therapy (HAART)	Regimens typically include a combination of at least three drugs, such as different association of protease inhibitors (PI), non-nucleoside reverse transcriptase inhibitors (NNRTI) and nucleoside reverse transcriptase inhibitors (NRTI).
WHO clinical stages of AIDS	Classification of the stages of HIV-associated clinical disease where stage1 indicates asymptomatic disease, stage 2 indicates mild disease, stage 3 indicates advanced disease and stage4 indicates severe disease.
Ambulatory	An individual able to perform activities for daily living.
Bedridden	An individual unable to perform activities of daily living.
Working	An individual able to perform usual work in and out of the house, harvest, go to school for children, normal activities or playing.
Viral Load Test	Refers to the amount of HIV in your blood. The results of these tests tell you whether your viral load is low, medium or high.
CD4 cell	A type of white blood cell that fights infection. Another name for them is T-helper cells.
CD4 Count	Measures the number of CD4 cells in a sample of your blood drawn by a needle from a vein in your arm. CD4 count tells how strong your immune system is.

DECLARATION

I, the undersigned, declare that this thesis is my original work, has not been presented for degrees in any other University and all source materials used for the thesis have been duly acknowledged.

Name: Haftu Legesse

Signature: …….

Date: ……

This thesis has been submitted for examination with my approval as a University advisor.

Name: Professor M.K Sharma

Signature: ……..

Date: ……..

Place: College of Natural Science, Addis Ababa University