Modelling of AIDs Disease Associated Risk Factors and Mortality in Gamo Gofa Zone, Ethiopia

Derbachew Asfaw Teni and Tilahun Ferede Asena

Department of Statistics, Arba Minch University, Arba Minch,
P.O.Box: 21, Arba Minch, Ethiopia

Abstract: Acquired Immune Deficiency Syndrome (AIDs) which is believed to be caused by the Human Immunodeficiency Virus (HIV) has been the major problem worldwide. The target population for this study was recorded patients of laboratory confirmed HIV/AIDs patients at Arba Minch Referral Hospital. ASRS method was adopted for selecting a representative sample of the patients based on their ART unique identification number/card for a cohort of patients from 1997 - 2005. Hence, the sample size with $N = 2447$ was $n = 136$. This study was aimed to estimate time to death of HIV/AIDs patients and determine significant risk factors for the survival times at Arba Minch Referral Hospital, Gamo Gofa Zone, Ethiopia. The findings based on semi parametric survival regression model on the data of AIDs patients reviled that the major risk factors determining the survival time are initial weight, past opportunistic infection, DM status, CD4 count, WHO clinical stage and adherence level. Butthe covariate sex, age, TB status and regimen were not significant factors. This study recommends Stakeholders, public health policy makers, researchers and the public at large, brings HIV/AIDs diseases and other chronic diseases to their agenda, so that appropriate prevention and control strategies are implemented along with a population wide surveillance intervention. Health care workers should anticipate and inform patients about the possible related risk factors of death through early diagnosis and appropriate intervention. Donors and government should understand the risk factors that influence the death of AIDs patients.

Key words: Ethiopia · Disease · Gamo Gofa Zone · HIV/AIDs · Mortality · Risk Factors

INTRODUCTION

Global Situation of HIV/AIDs: Acquired Immune Deficiency Syndrome (AIDs) which is believed to be caused by the Human Immunodeficiency Virus (HIV) has been the major problem worldwide. The rate of spread of the HIV/AIDs epidemic has reached a shocking level. The expansion of the epidemic has now become a burning issue globally and this is particularly so more important in developing countries. The disease being one without any cure is still accountable for economic, social and health crises in many developing countries. Its high prevalence and distribution among the youth made things even more complicated.

Since the beginning of the epidemic, almost 78 million people have been infected with the HIV virus and about 39 million people have died of HIV. Globally, 35.0 million [33.2–37.2 million] people were living with HIV at the end of 2013. An estimated 0.8% of adults aged 15–49 years worldwide are living with HIV, although the burden of the epidemic continues to vary considerably between countries and regions. More than two-thirds (70%) of all people living with HIV, 24.7 million, live in sub-Saharan Africa—including 91% of the world's HIV-positive children. In 2013, an estimated 1.5 million people in the region became newly infected. An estimated 1.1 million adults and children died of AIDs, accounting for 73% of the world's AIDs deaths in 2013 (WHO, 2013). TB is the leading cause of death among PLWHIV in Africa and a major cause of death elsewhere, accounted for almost 2 million deaths per year globally. It is also the most common cause of illness among people living with HIV/AIDs where HIV/AIDs has had its greatest adverse impact.

Corresponding Author: Derbachew Asfaw Teni, Department of Statistics, Arba Minch University, Arba Minch,
P.o.Box 21, Ethiopia. Tel: +251946555818.

HIV/AIDs in Ethiopia: Ethiopia is a large, predominantly rural, country covering an area of 1.13 million sq km. The country is a Federal Democratic Republic, with a decentralized administrative structure, composed of nine regional states and two city administrations. Ethiopia is currently the second most populous country in Africa (next to Nigeria) with a total population of about 79 million, growing at an annual rate of 2.6%, according to the 2007 national census [1].

The first time that HIV observed in Ethiopia was in 1984 and the first two AIDs cases were reported in 1986. Then, National AIDs Control Program (NACP) was established at the department level at the MOH in 1987 [1]. In 2005, the estimated national HIV prevalence was 3.5% (10.5% for urban and 1.9% for rural areas) according to Federal Ministry of Health report [1]. In 2010, in Ethiopia, there are 1.2 million people living with HIV (PLHIV), with an adult HIV prevalence of 2.4% (7.7% urban and 0.9% rural) and male-female ratio of 1.9%. A total of 397,818 people living with HIV are estimated to be in need of antiretroviral treatment (ART) and an estimated 137,494 new HIV infections are expected to join in the given year. The HIV/AIDs epidemic in Ethiopia is generalized with significant heterogeneity between regional states and population groups as [2].

According to Ministry of Health (FMOH, 2010), the three highest prevalence regions in the country in 2010 are Gambela (8.3% urban), Addis Ababa (7.5%) and Dire Dawa (4.2%). Other regions with HIV prevalence rates greater than the national estimate (2.1%) are Harari (3.2%), Amhara (2.7%) and Tigray (2.7%). Somali is the region with the lowest HIV prevalence estimate in the country (0.8%). By the year 2007, in UNAIDs report of [3], the number of PLWHIV requiring ART was 258,264. Of these, 242,548 were adults (24.8% of those 15 years and older living with HIV) and 15,716 were children (24.3% of HIV positive children in the age of 0 to 14 years).

Almost 27 years have now elapsed since the virus was first reported [4]. During these years HIV infection has changed from a fatal condition to a manageable chronic illness mainly due to the development of antiretroviral therapy (ART). The goal this therapy is to improve survival; to reduce HIV associated morbidity and mortality, to increase the quality of life, to restore immune function and to achieve maximal and sustained suppression of viral replication [4]. By 2010 WHO has planned to put 9.8 million people on ART with the goal of providing universal access to HIV care and ART [3].

Most of the studies in Ethiopia focused on the prevention [5] and factors that increase the chance of contracting the disease, mainly dealing prevention before a person is HIV positive and few dealing factors that influence the death status [6].

As reported in [7], we assume that an infected patient passes through the following immunological states related to CD4+ lymphocytes counts: state I (CD4>500 x 10^6cells/L), state II (350<CD4 = 500), state III (200<CD4 = 350), state IV (CD4 = 200). Moreover, we added an absorbing state (the patient death, denoted by D).

Acquired Immune Deficiency Syndrome (AIDs): AIDs is a chronic, life-threatening disease caused by the human immunodeficiency virus (HIV) that damage the immune system of human being. HIV interferes with the body's ability to fight off viruses, bacteria and fungi that cause disease and makes the infected person more susceptible to certain types of cancers and infections that the body would normally resist. The virus and the infection itself are known as Human Immune Virus (HIV). Acquired immunodeficiency syndrome (AIDs) is the name given to the later stages of an HIV infection [3].

General Overview of ART: Antiretroviral medications are designed to inhibit the reproduction of HIV in the body. The main effect of antiretroviral treatment is to suppress viral replication, allowing the individual's immune system to recover and protect him/her from the development of AIDs and death. The clinical benefit of ART for AIDs patients, in terms of mortality reduction and improved quality of life, is well established but shows regional variations, with higher case fatality rates in poor countries [8]. With the advent of antiretroviral therapy (ART), the morbidity and mortality rates of HIV infection are decreasing dramatically in Europe and the USA ([9]; [10]). However, according to the White House Office of National AIDs Policy (1999), in some of Africa and Asian countries, HIV morbidity and mortality rates are increasing due to the ineffective implementation of prevention and intervention policies.

In 2009 alone, 1.2 million people received HIV antiretroviral therapy for the first time, an increase in the number of people receiving treatment of 30% in a single year. Overall, the number of people receiving therapy has grown 13 fold, more than five million people in low and middle income countries, since 2004. Expanding access to treatment has contributed to a 19% decline in deaths among people living with HIV between 2004 and 2009.

This is just the beginning, as 10 million people living with HIV and who are eligible for treatment under the new WHO guidelines are still in need [11].

Major Factors for Survival Time of AIDs Patients: [12] and [13] examined the survival probability of AIDs patients using socio-demographic factors. According to their report age is the significant predicator of survival of HIV/AIDs patients. [11] from [14] from Korea indicated that the covariate diagnosis, age and sex are significant predicator for HIV/AIDs progression. According to their report younger patients have the advantage of surviving longer than older patients and male patients were dying at the rate which was 2 times more than female patients.

Opportunistic infections (OIs) are common causes of death in HIV infected patients. Antiretroviral therapy (ART) has reduced the incidence of opportunistic infections for certain patients with access to care. However, opportunistic infections may continue to cause substantial morbidity and mortality in patients with HIV infection (Holmes *et al*., 2003). Another explanation could be that, as an AIDs opportunistic infection, pulmonary tuberculosis affects 30.9% of the diagnosed cases and it is characterized as an AIDs defining disease. Disseminated tuberculosis associated with pulmonary tuberculosis occurs in 6.4% of the cases. Diagnostic procedures and treatment delay and the often concomitant occurrence of tuberculosis could explain the greater impact on survival from the second year onwards ([15]; [16]).

[13] Also showed that functional status had an impact on the survival of AIDs patients after conducting a study on 168 patients with AIDs in Rome and the result implied that lower ability to perform self-sufficiency in Activities of Daily Living (ADLs) were related to shorter survival by Cox regression. [17] also made a study on 305 persons with AIDs in Boston by reviewing their medical records and ascertaining the vital status of patients. The result demonstrated that measures of activities of daily living, functional status, had an impact to predict the survival of people with AIDs. Casalino *et al*., (1998) used cox regression and showed that short and long term survival are strongly associated with the preadmission health status, functional status and weight loss of people with HIV after examining 421 patients in two years study period.

A retrospective study which was done in Adama Hospital ART clinic with 259 HIV/AIDs patients [6] showed that condom use, alcohol, baseline CD4,

baseline weight, drugs, number of rooms and WHO stage were predicators of survival of HIV/AIDs patients. The comparable study done in Addis Abeba Tikur Ambesa hospital revealed that gender, age, functional status, ART, CD4 and weight were statistically significantly related to survival of HIV/AIDs patients.

This study describes the mathematical model, with a focus on Kaplan-Meier Curves, Cox (Proportional Hazards) regression,and Parametric Survival regression models, in predicting the time to death of AIDs patients and evaluating the association of the factors with survival/death status respectively.

Description of the Problem: HIV/AIDs is the major world's pandemics, with most of its effects felt in sub-Saharan Africa. In trying to control this pandemic, efforts have been largely aimed at prevention with little attention given to care. The ever increasing numbers infected with the diseases, makes it imperative (not to be avoided) for health scientists and researchers to redefine their position and goals in combating this disease. WHO has stirred up many developing countries to make ART available to their HIV infected patients. This development raises questions about when it is appropriate to start ARV therapy in those dually infected patients with HIV/AIDs, the difference in effect of ARV therapy between those only infected with HIV and those dually infected with HIV is also unclear and requires careful research.

The survival time for AIDs patients depends on socioeconomic, demographic and clinical factors. Therefore, this study is aggravated to identify the major risk factors associated with survival of AIDs patients which are also common to a number of other communicable diseases. Research Question:

• What are the significant risk factors for the survival of AIDs patients? And how long is the mean survival time probability?
• Which groups have high hazard of death among various levels of risk factors?
• Which parametric survival models are best fit for the anticipated data on survival time of HIV/AIDs patients?

General Objective: This study is anticipated to estimate time to death of AIDs patients and determine significant risk factors for the survival times of HIV/AIDs patients at Arba Minch Referral Hospital, Gamo Gofa Zone, Ethiopia: Specifically:

- To estimate and compare survival probability of AIDs patients who are on ART.
- To identify determining significant risk factors for HIV/AIDs.
- To fit a Cox proportional hazard and parametric regression models for the data of HIV/AIDs patients who are under follow up.
- To provide quantitative information from this study to researchers, to the society and any stakeholders.

MATERIALS AND METHODS

Sampling Design: The target populations for this study were recorded patients of laboratory confirmed HIV/AIDs patients at Arba Minch Referral Hospital. The study is a retrospective study that is all the events - exposure had already occurred in the past, which reviews the patient cards and patient's information sheet. The researcher merely collects the data and investigate the risk factors associated with the survival of patients with AIDs diseases. A simple random sampling method was adopted for selecting a representative sample of the patients based on their ART unique identification number/card for a cohort of patients from 1997 - 2005. Hence, the sample size with $N = 1244$ 1244 was n=36.

Variables in the Study
Dependent Variable: Survival time of the HIV/AIDs patients measured in months is considered as a response variable.

Independent Variables or Covariates: Age of patient, sex of patients (Female and Male), chew chat, initial weight of patients, CD4 cell count of patients, TB status of patient (positive and negative), WHO clinical stage at start of the treatment (stage I, stage II, stage III and stage IV),Regimen, Adherence level, patient's status (death, censored, transferred out and survived).

Methods of Statistical Data Analysis
Non Parametric Survival Models: Survival analysis is defined as a branch of statistics which deals with data related to time to an event. This topic is also called reliability analysis in engineering and duration analysis in economics or sociology.

The term survival analysis applies to techniques in which the data being analyzed represent the time it takes for a certain event to occur. The use of survival analysis, as opposed to the use of different statistical methods, is most important when there is no time-to-event record. In reality such situation can occur due to the following reasons: when an individual survive beyond the study period or the individual does not experience the event, lost to follow-up (i.e. drop out, transfer to other place) and deaths due to other causes different from that/those specified in the study. Therefore, survival data are almost always incomplete. The statistical terminology for such data is censoring. Censoring is common in survival analysis and it is considered as an important feature of survival data. Survival analysis is well suited to such data which are very common in medical research since studies in medical areas have a special feature that follow-up studies could start at a certain observation time and could end before all experimental units had experienced an event. The most common encountered form of a censored observation illustrated in Figure 1 below. An 'X' indicates that the subject has experienced the outcome of interest; a 'O' indicates censoring. Subject A experiences the event of interest on day 7. Subject B does not experience the event during the study period and is right censored on day 12. Subject C does not experience the event of interest during its period of observation and is censored on day 10. Subject D is interval censored. Subject E is left censored: it has been found to have already experienced the event of interest when it enters the study on day 1. Subject F is interval truncated: there is no way possible that the event of interest could occur to this individual between days 4–6. Subject G is left truncated: there is no way possible that the event of interest could have occurred before the subject enters the study on day 3.

Kaplan-Meier Estimator: The Kaplan-Meier estimator, or product limit estimator, is the estimator used by most software packages because of the unsophisticated step approach. It incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. The KM estimator consists of the product of a number of conditional probabilities resulting in an estimated survival function in the form of a step function. The Kaplan-Meier estimator of the survivorship function (or survival probability) $S(t) = P(T \geq t)$ is defined as:

$$\hat{S}(t) = \prod_{t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{t_j < t} \left(1 - \frac{d_j}{n_j} \right) \tag{1}$$
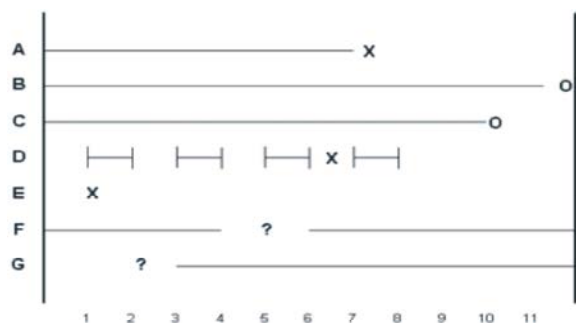
Fig. 1: Left-censoring, right-censoring and truncation.

With the convention that $\hat{s}(t) = 1$ *if* $t < t_{(i)}$ Where d_j is the number of individuals who experience the event at time t_j and n_j is the number of individuals who have not yet experienced the event at that time.

Comparison of Survivorship Functions

The Cochran-Mantel-Haenszel Log Rank test: The log rank test, sometimes called the Cox-Mantel test, is the most well-known and widely used test statistic. This test is based on weights equal to one, i.e. $w_i = 1$. Therefore, the log rank test statistic becomes:

$$Q_{LR} = \frac{\left[\sum_{i=1}^{m} n_i \left(d_{li} - \hat{e}_{li}\right)\right]^2}{\sum_{i=1}^{m} \hat{v}_{li}} \qquad (2)$$

The Generalized Wilcoxon test: generalized the Wilcoxon rank sum test to allow for censored data. This test uses weights equal to the number of subjects at risk at each survival time, i.e. $w_i = n_i$ and is called Wilcoxon or generalized Wilcoxon test in most software packages. Thus the Wilcoxon test can be defined as:

$$Q_{GWt} = \frac{\left[\sum_{i=1}^{m} n_i \left(d_{li} - \hat{e}_{li}\right)\right]^2}{\sum_{i=1}^{m} n_i^2 \hat{V}_{li}} \qquad (3)$$

Semi Parametric Survival Regression Model

The Cox Proportional Hazards Model: It is usually written in terms of the hazard model formula. This model gives an expression for the hazard at time t for an individual

with a given specification of a set of explanatory variables denoted by X and it is generally given by:

$$h(t, X_i, \beta) = h_0(t) \exp(\beta' X_i) \qquad (4)$$

where $h_0(t)$ is the baseline hazard function that characterizes how the hazard function changes as a function of survival time, X_i is the vector of values of the explanatory variables for the ith individual at time t and β is the vector of unknown regression parameters that are assumed to be the same for all individuals in the study, which measures the influence of the covariate on the survival experience.

A smart property of the Cox model is that, even though the baseline hazard part of the model is vague, it is still possible to estimate the $\beta' s$ in the exponential part of the model. So, it can equally be regarded as linear model, as a linear combination of the covariates for the logarithm transformation of the hazard ratio given by:

$$\log\left\{\frac{h(t, x, \beta)}{h_o(t)}\right\} = \beta' X$$

The cumulative hazard function is given by: $H(t) = H_o(t) \exp(\beta' X)$ and the survivorship function is: Where $S(t, X, \beta) = S_o(t) \exp(\beta' X)$ is a baseline survival function.

Fitting the Proportional Hazard Model: The Maximum Likelihood estimates of the Cox model parameters are derived by maximizing a likelihood function usually denoted as L. The likelihood function is a mathematical expression which describes the joint probability of obtaining the data actually observed on the subjects in the study as a function of the unknown parameters (the $\beta' s$) in the model being considered.

The partial likelihood can be written as the product of several likelihoods, one for each of, say k failure times. Thus, at the jth failure time, lj denotes the likelihood of failing at this time, given survival up to this time. Note that the set of individuals at risk at the ith failure time is called the "risk set," $R\left(t_{(j)}\right)$ and this set may change actually get smaller in size as the failure time increases.

$l(\beta) = \prod_{j=1}^{k} L_j$, here L_j is the ith failure time given the risk set $R\left(t_{(j)}\right)$.

In a very general sense, the partial likelihood is given by the expression

$$L_p(\beta) = \prod_{i=1}^{m} \left[\frac{e^{x_i\beta}}{\sum_{j \in R(t_i)} e^{x_i\beta}} \right]^{c_i} \quad (5)$$

where the summation in the denominator is over all subjects in the risk set at time t_i denoted by $R(t_i)$, the product is over the m distinct ordered survival time and $x_{(i)}$ denoted the value of the covariance for the subject with ordered survival time $t_{(i)}$ and c_i is an indicator of tied survival times. Once the likelihood function is formed for a given model, the next step for the computer is to maximize this function. This is generally done by maximizing the natural log of L, which is computationally easier. The log partial likely function is given by (assuming $c_i = 0$):

$$L_p(\beta) = \sum_{i=1}^{m} \left\{ X_{(i)}\beta - \ln \left[\sum_{i \in R(t_{(j)})} e^{x_j\beta} \right] \right\} \quad (6)$$

The maximization process is carried out by taking partial derivatives of log of L with respect to each parameter in the model and then solving a system of equations as shown here. This solution is carried out using iteration.

Parametric Regression Modeling: In previous topics it was focused entirely on the use of non-parametric model and proportional hazards Cox regression model, in the

analysis and prediction of the survival time of patients with HIV/AIDs. The basis of this method was to avoid having to specify the hazard function completely. However, there may be settings in which the distribution of the survival time is in specific parametric distribution that justifies the use of a fully parametric model to better address the goal of the analysis in accordance with [21].

Weibull Regression Model: Survival time is a positive random variable with Weibull probability density function can be expressed as:

$$f(t;\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \exp\left(\left(-\frac{t}{\mu}\right)^{\alpha}\right) \quad (7)$$

where, $\mu > 0$ and $\alpha > 1$ and the baseline hazard function of the distribution becomes:

$$h(t;\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \quad (8)$$

This yield the following survivorship functions: $S(t) = \exp\left(-\left(\frac{t}{\mu}\right)^{\alpha}\right)$ and the cumulative hazards function becomes: $H(t) = \left(\frac{t}{\mu}\right)^{\alpha}$

Depending on the value of α the hazard function can increase or decrease with increasing survival time. Hence the Weibull model can yield an accelerated failure time model. Independent observations (t_i, δ_i), $i = 1,...,n$ with survival time t_i and censoring indicator δ_i which has value of one if i^{th} observation is not censored and zero when the ith observation is censored and Let β be the unknown parameter. The likelihood function is

$$L(\beta) = \prod_{i=1}^{n} \left\{ f(t_i)^{\delta_i} (s(t_i))^{1-\delta_i} \right\} = \prod_{i=1}^{n} \left\{ \left(\frac{f(t_i)}{s(t_i)}\right)^{\delta_i} s(t_i) \right\} = \prod_{i=1}^{n} \left\{ \{h(t_i)\}^{\delta_i} s(t_i) \right\}$$

$$\prod_{i=1}^{n} \left\{ \left(\frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1}\right)^{\delta_i} \exp\left[-\left(\frac{t}{\mu}\right)^{\alpha}\right] \right\} \quad (9)$$

Reparametrizing the Weibull distribution using $\lambda = \mu^{-\alpha}$ then $h_o(t)$ $\lambda = \mu^{-\alpha}$ will be the baseline hazard function. Now incorporate covariates X in the hazard function, the Weibull regression models become:

$$h(t;X,\beta) = \lambda\alpha t^{\alpha-1} \exp(X\beta) \quad (10)$$

The model assumes that individual i and j with covariates x_i and x_j have proportional hazard function of the form:

$$\frac{h(t;x_i)}{h(t;x_j)} = \frac{\exp(x_i\beta)}{\exp(x_j\beta)} = \exp\left[(x_i - x_j)'\beta\right]$$

A different parameterization is used with intercept v and scale parameter σ and covariate effects γ_i having relationship with original parameterization as $\beta_j = \frac{-\gamma_j}{\sigma}, \alpha = \sigma^{-1}$ and $y = \ln(t)$.

The Exponential Regression Model: For the time data and skewed to the right, with distribution of the time is exponential, the time of survival for a set of covariates X, which is called, accelerated failure time, expressed as:

$$T = \exp\left(\beta'X + \varepsilon\right) \qquad (11)$$

where, ε is error component.

The exponential model $t \sim Exp(\alpha)$ is the simplest parametric model and assumes a constant risk or hazard over time, which reflects the property of the distribution appropriately called 'lack of memory' because the hazard function $h(t) = a$ does not depend on time. Hence the probability of failure in a time interval $[t, t + \delta t]$ does not depend on previous interval.

The survivorship function may be obtained by expressing in terms of time as: $S(t, X, \beta) = \exp\left(-te^{-\beta'X}\right)$ and the hazard function of the exponential regression model is $h(t, X, \beta) = e^{-(\beta'X)}$. For the exponential regression survival models the hazard ratio for the dichotomous covariate is $HR(x = 1, x = 0) = e^{\beta_1}$.

The Log-Logistic Regression Model: Multiple covariate log-logistic accelerated failure time may be expressed as:

$$\ln(t) = \beta'X + \sigma\varepsilon \qquad (12)$$

where σ is the scale parameter and ε is the residual (unexplained) variation in the transformed survival times.

The survivorship function for the model (12) is $S(t, X, \beta, \sigma) = (1 + \exp(z))^{-1}$ Where z is the standardized log-time outcome variable, that is; $z = \frac{(y - \beta_o - \beta_i X)}{\sigma}$ and $y = \ln(t)$.

The odds of a survival time of at least t are, OR$= \frac{s(t, x, \beta, \sigma)}{1 - s(t, x, \beta, \sigma)} = \exp(-z)$, assumes that the covariate is

dichotomous and coded 0 or 1. The odds- ratio at time t from the ratio the odds of a survival time evaluated at x= 0 and x= 1 is:

$$OR(x = 1, x = 0) = \frac{\exp\left[\frac{-(y - \beta_0 - \beta_1 x_1)}{\sigma}\right]}{\exp\left(\frac{-(y - \beta_0 - \beta_1 x_1)}{\sigma}\right)} = \exp\left(\frac{\beta_1}{\sigma}\right) \quad (12)$$

This is independent of time.

The Lognormal Regression Model: The log-normal model may take censored time dependent variable that allows the hazard rate to increase and decrease. The log-normal model assumes that $\varepsilon \sim N(0, 1)$. Let $h(t)$ be the hazard function of T for (12) when $\beta = 0$.e. $\beta_o^* = \beta_1 = \ldots = \beta_p$. Then, it can be shown that $h(t)$ has the following functional form:

$$h(t) = \frac{\phi\left(\frac{\log(t)}{\sigma}\right)}{\left[1 - \phi\left(\frac{\log(t)}{\sigma}\right)\right]\sigma t} \qquad (13)$$

where, $\phi\left(\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-t^2}{2}\right)\right)$ is the probability density function and $\phi(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-t^2}{2}\right)du$ is the cumulative distribution function of the standard normal distribution.

Obviously we no longer have a proportional hazards model. If the baseline hazard function is desired, it can be obtained from equation (13) by setting x = 0. The survival function $s(t/X)$ at any covariate x can be expressed as:

$$s(t/X) = \phi\left[\beta_o^* + \beta_1^* x_1 + \ldots + \beta_p^* x_p - \alpha\log(t)\right] \qquad (14)$$

Where, $\alpha = \frac{1}{\sigma}, \beta_j^* = \frac{\beta_j}{\sigma}$ for $j = 1, 2, 3, \ldots, p$. This is the final survival model with intercept depending with t.

Model Selection: [22] proposed an information criterion (AIC) statistic to compare different models and/or models with different numbers of parameters. For each model the value is computed as:

$$AIC = -2\log(likelihood) + 2(p + 1 + s) \qquad (15)$$

where p denotes the number of covariates in the model without including the constant term and s is the number of parameters minus one i.e. $s = 0$ for the exponential regression model and $s = 1$ for Weibull, log logistic and lognormal regression models. According to the criterion, a model with small AIC value will be considered as it fits for the data.

RESULTS AND DISCUSSION

Nonparametric Survival Models: In this study, a sample of 131 HIV/AIDs patients are considered, out of which 71(54.2%) are females and 60(45.8%) are males. Among these 22(16.8)% patents are died and the remaining 109 (83.2)% are censored. A death proportion of females which is 14.1% seem lower than males 20%.Patients who have past opportunistic infection have high proportion of death, 35.4%, than those who have no any disease before. The death proportions of patients who have low and moderate adherence are respectively 29.3% and 6.9%. But, no event is observed for high adherence patients. The mean survival time is 53.93 months and the median survival time is 68 months. The results together with its 95% confidence intervals are displayed in Table 1 below. The variance of the mean is based on the [21] estimator of the variance of the survival distribution.

Comparison of Survival Experience: When comparing levels (subgroups) of covariate, it is always a good idea to begin with a graphical display of the data in each level. Therefore, we should graph the Kaplan-Meier estimator of the survivorship function for each of the levels.

Female patients lived longer than male patients as shown in Figure 4. A; but the log-rank test suggests that there is no significance difference in survivorship functions between male and female patients. The value of the log rank test for covariate Chew chat in Table 2 and KM-estimate in Figure 4.B indicates that there is significant difference in survivorship function between levels of chat chewing status (p<0.000726). The estimated survivorship function for patients with low adhere, moderate adhere and high adhere, significantly different. This suggests, patients who were high adherent might have a more favorable survival experience than others. The result also confirmed by Log-rank test in Table 2 below. Comparing the survivor functions between different WHO stages of HIV/AIDs patients, Kaplan-Meier survivor estimates for the four AIDs stages are plotted in Figure 4.F. This Figure shows that patients at stage I and stage II had slightly higher survival compared with stage III patients. Statistical test is made by using log-rank (mantel-Cox) test in Table 2 and this shows that there is significant difference between patient's who's WHO stage was I, II, III and IV with respect to survival time.
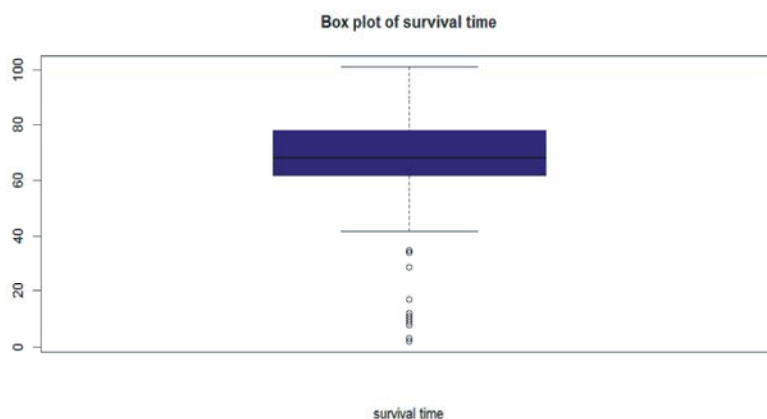

Fig. 2: Box plot of Survival Time of HIV/AIDS Patients

Table 1: Survival Status and Mean Survival Time for HIV/AIDs Patients

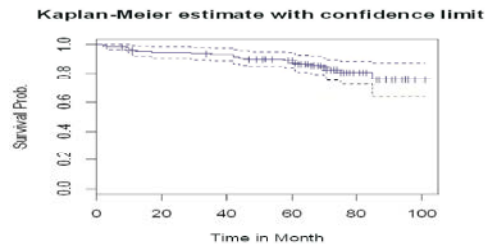| | | | | | | 95% C.I | |
						---	---
Death	Censored	Total	Mean of survival time	Median	Std.Error	Lower	upper
22(16.8)%	109(83.2)%	131	53.93	68	20.33116	9.25	95.00

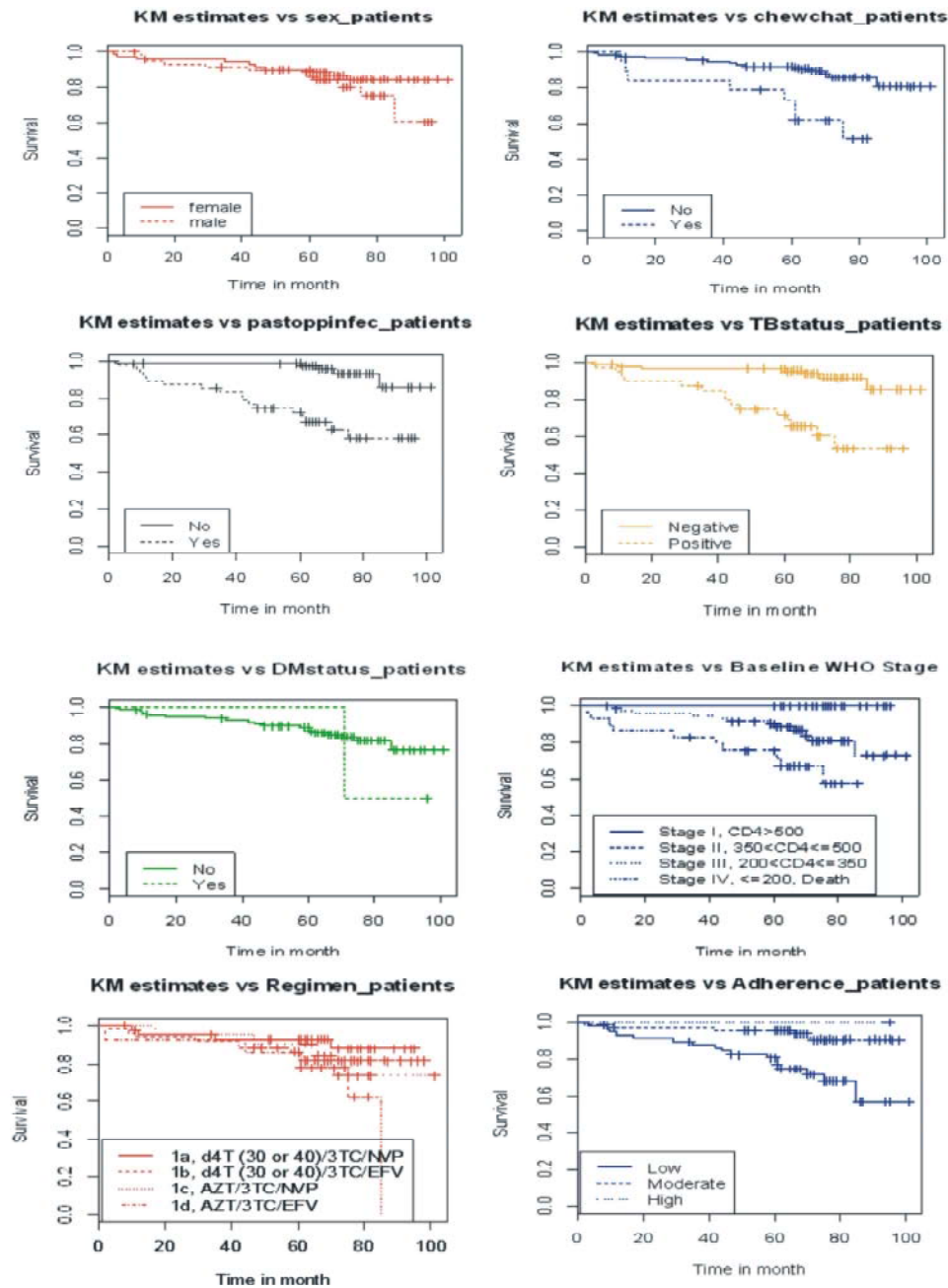Fig. 3: Kaplan-Meier Estimate with Confidence Limits

Fig. 4: The Kaplan-Meier survival function estimates

Table 2: Comparison of survival experience on HIV/AIDs patients using demographic, health and risk behavior variables

Covariate		N	Obs.	Exp.	(O-E)^2/E	(O-E)^2/V	Chisq	df	p
Sex	*female*	71	10	12.49	0.498	1.16	1.2	1	0.281
	male	60	12	9.51	0.654	1.16			
Chewchat	*no*	112	14	19.23	1.42	11.4	11.4	1	0.0007
	yes	19	8	2.77	9.88	11.4			
Pastoppinf	*no*	83	5	15.04	6.7	21.4	21.4	1	3.81e-06
	yes	48	17	6.96	14.5	21.4			
Tbstatus	*negative*	91	7	16.35	5.35	21.1	21.1	1	4.45e-06
	positive	40	15	5.65	15.47	21.1			
Dmstatus	*no*	129	21	21.535	0.0133	0.632	0.6	1	0.426
	yes	2	1	0.465	0.6152	0.632			
WHO stage	*I*	5	0	1.08	1.07887	1.143			
	II	24	0	4.64	4.63629	5.902	15.1	3	0.00174
	III	73	12	12.32	0.00832	0.019			
	IV	29	10	3.96	9.18730	11.314			
Regimen	*1a*	43	4	7.55	1.6697	2.5570			
	1b	51	9	8.50	0.0292	0.0477	4.9	3	0.178
	1c	23	4	3.64	0.0347	0.0419			
	1d	14	5	2.30	3.1576	3.5386			
Adherence	*low*	58	17	9.258	6.474	11.23			
	moderate	72	5	12.468	4.473	10.36	11.3	2	0.00356
	high	1	0	0.274	0.274	0.28			

Table 3: Univariate analysis of Cox proportional hazards on the time to event of HIV/AIDs patients

Covariate	$\hat{\beta}$	SE	Wald	df	Sig.	Exp($\hat{\beta}$)	95% CI for Exp($\hat{\beta}$)
Age	-0.03053	0.02815	1.18	1	0.278	0.96993	(.9179, 1.025)
Sex *(Female)*	0.4596	0.4297	1.14	1	0.284	1.5835	(0.6822, 3.676)
Weight	-0.0704	0.02874	6.01	1	*0.014*	0.9320	(0.8809, 0.986)
Chewchat *(No)*	1.4075	0.4501	9.78	1	*0.001*	4.0858	(1.691, 9.872)
Pastoppinfec *(No)*	2.0111	0.5103	15.53	1	*8.12e-05*	7.4717	(2.748, 20.31)
TBstatus *(Negative)*	1.8506	0.4607	16.14	1	*5.892e-05*	6.3634	(2.58, 15.7)
DMstatus *(No)*	0.7955	1.0284	0.6	1	*0.04392*	2.2155	(0.2952, 16.63
CD4count	-0.0112	0.00269	17.16	1	*3.439e-05*	0.9889	(0.984, 0.994)
WHO stage *(IV)*			12.014	3	0.006		
I	-1.304	0.498	4.311	1	0.033	0.271	(0.086,0.894)
II	-1.105	0.392	7.946	1	0.001	0.301	(0.147,0.660)
III	-1.041	0.356	8.55	1	0.004	0.353	(0.175,0.710)
Regimen *(1d)*			4.49	3	0.2132		
1a	0.6965	0.6013	1.341	1	0.2468	2.0067	(0.6175, 6.521)
1b	0.7351	0.7094	1.073	1	0.3001	2.0857	(0.5193, 8.376)
1c	1.4189	0.6724	4.452	1	0.0348	4.1326	(1.1062, 15.439)
Adherence *(high)*			8.99	2	0.01120.0		
low	-1.53e+00	5.092e-01	8.98	1	0272	2.173e-01	(0.0801, 0.5894)
moderate	-1.689e+01	6.486e+03	9e-6	1	0.99792	4.609e-08	(0.00000, inf.)

Table 4: Cox Proportional Hazard Model for HIV/AIDs Patients under ART at Arba Minch referral Hospital from (1997-2005)

Covariate	$\hat{\beta}$	SE	Wald	df	Sig.	Exp($\hat{\beta}$)	95% CI for Exp($\hat{\beta}$)
Weight	-.058	.028	4.385	1	0.036	.944	(.894,. 996)
Pastoppinfec	-1.709	.548	9.740	1	0.002	.181	(.062, 0.530)
DMstatus	-2.430	1.176	4.268	1	0.039	.088	(.009,. 883)
CD4	-.006	.003	5.941	1	0.015	.994	(.989,. 999)
WHO stage *(IV)*			12.014	3	0.006		
I	-1.304	0.498	4.311	1	0.033	0.271	(0.086,0.894)
II	-1.105	0.392	7.946	1	0.001	0.301	(0.147,0.660)
III	-1.041	0.356	8.550	1	0.004	0.353	(0.175,0.710)
Adherence level *(High)*			16.499	2	0.000		
Low	1.567	0.443	12.510	1	0.000	4.794	(2.011,11.42)
Moderate	0.850	0.314	7.311	1	0.007	2.339	(1.263,4.330)

Similarly, the covariates past opportunistic infection and TB are significantly associated with the survival of AIDs patients stratified by sub-groups this is confirmed by Log-rank test results in Table 2.

Kaplan-Meier Estimate with Confidence Limits
The Kaplan-Meier Survival Function Estimates:
Single Covariate Cox Proportional Hazards Model:
Single covariate Cox proportional hazards model analysis is an appropriate procedure that is used to screen out potentially important variables before directly included in the multivariate model. The relationship between each covariates and survival time of HIV/AIDs patients are presented in Table 3. As can be seen from this Table, survival of the patients is significantly related with weight, chew chat, past opportunistic infection, TB status, CD4 count, WHO stage and adherence. But the covariate like regimen, age, sex and DM status are not statistically significant at 5% significant level. Furthermore, using a modest level of significance 25% to include in the multiple covariates model for further investigation are weight, chew chat, past opportunistic infection, TB status, CD4 count, WHO stage, adherence and regimen.

Multiple Covariates Cox Proportional Hazards Model:
One problem of single covariate approach is that it ignores the possibility that a collection of variables, each of which is weakly associated with the outcome, can become an important predictor of the outcome when taken together. It is for this reason that we used p-value of 0.25 for selection of variables that are candidates for the multiple covariate analysis from single covariate findings.

Based on Table 4, Survival of HIV/AIDs patients was significantly related with initial weight, adherence, past opportunistic infection, CD4 count, DM status and WHO stage. The values of the Wald statistic for individual β coefficients support that the estimated values $\hat{\beta}'_{i\,s}$ are significantly different from zero at $a = 5\%$ level of significance for all the above covariates.

Checking for the Linearity of Continuous Covariates in the Model: A number of techniques are available, all of which are designed to determine whether the data support the hypothesis that the effect of the covariate is linear in the log hazard and, if not, which transformation of the covariate is linear in the log hazard. The graphical method of testing linearity for continuous covariates was used. The martingale residuals may be plotted against covariates to detect for the correctness of the functional form.

Assessment of Model Adequacy: A check of the proportional hazards assumption can be done by looking at the parameter estimates β_1, …, β_q over time. We can safely assume proportional hazards when the estimates don't vary much over time. The null hypothesis of constant regression coefficients can be tested, both globally as well as for each covariate, by using the cox. z ph function.

The formal test applied to the model presented in Table 5, shows the time-dependent covariates (interaction of covariates with logarithm of time) were not significant for weight, past opportunistic infection, DM status, CD4 count, WHO clinical stage and Adherence level which justifies the proportional hazard assumption holds at 5% level of significance.
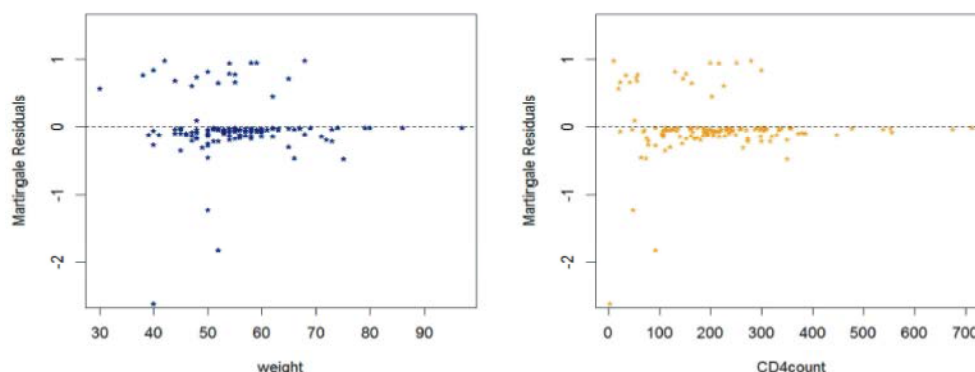

Fig. 5: Plots of the Martingale Residuals against the covariate initial weight and CD4 count in the final model

Table 5: Statistical test for proportional hazards assumption of the covariates

Covariate	rho(ρ)	chisq	P
Weight	0.3740	2.87e+00	0.0905
Pastoppinfec	0.0428	1.09e+00	0.2966
DMstatus	0.2203	1.09e+00	0.5429
CD4	0.1285	3.70e-01	0.5429
WHO stage *(IV)*			
I	-0.0491	8.46e-11	1.0000
II	0.3957	1.01e-08	0.9999
III	0.2312	5.27e-09	0.9999
Adherence level *(High)*			
Low	-0.4028	3.79e+00	0.0517
Moderate	-0.5055	2.13e-09	1.0000
GLOBAL	NA	1.06e+01	0.3077

Goodness of fit: A perfectly adequate model has low R^2 due to high percent of censored data (Cox, 1972). Thus, the model fitted in this study the value of R^2 statistic is 0.39, implying a good fit of the model. In addition to R^2, the results of the likelihood ratio test (chi-square =64.65 on 14 df, p=1.763e-08), Score test (chi-square=59.02 on14 df, p=1.739e-07) and Wald tests (chi-square= 24.66 on 14 df, p=0.03804) these all suggests that model is in good fit, i.e. significant at 5% level of significance. Thus, all in all we can say that our model fits the data very well.

Interpretation of the Final Model: The estimated coefficient for continuous risk factor initial weight being $\hat{\beta}$ = -0.058 with the hazard ratio exp($\hat{\beta}$) = 0.944, 95% CI: 0.894, 0.996). The 95% confidence interval suggests that an increase in the hazard rate may be as high as 0.996(99.6)% or even a decrease rate of 0.894(89.4%) in consistent with the data.

Looking at past opportunistic infection (past oppin fec), after adjusting other covariates, patients who had no any past opportunistic infection are found to be associated with high survival time, whose hazard rate is 0.181 times lower than those who have past opportunistic infection (adjusted HR=0.181, 95% CI: 0.062, 0.530) which means the survival time of patients who had no past opportunistic infection is increased by 0.819(81.9%) and the increment could be as low as 47% and as high as 93.8%. The estimated coefficient for continuous risk factor CD4 count being $\hat{\beta}$ = -0.006, which implies the hazard ratio is exp($\hat{\beta}$) = 0.994. This indicates the change of hazard rate for every unit increase or decrease in the CD4 count of HIV/AIDs patients (adjusted HR=0.994, 95% CI=0.989,0.999).

Patients with moderate adherence are dying at a rate which is 2.3 times greater than patients with high adherence. The estimated 95% confidence interval for hazard ratio indicates, the hazard rate for patient with moderate adherence could be as high as 4.330 times or as low as 1.263 times high adherent patients. HIV positive patients who are low adherent to ART are dying at a rate which is 4.7 times greater than patients who are high adherent to ART. The 95% confidence interval for hazard ratio of low adherence level is 2 to 11. This suggests those patient who are low adhere to ART may die as much as 11 times higher than or as low as 2 times higher than patients who are high adherent to ART.

The estimated hazard ratio of patients who are at stage II is 0.301, which implies that the death rate of WHO stage II patients is 0.301 times lower than the death rate of WHO stage IV patients or the hazard rate of patients who are at stage IV are 0.699 times greater than those who are at stage II with the 95% confidence interval for the hazard rate is as large as 0.86(86%) and as low as 0.34(34%).

Finally, family history of DM is another predictor variable related with risk of death of patients. The hazard of patients who had DM were found to be 0.088 times the hazard of those who do not have any history of DM (Adjusted HR=0.088, 95% CI=0.009-0.883).

Parametric Regression Modeling: For the data on HIV/AIDs patients the parametric models were fitted. The common applicable criterion to select the model is the Akaikie information criterion (AIC) statistic proposed by Akaikie (1974). From Table 6 the Log-logistic regression model has the least AIC value which shows that the data of HIV/AIDs patients fits for the Log-logistic regression model. Indifferent to [21], ensuring that model choice depends on the type of data.

Table 6: The AIC value for different parametric regression models

Model type	Exponential	Weibull	Log-logistic	Lognormal
AIC value	280.8389	280.7896	278.0451	279.9349

Table 7: Parameter estimates, standard errors and the hazard ratios in the final Log-logistic regression model

Covariate	$\hat{\beta}$	SE	Waldp-value	Exp ($\hat{\beta}$)	
Intercept	-0.712	1.927	0.136	0.712	0.488
Sex	0.258	0.38	0.461	0.497	1.294
Age	0.031	0.024	1.671	0.196	1.031
Weight	0.094	0.025	14.769	0	1.098
Chewchat	-1.169	0.508	5.295	0.021	0.311
Pastoppinfec	-1.561	0.629	6.15	0.013	0.21
TBstatus	0.353	0.551	0.411	0.521	1.424
DMstatus	-2.746	0.985	7.767	0.005	0.064
CD4count	0.004	0.003	1.682	0.019	1.004
WHOstage	0.046	0.383	0.015	0.09	1.048
Regimen	-0.209	0.183	1.295	0.255	0.812
Adherence	0.3	0.431	0.5	0.479	1.356

Table 8: The likelihood ratio and significance of the Log-logistic regression model

Loglik (intercept only)	Loglik (model)	Chi-square	df	p	Scale	Intercept
-153.6	-126	55.08	11	7.5e-08	0.58	-0.71178

Multivariate Analysis of Log-Logistic Regression Model:
Results presented in Table 7 indicate the parameter estimates of coefficients $\hat{\beta}$ for the covariates in the final Log-logistic regression model along with the associated standard error, significance level and hazard ratio. Survival time of HIV/AIDs patients were significantly related with initial weight, chew chat, past opportunistic infection, DM status, CD4 count and WHO clinical stage as can be seen from Table 7. These results are comparable with the Cox regression model in Table 5 above.

Assesment of Adequecy of the Log-logistic Regression Model: From the likelihood ratio test Table 8, it can be seen that the model is significant and in using the log likelihood values of the null model and the full model it can be seen that the model has a significant improvement after the covariates are added in the model.

CONCLUSION

This study was aimed to estimate time to death of HIV/AIDs patients and determine significant risk factors for the survival times at Arba Minch Referral Hospital, Gamo Gofa Zone, Ethiopia. For determining the risk factors and modeling the survival time, a total of 131 patients were considered. 71 (54.2%) are females and 60(45.8%) are males. Among those patients 22(16.8)% were died and the rest were censored.

The findings based on semi parametric survival regression model on the data of AIDs patients reviled that the major risk factors determining the survival time are initial weight, past opportunistic infection, DM status, CD4 count, WHO clinical stage and adherence level. But the covariate sex, age, TB status and regimen were not significant factors. Among various parametric regression models, Exponential, Weibull, Lognormal and Log logistic, the Log-logistic regression model found to have small AIC. Therefore, the Log-logistic regression model better fits the data of AIDs patients.

The hazard rate of patients who have any past opportunistic infection, positive family history of DM, low CD4 count, being at the worse stage (i.e. stage III and IV), initial weight and low adherence have relatively increased. This result confirmed with both Cox proportional hazard regression model and the Log-logistic regression model.

Recommendation: Stakeholders, public health policy makers, researchers and the public at large, brings HIV/AIDs diseases and other chronic diseases to their agenda, so that appropriate prevention and control strategies are implemented along with a population wide surveillance intervention.

Health care workers should anticipate and inform patients about the possible related risk factors of death through early diagnosis and appropriate intervention. Donors and government should understand the risk factors that influence the death of AIDs patients.

Future works required to assess the survival time of HIV/AIDs patients having considered another important risk factors and further extending model fitting paradigm towards current statistical topics like; Bayesian Survival Analysis.

REFERENCES

1. Federal Ministry of Health, 2010. Antiretroviral Therapy in Ethiopia. Addis Abeba, Ethiopia.
2. Federal Ministry of Health, 2007. Antiretroviral Treatment for Adult and Adolescence. Addis Abeba, Ethiopia.
3. UNAIDs., 2008. Report on the Global AIDs Epidemic. Geneva. http://www.unaids.org [accessed 10 May 2011].
4. FHAPCO and MOH, 2008. Report on Progress Towards Implementation of the UN Declaration of Commitment on HIV/AIDs.
5. Buve, A., M. Cara el and R.J. Hayes, 2001. Multi-center Study on Factors Determining Differences in Rate of Spread of HIV in sub-Saharan Africa: methods and prevalence of HIV Infection. AIDs, 15(4): 5-14.
6. Nuredin Ibrahim, 2007. Evaluation of Factor Affecting the chance of survival/death Status among HIV positive people under the antiretroviral treatment program: the case of Adama Hospital. A Journal of Ethiopian Statistical Association.
7. Giuseppe, S., 2000. Modelling the HIV/AIDs epidemic via survivor functions, European Journal of Epidemiology, 16: 573-579.
8. Braitstein, P., M. Brinkhof, F. Dabis, 2006. Mortality of HIV Infected Patients in the First of Antiretroviral Therapy: Comparison between Low Income and High Income Countries.
9. Palella, F.J., J. Delaney, K.M. Moorman, A.C. Loveless, M.O. Fuhrer J. and G.A. Satten, 1998. Declining Morbidity and Mortality Among Patients with Advanced Human Immunodeficiency Virus Infection. HIV Outpatient Study Investigators. North Engl, Journal of Medicine, 338(13): 853-60.
10. United Nations Programme on HIV/AIDs (UNAIDs)., 2010. Report on the Global AIDs Epidemic, Geneva.
11. Zhang, X. 2007. HIV/AIDs Relative Survival Analysis. Master's Thesis Published in Georgia University. USA.
12. Monica, M. and P. Patrizio, 2006. Comparing Costs and Effectiveness of Different Starting Points for Highly Active Antiretroviral Therapy in HIV Positive Patients. Acquir Immune Deficiency Syndrome, 11: 197-202.
13. Dlodlo, R. and A. Fujwara, 2005. Should Tuberculosis Treatment and Control be Addressed Differently in HIV-infected and Uninfected Individuals Eur Respir, 25: 751-757.
14. Bucher, H.C., L.E. Griffith and G.H. Guyatt, 2003. Isoniazid Prophylaxis for Tuberculosis in HIV Infection: A Meta-Analysis of Randomized Controlled Trials., 13: 501-507.
15. Gandhi, N., M. Anthony, A. Willem, R. Sturm, B. Pawinski and G. Thiloshini, 2006. Extensively Drug-resistant Tuberculosis as a Cause of Death in Patients Co-Infected with Tuberculosis and HIV in a Rural Areaof South Africa, South Africa.
16. World Health Organization, 2008. Report of a Joint World Health Organization HIV/AIDs and TB Department Meeting. Geneva, Switzerland.
17. Zelalem Getahun, 2010. Statistical Modelling of HIV/AIDs Evolution and Survival of AIDs Patients: A Case Study at Bahir-dar Felege-Hiwot Referal Hospital, Ethiopia.
18. Ministry of Health, 2006. AIDs in Ethiopia, Sixth report, Addis Ababa, Ethiopia, pp: 150.
19. United Nations Program on AIDs and World Health Organization. 2006. AIDs Epidemic Update, Brussels, Belgium.
20. Zeh, A., 2008. Outcome of HAART in Patients with Tuberculosisin Themba Lethu Clinical Cohort, South Africa
21. Derbachew, A., E. Fikre and A. Cheru, 2015. Survival Analysis of Diabetes Mellitus Patients Using Parametric, Non-Parametric and Semi-Parametric Approaches: Addis Ababa, Ethiopia. Ee-JRIF, 7(1): 20-39.
22. Akaike, H., 1974. A new look at the statistical model identification, IEEE Transactions on Automatic control 19(6):716-723,doi:1.1109/TAC.1974.1100705.