HAWASSA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES
SCHOOL OF MATHEMATICAL AND STATISTICAL SCIENCES

STATISTICAL ANALYSIS OF FACTORS THAT AFFECT SURVIVAL TIME OF HIV/AIDS PATIENTS AT SIDAMA ZONE

Principal investigator: Tofik Mussa (MSc.)
Co-investigator: Abdurehman Kassa (MSc.)

MAY, 2014
HAWASSA, ETHIOPIA

HAWASSA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES
SCHOOL OF MATHEMATICAL AND STATISTICAL SCIENCES

STATISTICAL ANALYSIS OF FACTORS THAT AFFECT SURVIVAL TIME OF HIV/AIDS PATIENTS AT SIDAMA ZONE

Principal investigator:

Name	Signature	Adress
Tofik Mussa	————	toffamr@gmail.com, 0911878075

Co-investigator:

Name	Signature	Adress
Abdurehman Kassa	————	abdut2@gmail.com, 0913368728

ACKNOWLEDGMENT

I would like to take this opportunity to thank Hawassa university to allow me to do this research. My gratitude to Abdurehman Kassa who is the co-investigator of this work. And also I thank Statistical Program staffs that they help me in sharing ideas and materials specially Denekew Bitew for his advice and comments.

Table of Contents Page

SUMMARY

AIDS is an epidemic that affect many people living in every part of the globe. The epidemic has immense impacts on economy of the world. Ethiopia is one most affected countries in the disease. An ART program supports HIV/AIDS patients require accurate data regarding patients.

In this study survival analysis methodology is used to analyze the probability of patients survival time given their status in different categories and practices. These categories and practices may be regarded as behavioral and natural factors.. Using parametric and semi-parametric cox proportional regression we are tried to identify factors that affect the survival time of HIV/AIDS patients. For parameter estimation we used maximum likelihood estimation method and Newton-Raphson iteration method. Cox-Snell method of model adequacy shows the model is good fit to the observed data.

The survival probability of patients can be affected in different behavioral and natural factors. We found that smoking, Imbibe alcohol, use of chat and TB treatment highly affect the survival time of The patients. The study is conducted in two hospitals of Sidama zone, Yirgalem hospital and Adare hospital. The study coast the researcher about 17,000 birr which Hawassa university was the sponsor for this coast.

1 INTRODUCTION

1.1 Background of the Study

AIDS, stands for Acquired Immune Deficiency Syndrome, is caused by HIV (Human Immunodeficiency Virus) is an epidemic that affects people living in every part of the globe. According to (UNAIDS, 2008) report an estimated 33 million people living with the disease world wide among which 22 million are in sub-Saharan Africa. According to the report the infection was 30.8 million in adults and the rest 2.2 million in children.

The epidemic is in spreading. There were 2.7 million new infected people world wide (UNAIDS, 2008). According to the report sub-Saharan Africa accounts 67%. The report shows over 2 million people died with the epidemic.

Ethiopia, belongs to sub-Saharan Africa, is the second largest country in population number in the region. It has large number of peoples infected by the disease. According to (MOH, 2006) 84.6% of HIV infected people are live in four regions. Addis Ababa, Amhara, Oromia and SNNPR with 30.2%, 22.7%, 22.2% and 9.5% respectivelly.

The epidemic worse poverty and it increases the vulnerability to the future collapse of livelihood. MOH reports, in 2005 there were $175,877$ peoples live with HIV/AIDS in SNNPR from which $20,305$ were children. The report shows in 2005 there were $18,496$ new infections, among which 4745 were children. It estimates 17866 deaths of all age group annually.(MOH, 2006)

1.2 statement of the problem

There are different factors that affect the survival time of patients. An ART program provides Medicines and some supports to patients regarding beviaral changes. Therefore it is important to identify these factors to ART program that provide clinical and non-clinical measures. Moreover this study motivate to identify the major factors that affect the survival time of HIV patients in addition to it inspire to answer the following questions.

- What are the the major factors that affect the survival of HIV patient.
- What is the survival of patient given the status.
- How can we compare the survival of different groups

1.3 Significance of the Research

Survival analysis has diverse applications in several disciplines. The outputs of the research will give factors that affect the survival time of HIV/AIDS patients. The survival analysis accuracy can then be integrated in to health science to identify those factors that affect survival time. In addition to the study gives recent information.

Moreover there has been an increasing interest in survival modeling. In the study theories computer algorithms has been used. Therefore the importance of this study may serve as basis for further research analysis and it will have paramount importance for theoretical frame work and practices.

1.4 Objectives

1.4.1 General Objective

The genera objective of this study is to identify factors that affect the survival time of HIV/AIDS patients

1.4.2 Specific Objectives

The following important points are included in the specific objectives of the study

- to analyzing the probability of the survival time of patients
- to identify some factors that affect the probability of survival time of patients
- to develop model that used to predict survival time of patients
- to compare the survival probability of different group patients

2 MATERIALS AND METHODS

2.1 Study Area

The study is carried out in Southern Nation Nationalities Peoples Region (SNNPR) State specifically Sidama zone. Sidama is one of the zones in SNNPR, which SNNPR is located in Southern and South-Western part of Ethiopia. Astronomically it roughly lies between $4^0 43' - 8^0 58'$ North latitude and $34^0 88' - 39^0 14'$ East longitude. It is bordered with Kenya in south, Sudan in south-west, Gambella region in north-west and surrounded by oromia region in the north-west, north and east directions.

The capital city of the region, Hawassa city, is 270 km from Addis ababa. SNNPR is divided in to 13 zonal administrations has $110,931.9 KM^2$ total area, which accounts 10% of the total area land of the country. According to the report of Central Statistical Agency of Ethiopia SNNPR has $15,760,743$ peoples live in the region (in 1999E.C). Accordingly this figure is nearly 20% of the country population. Among the total populatio 49.73% percent were males and 50.27% were females.

2.2 Target Population and Study Variables

The target population for this study is all patients under follow up of anti retriovairal treatment(ART) at Sidama zone in two hospitals, Yirgalem Hospital and Adare Hospital.

The outcome of this study is the survival time of HIV/AIDS patients under follow up of ART initiate the treatment in ten years, 1995-2005 EC. The survival time is the amount of time that a patient follow up ART until event with in the time 1995 to Megabit 15/2006EC. Event is the death of patient

in the study time, termination of follow up, initiate the treatment and transfered to other hospitals or survive until the end of the study. The event, termination of the treatment,transfer to other hospitals or survive until the end of the study is called right censored. A patient initiate ART else where and transfered to the two hospitals for treatment until event is called left censored. The event is dependent variable denoted by Y

$$Y = \begin{cases} 1 \;\; if \;\; a \;\; patient \;\; died \\ 0 \;\; other \;\; wise \end{cases}$$

Variables expected to relate with the dependent variables, Independent variables are

1. Continuous variables
    - Age
    - Weight
2. Catagorical variables
    - Gender $1 = male, \;\; 0 = female$
    - Imbibe alcohol $1 = yes, \;\; 0 = no$
    - Chat use $1 = yes, \;\; 0 = no$
    - Smoke cigarate $1 = yes, \;\; 0 = no$
    - WHO catagory $1 = stageI, \;\; 2 = stageII, \;\; 3 = stageIII, \;\; 4 = stageIV$
    - Treatment of TB $1 = yes, \;\; 0 = no$

2.3 sampling frame

The target population is all HIV/AIDS patients under the follow up of ART in two hospitals. The total population is 3804, among which 1560 are in Yirgalem hospital and 2244 are in Adare hospital

2.3.1 Sample size determination

The sampling technique is simple random sampling. The number of samples in each hospital is based on proportional allocation

In this study the sample size will be calculated according to Cochran 1977 formulation

$$n = \frac{\frac{Z^2 p(1-p)}{\Delta^2}}{1 + \frac{1}{N}\{\frac{Z^2 p(1-p)}{\Delta^2}\}}$$

Where Z is standardized upper value of probability $\alpha/2$, while α is level of significance value. p represents the death proportion and N is population size. Δ is level of precision.

the sample size is determined under the consideration of objective of the study, time constraint and budget constraint and margin of error

using N=3804 $Z_{0.025} = 1.96, \Delta = 0.0305$ and p=0.07(obtained from previous study Nuredin(2007) and Denekew) the sample size is n=252 and the sample in adare $n_a = 149$ and the sample size in Yirgalem $n_y = 103$

2.4 Methodology

The observed time will be values of a random variable T. The distribution of the random variable T can be described in a number of equivalent ways. There is of course the usual (cumulative) distribution function.

$$F(t) = P(T \leq t), \quad t \geq 0$$

When T is a survival time, $F(t)$ is the probability that a randomly selected subject from the population will die before time t. If T is a continuous random variable, then it has a density function $f(t)$, which is related to $F(t)$ through following equations [2]

$$f(t) = \frac{dF(t)}{dt} \qquad F(t) = \int_0^t f(u)du$$

Then it is often common to use the survival function

$$S(t) = P(T \geq t) = 1 - F(t)$$

The survival function $S(t)$ is a non-increasing function over time taking on the value 1 at t = 0, i.e., $S(0) = 1$. For a proper random variable T, $S(\infty) = 0$, which means that everyone will eventually experience the event. Obviously if T is a continuous random variable, we have

$$S(t) = \int_t^\infty f(u)du, \qquad f(t) = \frac{-dS(t)}{dt}$$

That is, there is a one-to-one correspondence between f(t) and S(t).

mean survival time

The mean survival time is the expected survival time of a patient. Which can be obtained using:

$$E(T) = \int_0^\infty S(t)dt$$

Median Survival Time

Median survival time m is deŕned as the quantity m satisfying S(m) = 0:5. Sometimes denoted by $t_{0.5}$. If S(t) is not strictly decreasing, m is the smallest onesuch that $S(m) \leq 0.5$.

pth Quantile survival time

pth quantile of survival time ($100pth$ percentile): t_p such that $S(t_p) = 1 - p(0 < p < 1)$. If $S(t)$ is not strictly decreasing, $tp$is the smallest one such that $S(t_p) \leq 1 - p$.

2.4.1 Kaplan-Meier Estimation

Description of survival data in numerical figures and graphical figures is a prerequisite for inferential statistics. This description includes survival distribution and Kaplan-Meier(KM) estimation of survival function. The KM estimation of survival function, non-parametric estimator, is the product of conditional probabilities. It is given by

$$\hat{S}(t) = \prod_{t_j<t}(1 - \frac{d_j}{n_j})$$

Where d_j the number of events at time t_j and n_j is the number of individuals have not yet the event exprienced.

2.4.2 Proportional Hazards model

Proportional hazard model proposed by Cox(1972) known as Cox regression is the basic model for survival data. This model is semi-parametric as it relate life time parametrically with individuals characteristics and does not specify the distribution of life time. The hazard model written in terms of hazard function is given by

$$h(t, X, \beta) = h_0(t) \exp(X\beta)$$

Where X is a matrix which is a vector of values of individual explanatory variable at time t and β is the vector of unknown parameters. $h_0(t)$ is the baseline hazard function, that is the value of hazard function when the matrix X entrly zero.
Alternatively the model can be described as the logarithm of the ratio of hazard function is the linear combination of covariates

$$\log\{\frac{h(t, X, \beta)}{h_0(t)}\} = X\beta$$

The cumulative hazard function is

$$H(t, X, \beta) = H_0(t) \exp(X\beta)$$

Then the survival function is given by

$$S(t, X, \beta) = [S_0(t)]^{\exp X\beta}$$

where $S_0(t)$ is the baseline survival function

2.4.3 Model Fitting

Maximum likelihood estimation is used to estimate the parameters. This method estimates regression parameters β by maximizing a likelihood function denoted by L or $L(\beta)$. Actually we use "partial" likelihood function as the Cox model do not consider probabilities of all subjects. The "partial" likelihood function is written as the product of all likelyhood functions. Each likelihood function is for failure time. Then the likelihood function of k failure times is

$$l(\beta) = \prod_{j=1}^{k} l_j$$

where l_j is the j^{th} failure time.

2.4.4 Hypothesis Testing

The significance of the effectiveness of the independent variable need to be tested. We always made two complementary statements about the independent variable in the significance. This significance is checked by testing the following hypothesis.

$$\gamma = \begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases}$$

Where β_i is the coefficient of the independent variable whose significance is to be tested. We will reject a statement that the independent variable is not significant, i.e $H_0 : \beta_i = 0$ if the $p-$value is very small compared to our $\alpha-$value. While α is the probability of rejecting H_0 when it is true. A statement corresponds to H_0, the independent variable does not significantly affect the dependent variable, is called null hypothesis. And the negation of this statement, H_1, is called alternative hypothesis

2.4.5 Most Powerful Test

There is common to do errors in statistical estimation and investigation. The object is to minimize the errors. These errors are when we accept false statement or/and when we reject true statement. These errors are called TypeII error and Type$I$ errors respectively.

A test $\gamma^* = \begin{cases} H_0 : \beta_i = \beta_0 \\ H_1 : \beta_i = \beta_1 \end{cases}$ is most powerful test if it satisfies two conditions

- $\pi_{\gamma^*}(\beta_0) = \alpha$
- $\pi_{\gamma^*}(\beta_1) \geq \pi_{\gamma}(\beta_1)$ where $\pi_{\gamma}(\beta_0) \leq \alpha$

  Where $\pi_{\gamma}(\beta_0)$ the probability of rejecting the null hypothesis when it is true and $\pi_{\gamma}(\beta_1)$ the probability of rejecting the null hypothesis when it is false. Therefore a test of hypothesis with maximum probability of rejecting the null hypothesis while it is false when the probability of rejecting null hypothesis while it is true is not less than any other test is most powerful test. That is a hypothesis with minimum probability of committing Type$I$ error and TypeII error

2.4.6 Group Comparison

The survival time of one group distribution(say group 1) is stochastically larger than the other group(say group 2) if $S_1(t) \geq S_2(t)$ for all $t \geq 0$. where $S_i(t)$ is the survival function for group i. If T_i is the corresponding survival time for groups i, we also say that T_1 is stochastically (not deterministically)

larger than T_2. Note that $T1$ being stochastically larger than T_2 does NOT necessarily imply that $T_1 \geq T_2$.

2.4.7 Model Adequacy

After the model is developed we need to know how the model is effective. The effectiveness of the model can be assessed by several methods. The model is accurate if it meets the assumption.

2.4.8 Assumption Checking

Certain assumptions are required for proportional hazard regression model. Here we used testing goodness of fit and partial likelihood ratio test. Test of goodness of fit R^2 formulated by Cox and Snell(1989) is:

$$R^2 = 1 - \exp\,[\frac{2}{n}(LL_0 - LL_{\hat{\beta}})]$$

Where LL_0 is the log likelihood of zero model and $LL_{\hat{\beta}}$ is the log likelihood of the model including covariates.
The partial likelihood model requires to fit restricted and unrestricted models. We want to test how much unrestricted model fits the data better than the restricted moel. And the partial likelihood model is checked by Chi-square test.

$$Q_{LR} = 2[LL_p(\beta) - LL(0)] \sim \chi^2(p)$$

3 Result and Discussion

The purpose of the study to estimate the survival time of HIV/AIDS patients under the follow up of ART. Kapplan-meir and parametric family of survival function are used to estimate the survival time and hazard function. an attempt has been made to identify the best survival and hazard functions based on different model selection tools. R-statistical software and STATA are used for computition and parameter estimation.

Factors like gender, cigarette smoking, using chat, imbibe alcohol, and treatment of Tb are analyzed as categorical variables whether they affect the survival time of the patient. Age and weight are also considered whether they affect survival time.

From total population 3804 patients under follow up of ART in Adare and Yirgalem hospitals 252 samples are taken. The sample shows 41 patients are died in the follow up time and the remaining 211 are censored. The follow up time is 11 years(1995-Megabit 15/2006 EC). The maximum time record time until the event happen under the follow up of ART is 10.55 years.

The mean time of patients in the ART follow up is 3.92 years and standard deviation 2.53 years. The minimum and maximum age records are 18 and 64 respectively. The minimum and the maximum weight are 25 and 102 kg respectively.

Table 3.1 shows the descriptive summary of categorical variables of patients under ART follow up in Adare and Yirgalem hospitals.

Table 3.1 Summary of the data in two hospitals

	Categories											
	stage				gender		Smoke		Alcohol		Chat		TB			
Hospital	I	II	III	IV	Male	Female	Yes	No	Yes	No	Yes	No	Yes	No	Total	Death
Adare	31	39	64	15	63	86	43	106	46	103	28	121	12	137	149	22
Yirgalem	28	34	28	13	49	54	34	69	13	90	27	76	13	90	103	19
Total	59	73	92	28	112	140	77	175	59	193	55	197	25	227	252	41

There are more female patients than male patients under the follow up of ART. In two hospitals there is no significance difference between patient proportions of stages, smokers, chat user and alcoholics, except that the proportion of patients with TB treatment in Yirgalem hospital is high relative to the group.

Table 3.2 shows the death and censored description with categories of study variables.

Table 3.2 Summary of the death data in catagories

	Categories												
	stage				gender		Smoke		Alcohol		Chat		TB		
	I	II	III	IV	Male	Female	Yes	No	Yes	No	Yes	No	Yes	No	Total
death	4	15	18	4	23	18	26	15	21	20	20	21	15	26	41
censored	55	58	74	24	89	122	51	160	38	173	35	176	10	201	211
Total	59	73	92	28	112	140	77	175	59	193	55	197	25	227	252
Death/total%	6.8	20.5	19.6	14.3	20.5	12.9	33.8	8.8	35.6	10.4	36.4	10.7	60	11.4	

Table 3.2 shows more death proportions are recorded in male than female compared to the total male and female in the follow up of ART. The proportion death in cigarette smokers, chat users and alcoholics are high compared to non-smokers, non-chat users and non-alcoholics respectively (33.8% Vs 8.8%, 36.4% Vs 10.7% and 35.6% Vs 10.4% respectively). The percentage of death in treatment of TB from total TB patients are highest record.

3.1 Assumption checking

Some assumptions are made concerning the data. The data is a single failer data (i.e single event per subject). in explanatory variables there is no time varying covariate. Furthermore the covariates are fixed for each individual. i.e once the random variable is detected the value do not have stochastic property. The basic assumption is the time event censored observations are random. Figure 3.1 indicates the time event against individuals ID number.

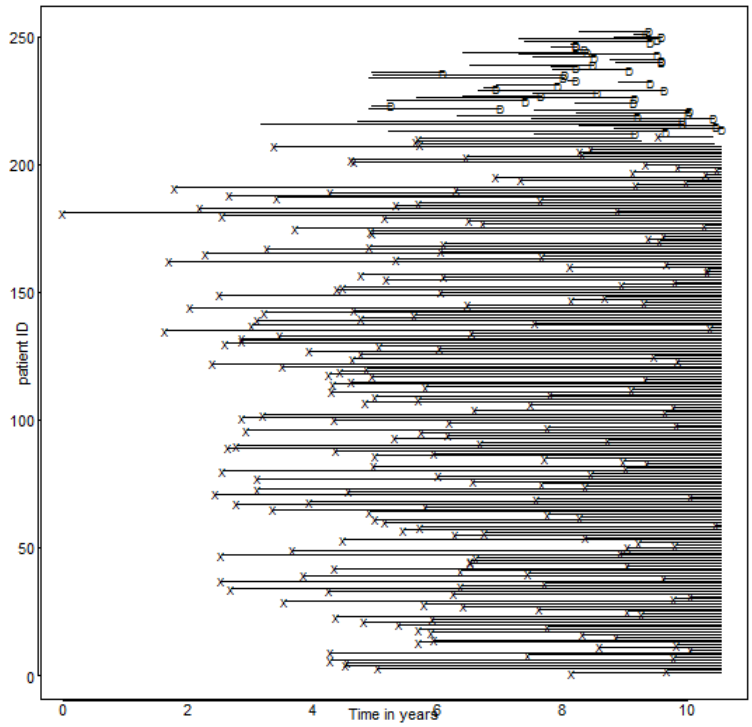
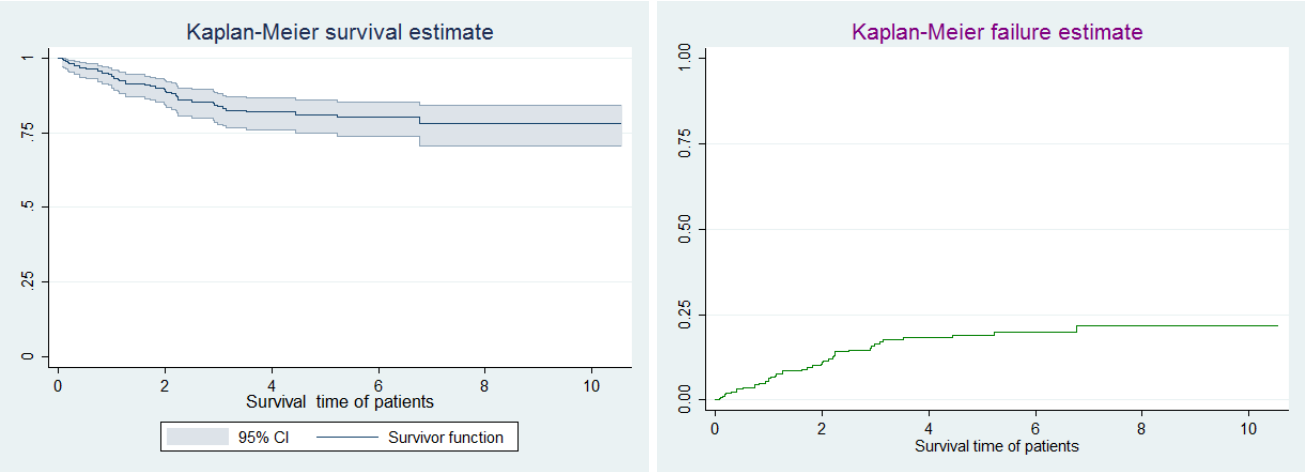


Figure 3.1: Survival time of patients, X indicates the start time of censored individual, D indicates the end time of dead individual

Figure 3.1 shows the events occur at random time. Furthermore the randomness of the series is tested using turning point test. We find the calculated value $Z_c = \frac{T-E(T)}{\sqrt{Var(T)}} = -0.3998501$, which is very small to compare $Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$. Therefore the time to event is random at 0.05 level of significance.

3.2 Kaplan-Meier estimation

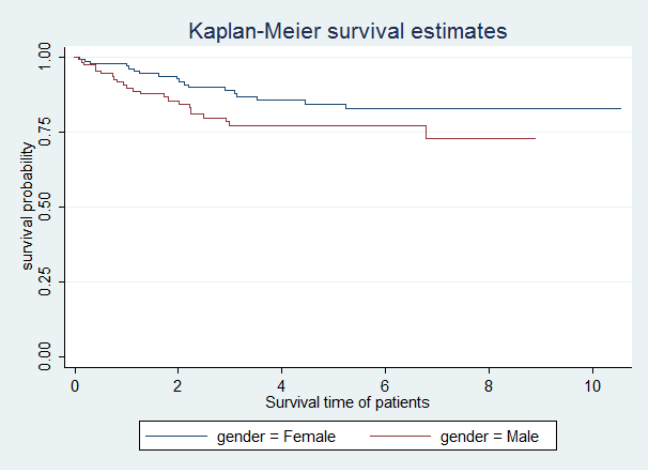
Estimation of survival time and cumulative hazard of patients are described. Neglecting the covariate effects on survival time and hazard Kapplan-Mier method is used. The Kaplan-Meier estimate of the survival time of patients and the cumulative hazard function at mean covariate given in Figure 3.2



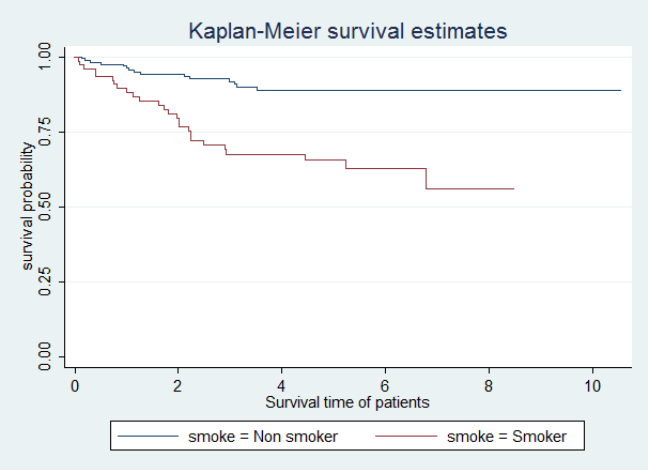
(a) (b)

Figure 3.2: Kapplan-Mier estimate of survival time and cumulative hazard of patients: (a) Kaplan-Meier estimate of Survival time under follow up of ART patients with upper and lower 95% confidence interval at mean covariate: (b)Kaplan-Meier estimate of cumulative hazard of patients at mean covariate.

The objective of the study is to see some factors that affect survival time of HIV/AIDS patients. Figure 3.3 the survival time of patients catagorized by gender, smoking, use of chat, imbibe alcohol and treatment of TB.

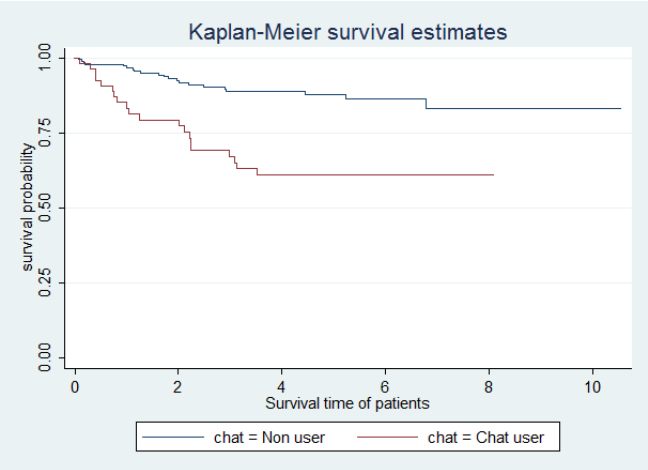


(a)

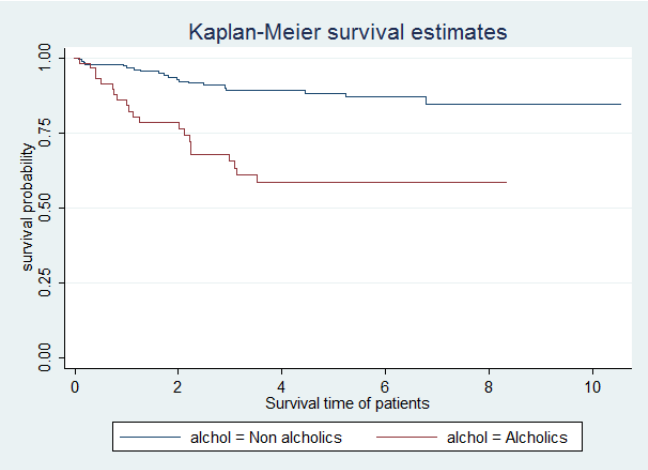


(b)

(a) Kaplan-Meier estimate of Survival time of patients by gender at mean covariate: (b) Kaplan-Meier estimate of Survival time of patients by smoking at mean covariate.

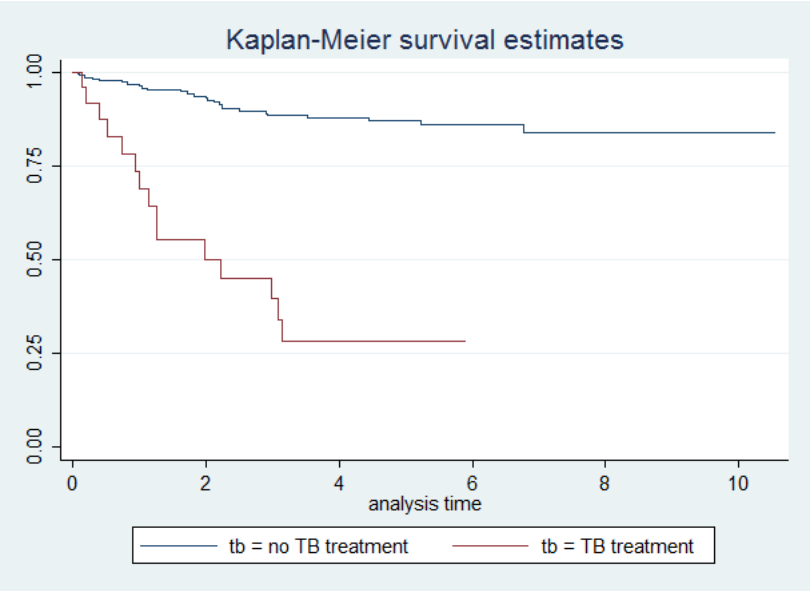


(c)



(d)

(c) Kaplan-Meier estimate of Survival time of patients by chweeing chat at mean covariate: (d) Kaplan-Meier estimate of Survival time of patients by imbibe alcohol at mean covariate.



(e)

(e) Kaplan-Meier estimate of Survival time of patients by treatment of TB at mean covariate.

Figure 3.2: The effect of covariates on Kaplan-Meier estimate of Survival time of patients.

From K-M estimate of the above figure we observe that the survival time of patients are affected by covariates by gender, smoking, imbibing alcohol, using chat and treatment of TB. Female patients have high probability of survival than males. Smoking, use of chat and imbibe alcohol seriously affects survival time. The survival time of HIV/AIDS patients with treatment of TB patients affected. From K-M estimate we see that the patients with TB treatment has high probability of to end their life time before six years.

3.3 Model fitting

For parametric family estimation of baseline hazard function or/and survival functions parameters are estimated using maximum likelihood. Proportional hazard models, which is the the combined effect of the baseline hazard and the

exponential of covariates, the baseline hazard function parameters assumed to take exponential, Gompertz, log logistic and Weibull distributions. In semi-parametric estimation, i.e. cox-regression model, the coefficients of covariates (factors) we use partial maximum likely hood estimation method. In Cox regression the baseline hazard model is estimated non-parametrically. The estimation of factor coefficients are independent of the baseline hazard, that is no assumption is made on the baseline distribution of cox regression.

3.4 Parametric models

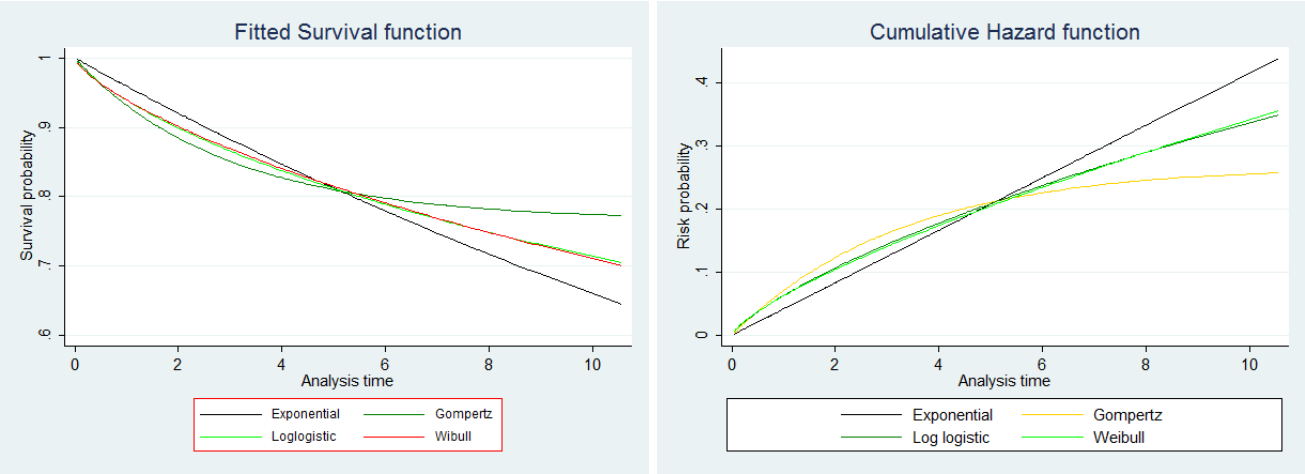
To describe the survival and hazard functions of the patients four types of density functions, exponential, Gompertz, log logistic and Weibull are taken to compare and explain. Table 3.3 gives the baseline function parameters using different distributions combined by model selection criteria. The parameters are estimated neglecting the covariate effect, which is the proportional hazards are 1 for all factors(i.e. the coefficients of all factors are zero) using maximum likely hood estimation method.

Table 3.3: The maximum likelihood estimate parameters using Newton-Raphson method

Model	Survival function($S(t)$)	parameter(s)	log likelyhood
Exponential	$\exp^{-\lambda t}$	$\lambda = 0.042$	-165.12
Gompertz	$\exp^{(\frac{\theta}{\alpha}(1-\exp^{\alpha t}))}$	$\theta = 0.0375, \alpha = 0.0025$	-159.54
Log logistic	$\frac{1}{1+\lambda t^{\alpha}}$	$\alpha = 0.79, \quad \lambda = 0.065$	-162.02
Weibull	$\exp^{-\lambda t^{\alpha}}$	$\alpha = 0.095, \quad \lambda = 0.021$	-162.54

The Gompertz model gives the large log likelihood values, Which corresponds to least AIC or BIC, implies it explains the situation better. However all likely hood functions do not have significance difference relative to each other. Therefore any of them could be used for the description, estimation and prediction. We will use Gompertz model for this study.

Figure 3.3 describes the status of survival function and cumulative hazard function using the parameters estimated in table 3.3. The functions are using the assumption of the density function takes exponential, Gompertz, log logistic and Weibull distributions.



(a) (b)

Figure 3.3: Estimated Functions using different density functions (a) The estimated Survival function : (b) The estimated cumulative hazard function

From the table we observe that the survival probability decrease through time and the hazard increases with increasing time regardless of covariates considered in this study.

3.5 Proportional hazard model

Table 3.4 gives the summary output STATA for proportional hazard model. The model is assumed the baseline hazard function estimated parametrically with exponential distribution. The combined effect of covariate effects are analyzed how they affect the survival time and cumulative hazard function. Coefficients, hazard ratios standard error, test statistic, and 95% CI ,p-value and Model selection criteria are given in the table.

Table 3.4: Proportional hazard regression with parametric baseline hazard

covar	Coef	Haz. Ratio	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
gender	.4628209	1.588549	.5514715	1.33	0.182	.8044561	3.136887
age	.0103008	1.010354	.0189396	0.55	0.583	.9739067	1.048165
weight	-.0159984	.9841289	.0194042	-0.81	0.417	.9468227	1.022905
chat	.471897	1.603032	.6141903	1.23	0.018	.7565027	3.396832
smoke	1.553601	4.728465	1.565477	4.69	0.000	2.471204	9.047565
alchol	1.099833	3.003665	1.055702	3.13	0.002	1.508268	5.981696
tb	1.971160	7.178998	3.473702	5.86	0.000	4.371862	19.27188
cons	-4.548443		1.1791	-3.86	0.000	-6.859436	-2.237451

log likelihood= -124.02845 $LR\chi^2_{(7)} = 82.19$ $Prob > \chi^2 = 0.0000$

The covariates observed in Table 3.4 have effects on survival or hazard function of HIV/AIDS patients. The $p - value = 0.000$ which is corresponds to $\chi^2_{cal} = 82.19$ suggests the model is significantly fit at 95% to the observed data. The $p - value$ corresponds to each covariate for smoking, imbibe alcohol, chat use and TB treatment shows these factors are significantly affect survival time of patients at 95%. chat users has 0.6 more hazard rate than non users, smokers have 3.7 more hazard rate than non smokers, and alcoholics have 2.003 more hazard rate than non-alcoholics. TB treatment is the most affecting factor on survival probability. The hazard rate of patients with TB tretment is 6.17 more than non-TB treatment patients.

From the result of the analysis of covariates gender, age, and weight are not significant factors at 95% confidence interval. However these factors are significant, gender at 81.8%, age at 41.7%, and weight at 58.3% probabilities. The effect of these covariates at thier significance percentage: males have 0.588 more hazard rate than females, increasing age by one year increases hazard rate by 0.01 and increasing weight by one unit decreases hazard by 0.015

3.6 Semi-parametric Cox Regression

To fit multiple cox regression model first we have to check the significance of each covariates separately on survival probability or hazard function. The following table gives covariate coefficients and the significance of covariates after they filter using univariate analysis. When we analyze individually all factors are significant. Table 3.5 gives the partial maximum likelihood estimation of parameters of Semi-parametric Cox regression using Newton-Raphson method.

Table 3.5: The partial maximum likelihood estimate of Semi-parametric cox regression using Newton-Raphson method

cova	coef	Haz. Ratio	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
gender	.429054	1.535804	.5391012	1.22	0.222	.7718719	3.05581
age	.0103976	1.010452	.0192444	0.55	0.585	.9734287	1.048883
weight	-.0162113	.9839194	.0193314	-0.83	0.409	.9467508	1.022547
chat	.469243	1.598783	.6172871	1.22	0.024	.7501402	3.407508
smoke	1.573575	4.823862	1.624751	4.67	0.000	2.492856	9.334534
alchol	1.036518	2.819383	1.017254	2.87	0.004	1.390062	5.718392
tb	2.020917	7.54524	2.956992	5.16	0.000	3.500143	16.26523

Log likelihood = -176.55254 $LR\chi^2_{(7)} = 72.85$ $Prob > \chi^2 = 0.0000$

There is a change in the output of Table 3.5 compared to the output of Table 4.4 that the significance of the variables are a little bit changed

3.7 Model adequecy

R^2 (formulated by Cox and Snell) is used to check the goodness fit of the model to the given data. For proportional hazard model, even if the model perfectly adequate, Hosmer and Lemshow state that the value of R^2 may be small due to the censored observations appeared in the data. In the fitted

model we calculated $LL_0 = -212.9773$ and $LL_B = -176.5525$ and then $R^2 = 25.1052$. which implies the model is adequate and fit good to the data.

One can also compute the goodness fit by LR test. The LR, which follows χ^2 distribution with degrees of freedom the number of parameters involved in the equation, value of $\chi^2 = 72.8496, p = 0.000$ which implies the model is good fit.

4 Conclusion and Recomendation

4.1 Conclusion

The study is conducted in SNNPR Sidama zone in two hospitals, Yirgalem Hospital and Adare Hospital. The study focus on identifying some factors like gender, age, weight, treatment of TB, cigarrate smoking, imbibe alcohol and chewing chat, whether they affect the survival time (hazard rate) of HIV/AIDS patients under the follow up of ART.

To explain the given data hazard and survival time Kapplan-Mier estimation is used. For model fitting and inference some parametric models, Cox-regression, proportional hazard model, are used. For baseline hazard model parametric family like exponential, Weibull, log normal, log logistic functions are used. The model gives the best fit is selected using model selection criteria.

Some of factors considered in this study significantly affect the survival time of patients (i.e. increase hazard rate). Males have high survival time than females. Patients who imbibe alcohol, smoke, chewing chat have high hazard (less survival time) than those who do not use.

4.2 Recomendation

ART program provides clinical and non-clinical supports to HIV/AIDS patients to improve their health and give long live. In this research an attempt has been made to identify some factors that affect the survival time of HIV/AIDS patients. Therefore some recommendations are made related to these risk factors.

- Cigarrate smoking, Chat use and imbibe alcohol seriously affect the survival time. Thus patients have to be aware of these matter and health workers are expected to consult patients for making behavioral changes.
- TB co-infection also affects the survival time. Therefore health workers have to take attention when patients face this problem.

4.3 Limitation

The outcome of the result is to represent only the situation of two hospitals. An attempt was made to increse the number of hospitals to be included in the sample, which is unsuccessful . The covariates attempt to be identified in this study may not be the only factors that affect survival time. In addition the limitations the ART stage are is considered because they may have time covariate.

References

[1] CSA, 1999: The 1994 Population and Housing Census of Ethiopia; Results at Country Level, Vol. 2, Analytical Report, June 1999, Addis Ababa.

[2] Daewon Zhang, 2005: Analysis of Survival Data; Lecture Notes

[3] MOH, 2006: AIDS in Etiopia; National HIV/AIDS Prevention and control Office; six Report

[4] UNAIDS, 2008: Reports on the Global HIV/AIDS Epidemic; Geneva, 362p.

[5] Denekew Bitew Belay, September 2012: Statistical Modelling of the Survival of HIV/AIDS Patients and Identi

cation of Early Mortality Changepoints after Initiation of ART : A Case Study at Hawassa University Referral Hospital.

[6] Nuredin Ibrahim, (2007): Evaluation of Factors Affecting the Change of Sur- vival/Death Status Among HIV Positive People Under the Antiretroviral Treatment Program : The case Study of Adama Hospital; Department of Statisics, Addis Ababa University, MSc.

[7] Daowen Zhang,2005: Analysis of Survival Data, Department of Statistics North Carolina State University

[8] Elias T. Lee, 2003: Statistical Methods for Survival Data Analysis,