Available online at www.elixirpublishers.com (Elixir International Journal)

Statistics

Elixir Statistics 117 (2018) 50413-50419

ISSN: 2229-712X

A Comparative Study of the Performance of Semi-Parametric and Parametric Survival Analysis Models in Assessing the Prognostic Factors of HIV/AIDS Mortality

James Friday E[1] and Jolayemi E. T[2]

[1]Department of Mathematics and Statistics, Federal University Wukari, P.M.B 1020, Wukari –Taraba State, Nigeria.

[2]Department of Statistics, University of Ilorin, P.M.B 1515, Ilorin –Kwara State, Nigeria.

ARTICLE INFO

Article history:

Received: 24 February 2018;

Received in revised form:

25 March 2018;

Accepted: 6 April 2018;

Keywords

Survival Analysis,
Event,
Prognostic factors,
Mortality,
HIV/AIDS,
Kaplan Meier,
Cox Proportional Hazard,
Parametric Models.

ABSTRACT

Survival analysis as a field of statistics has had a remarkable growth in the latter half of the last century. The methods used by survival analysis that have led to this insightful impact are the Kaplan-Meier method used in estimating the survival function, the log-rank test commonly used to determine the differences between two or more survival distributions and the Cox Proportional Hazard model used for investigating the effects of covariates on the hazard function. The parametric methods are rarely used. In this work, we present the basic concepts of survival analysis. Furthermore, we compare the semi-parametric (Cox Proportional Hazard) and parametric methods for analyzing survival data. We apply these methods to HIV/AIDS infected patients from secondary data to determine the prognostic factors of the event of mortality of the patients. It was found that the parametric method of survival analysis is good and probably preferred if the correct distribution is identified; and that drug abuse history of patients, tuberculosis status, patients' adherence to drugs and the patients' CD4 cell counts are prognostic factors for mortality experience of HIV/AIDS patients.

1.0 Introduction

HIV is an acronym for Human Immunodeficiency Syndrome, implying that the HIV is actually a virus. It is a virus that belongs to a class of viruses known as retroviruses. HIV is the virus that causes the Acquired Immunodeficiency Syndrome (AIDS).It attacks the human immune system directly, reproduces and multiplies by using the food and nutrients supplied by the body cells. HIV cannot grow or reproduce on their own. They need to infect the cells of a living organism in order to replicate (make more copies of themselves).The human immune system usually finds and kills viruses fairly quickly, but HIV attacks the immune system itself – the very thing that would normally get rid of a virus [5] With around 2.5 million people becoming infected with HIV in 2011, there are now an estimated 34 million people around the world who are living with HIV, including millions who have developed AIDS.

A peculiar characteristic of a person infected with the HIV is the progressive loss of an important kind of cells known as CD4 cells, otherwise called the T-helper cells. The virus increases the surface area of these cells, thereby reducing the cell counts while reproducing itself among the cells. Certain prognostic factors are said to be responsible for the mortality of HIV/AIDS patients and this study, by employing three survival analysis methods, determines the factors of mortality.

Survival Analysis, a statistical procedures of data analysis for which the outcomes of interest is time until an event occurs, has experienced remarkable growth during the latter half of the twentieth century. This field of statistics which is also referred to as "time-to-event" analysis has been found applicable to diverse fields such as Medicine, Social Sciences and Engineering. In Medicine, time-to-event could be the time from HIV to AIDS or death arising from AIDS or time to relapse of a cancer disease just to mention but a few; in the social sciences, the outcome could be time to events such as demand or supply changes, job changes, time to marriage among singles, time until birth of children among the married, and so on. Hence, survival analysis is a statistical method for data analysis where the outcome variable of interest is the time to the occurrence of the event under consideration [7]. In the Engineering field survival analysis is referred to failure time analysis because the main focus is on studying the time until failure of machines or electronic components. Because these methods have been used by experts in different fields, they also have several names. In sociology it is called event-history analysis and Economists put it as duration or transition analysis. However, these differences in the names do not suggest any difference in procedures, although different disciplines may stress somewhat different approaches. Survival analysis is the name that is most extensively used and acknowledged. The difficulties associated with the presence of censored observations steered up the development of a new field of statistical procedure. The methodological developments in survival analysis were largely achieved in the latter half of the 20th century.Although, Bayesian methods in survival analysis are well developed and are becoming quite common for survival data [2], this work will focus on the methods that are often used.

Tele: +234(0)7063706382, +234(0)8152257851
E-mail address: james@fuwukari.edu.ng

The earliest methods of survival analysis is the use of life tables, which was proposed by Berkson and Gage [1] in their work on survival rates for cancer. Kaplan and Meier (1958) [6] came up with an important method of estimation from incomplete observations and it is commonly known as the Kaplan-Meier (K-M) method. The non-parametric methods is very effective for homogeneous samples. However, they do not determine the relationship between the variables and the survival times. This requirement leads to the application of regression methods in survival analysis. In as much as the use of regression models were employed, they are not well suited for analysing survival data for certain reasons suggested by Klembaum [7].First, survival times are hardly normally distributed. Secondly, due to censored observations which result in missing values for the dependent variables. The Cox proportional hazard (PH) model has become the most commonly used model for analysis of survival data in the existence of covariates. The popularity of the model could be said to be due to its simplicity and the point that it is not centred on any assumptions about the survival distributions, hence its name – semi-parametric survival model.

When the assumptions are made about the shape of the hazard function, it makes the model a parametric one. The Accelerated Failure Time (AFT) model is also an alternative methods of survival analysis.

2.0 HIV/Aids Predictors of Mortality

In a retrospective cohort study conducted by Sibhatu, et al (2012),aimed at determining the predictors of mortality among HIV infected patients taking Anti-Retroviral treatment in Ethiopia and followed up to five years period, a total of 1540 study participants were included in the study and Cox regression as well as the Kaplan-Meier analysis were performed to investigate the factors that influenced time to death and mortality [14].The results showed that from the registered patients in the cohort, out of a total of 86 deaths in over 60 months ,63(73.3%) died during the first 12 months, 10 (11.6%) during the second year and 10(11.6%) in the third year of follow up. In the multivariate analysis, the independent predictors for mortality were more than 10% weight loss, bedridden functional status at baseline ,less than or equal to 200 CD4 cell count /cubic millimetres and advanced World Health Organisation clinical stage patients.

A research was carried out by Moore,et al (2006) where they aimed to determine the prognostic value of baseline CD4 percentage in terms of patient survival in comparison to absolute CD4 cell counts for HIV-positive patients initiating HAART (Highly Active Anti-Retroviral Therapy) [11].In the study,a cohort of 1623 ART-naïve HIV positive individuals who initiated HAART between 1st of August 1996 and 20th of June, 2002 was conducted. Cumulative mortality rates were estimated using Kaplan-Meier. Then Cox Proportional Hazard model was used to model the effect of baseline CD4 strata and CD4 percentage strata and other prognostic variables on survival. In the model, low CD4 percentages were associated with increased risk of death (CD4 percent less than 5 and p-value less than 0.01).Poor adherence to therapy were also associated with an increased risk of death.

In an attempt to compare the possible causes of death and mortality rates among HIV-infected persons, Crum, et al, (2006) conducted a research considering three eras (Pre-, Early and Late HAART–Highly Active Anti-Retroviral Therapy).It was reported that the number of deaths declined over the study period with 987 deaths in the Pre-HAART era, 159 deaths in the early HAART era and 78 deaths in the late HAART era. The researchers concluded that despite increasing concerns regarding antiretroviral resistance, the death rates among HIV-infected persons in the cohorts continues to decline, a result which can be attributed to open access to health care [4].

3.0 Description of Dataset

The data set used in this study was obtained from the Kirby Institute for infection and immunity in society, Australia and consists of records of patients infected with HIV/AIDS from 2008 to 2011, with all cases of HIV infected persons older than eighteen (18) years of age and followed for the outcomes of death.

3.1 Covariates in the Study

Covariates are simply the predictor variables used in the analysis of survival data. These are explanatory variables which influence the survival of patients and are given below:

1. Age at diagnosis (in years)……A
2. Gender (Male, Female)
   - Male……...G_{male}
   - Female……...Reference Category
3. Body Mass Index (measured in Kgm^{-1})

Below 18…… B_1
18 – 25…….Reference Category
26 – 30 …….. B_2
31 – 35 …….. B_3
36 – 40 …….. B_4

4. CD4 cell count (counted in cells/mm^3)

1 – 200 ……Reference Category
201 – 400 …….. C_1
401 – 600 …….. C_2
601 – 800 …….. C_3
801 – 1000 …… C_4
Above 1000…… C_5

5. History of Drug Abuse (Yes, No)

Yes……$DABUSE_{yes}$
No……Reference Category

6. Tuberculosis Treatment Status (Yes, No)

Yes……TB_{yes}

No…….Reference Category
7. Adherence to therapy (Poor, Good)
Poor…… ……DA_{poor}
Good……Reference Category
8. Clinical Staging
Stage 1……Reference Category
Stage 2…… S_2
Stage 3…… S_3
Stage 4…… S_4
9. Regimen Type
TDF+3TC+EFV……. R_1
AZT+3TC+NVP……Reference Category

4.0 The Method

4.1 Survival Time Distribution

Assume T is a random variable that denotes the survival time. The distribution of T is characterized by any of these three functions: the probability density function, the survival function, or the hazard function. The survival function exists for both discrete and continuous cases of the survival time, T, as well as the probability density and hazard functions.

Survival function is the probability that the survival time T is greater than or equal to t .that is, the probability that an individual survives up to a time represented by t.

$$\boldsymbol{S(t) = P(T \geq t), \quad t \geq 0}$$

4.1.1 For Discrete Case Of Survival Time, T

If T takes a discrete random variable with values in the order: $\boldsymbol{0 \leq t_1 < t_2 < t_3 \ldots\ldots\ldots < t_n}$, let the probability mass function (pmf) of T be given by $P(T = t) = f(t_i), i = 1,2,3, \ldots$then the survival function is:

$$\boldsymbol{S(t) = \sum_{i/t_i \geq t} f(t_i)}$$
$$\boldsymbol{= \sum f(t_i).\, I_{t_i \geq t}}$$

Where $\boldsymbol{I_{t_j \geq t}}$ is called the indicator function defined as:

$$\boldsymbol{I_{t_i \geq t} = \begin{cases} 0 \; if \; t_i < t \\ 1 \; if \; t_i \geq t \end{cases}}$$

Here, the hazard function given as *h(t)* is the conditional probability of the event occurring at time t_j given that the individual has survived up to time t_j

$$\boldsymbol{h(t_j) = P(T = t_j | T \geq t_{j-1})}$$

$$\boldsymbol{= \frac{f(t_j)}{S(t_j)}}$$
$$\boldsymbol{= \frac{S(t_j) - S(t_{j+1})}{S(t_j)}}$$
$$\boldsymbol{= 1 - \frac{S(t_{j+1})}{S(t_j)}}$$
$$\boldsymbol{1 - h(t_j) = \frac{S(t_{j+1})}{S(t_j)}}$$
$$\boldsymbol{\prod_{t_j} (1 - h(t_j) = \frac{S(t_2)}{S(t_1)} \times \frac{S(t_3)}{S(t_2)} \times \ldots \times \frac{S(t_{j+1})}{S(t_j)}}$$
$$\boldsymbol{= S(t)}$$

4.1.2 For Continuous T

For an absolutely continuous variable T, the probability density function of T is:

$$\boldsymbol{f(t) = F'(t) = -S'(t), \quad t \geq 0}$$

Here, the hazard function gives the instantaneous failure rate at t given that the individual has survived up to time t.

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

thus, there is a clearly defined relationship between S(t) and h(t) which is given by the formula:

$$\boldsymbol{h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} log S(t)}$$

$\boldsymbol{S(t) = e^{[-\int_0^t h(u)du]} = e^{-H(t)} \qquad t \geq 0}$,

where $\boldsymbol{H(t) = \int_0^t h(u)du}$ is called the cumulative hazard function which can be obtained from the survival function since $\boldsymbol{H(t) = -log S(t)}$.Thus, the probability density function of T can be written as:

$$\boldsymbol{f(t) = h(t) e^{-\int_0^t h(u)du}, \quad t \geq 0}$$

These three functions give mathematically equivalent specification of the distributions of the survival time T. if one of them is known, the other two are determined. Survival function is most useful for comparing the survival progress of two or more groups. The hazard function gives a more useful description of the risk of failure at any time point.

4.2 The Cox Proportional Hazard Model

A convenient approach to analysing a data set that consists of several risk factors is the use of regression analysis for survival data. Many different models can be used to relate survival to a collection of risk factors. One of the most frequently used models first produced by D.R. Cox (1972), is called a Cox Proportional Hazard Model usually described as the Cox PH Model.

Most interesting survival analysis research examines the relationship between survival – typically in the form of the hazard function and one or more explanatory variables (or covariates).The Cox PH Model is given by:

$$\boldsymbol{h(t;x) = h_0(t)\exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k\}}$$

where $\boldsymbol{x_1, x_2, x_3 \ldots x_k}$ are collection of independent variables,

$\boldsymbol{\beta_1, \beta_2, \beta_3 \ldots \beta_k}$ are the regression coefficients and

$\boldsymbol{h_0(t)}$ is the baseline hazard at time t.

$\boldsymbol{h_0(t)}$ represents the hazard for a person with the value 0 for all the independent variables.

By dividing both sides of the model by $\boldsymbol{h_0(t)}$ and taking the logarithms,a proportional hazards model can be written in the form:

$$\boldsymbol{ln\left[\frac{h(t)}{h_0(t)}\right] = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k}$$

The model makes no assumptions about the form of $\boldsymbol{h_0(t)}$ [non-parametric part of the model] but assumes a parametric form for the effect of the predictors on the hazard [parametric part of the model].The model is therefore referred to as a semi-parametric model. The attractiveness of the Cox approach is that this ambiguity creates no problems for estimation. Even though the baseline hazard is not specified, one can still get a good estimate for regression coefficients,β.

A measure of the effect of this model is called hazard ratio. The hazard ratio of two individuals with different covariates say X and X^* will be given by:

$$\widehat{\boldsymbol{HR}} = \frac{\boldsymbol{h_0(t)\exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k}}{\boldsymbol{h_0(t)\exp\{\beta_1 x_1^* + \beta_2 x_2^* + \beta_3 x_3^* + \cdots \beta_k x_k^*}}$$

$$= \frac{\boldsymbol{\exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k}}{\boldsymbol{\exp\{\beta_1 x_1^* + \beta_2 x_2^* + \beta_3 x_3^* + \cdots \beta_k x_k^*}}$$

$$= \boldsymbol{\exp\{\widehat{\beta}'(X - X^*)\}}$$

4.3 Parametric Survival Models

The Semi-parametric model (Cox PH) is the most frequently used for analysing prognostic factors in clinical data. This is possibly due to the fact that this model permits us to estimate and make inference about the parameters without assuming any distribution for the survival time. However, these models will not be suitable when the proportional hazard assumption is not justifiable.

The parametric proportional hazard model is the parametric form of the Cox Proportional Hazard model. The parametric PH model is:

$\boldsymbol{h(t;x) = h_0(t)\exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k\}}$ with $\boldsymbol{h_0(t)}$ known or specified.The key difference between the two kinds of models is that the baseline hazard function is assumed to follow a specific distribution when a fully proportional hazard model is fitted for the data whereas the Cox model has no such constraint.

4.3.1 The Exponential Survival Model

The exponential distribution is a one-parameter distribution. Lawless (2003) discovered that statisticians use the exponential distribution to model life data because the statistical methods involved were fairly simple. The probability density function for the exponential distribution is:

$$\boldsymbol{f(t) = \lambda \exp\{-\lambda t\}, t > 0 \qquad where\ \lambda\ is\ the\ scale\ parameter}$$

Under the exponential model, the hazard function of particular patient with covariates $\boldsymbol{x_1, x_2, x_3 \ldots x_k}$ is given by:

$$\boldsymbol{h(t;x) = \lambda \exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k\}}$$

where $\boldsymbol{\lambda}$ is the baseline hazard function for the exponential distribution. This implies that a large value of λ suggests a high risk and short survival while a small value of λ indicates a low risk and a long survival. However, the exponential distribution is limited. This is because it has only one parameter – the scale parameter, λ.By adding a shape parameter to the distribution makes it more flexible and increased ability to fit more kinds of data, which gives rise to the Weibull distribution.

4.3.2 The Weibull Survival Model

The Weibull distribution is an extension of the exponential distribution. It is the generalization of the exponential distribution to include the shape parameter, thus it has both the scale and the shape parameters. The probability density function of the Weibull distribution is:

$$\boldsymbol{f(t) = \lambda p t^{p-1} \exp\{-\lambda t^p\}, \quad t > 0}$$

$$\boldsymbol{where\ \lambda\ and\ p\ are\ the\ scale\ and\ shape\ parameters\ respectively}$$

For the Weibull model, the hazard function of particular patient with covariates $\boldsymbol{x_1, x_2, x_3 \ldots x_k}$ is given by:

$$\boldsymbol{h(t;x) = \lambda p t^{p-1} \exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k\}}$$

where $\boldsymbol{\lambda p t^{p-1}}$ is the baseline hazard function for the Weibull distribution.

4.3.3 The Log-Normal Survival Model

A random variable denoted by X is said to be Log-Normally distributed if $\boldsymbol{\ln(X)}$ is normally distributed.That is, if $\boldsymbol{Y = \ln(X)}$ and $\boldsymbol{Y}$ follows a normal distribution parameterized with mean μ and standard deviation σ,then the distribution of $\boldsymbol{X}$ is said to be Log-Normal.

The probability density function for the Log-Normal distribution is given as:

$$\boldsymbol{f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(lnx - \mu)^2}{2\sigma^2}, \qquad x > 0, \sigma > 0, -\infty < \mu < \infty}$$

Since $\boldsymbol{X}$ is an antilogarithmic function of a normal random variable, some researchers refer to it as AntiLog-Normal distribution.

Under the Log-Normal model, the hazard function of particular patient with covariates $x_1, x_2, x_3 \ldots x_k$ is given by:

$$h(t;x) = \frac{\frac{1}{t\sigma}\phi(\frac{lnt}{\sigma})}{\phi(-\frac{lnt}{\sigma})}\exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \beta_k x_k\}$$

where $\frac{\frac{1}{t\sigma}\phi(\frac{lnt}{\sigma})}{\phi(-\frac{lnt}{\sigma})}$ is the baseline hazard function for the Log-Normal distribution.

5.0 Analysis and Results

5.1 Cox Proportional Hazard Regression R Output

Call:

coxph(formula = Surv(time,status) ~ age + sex + drugabuse + tbtreatment + drugadherence + BMI + regimen + cd4count + staging, data = james)

	coef	exp(coef)	se(coef)	z	p
age	-0.002197	0.998	0.00475	-0.46298	0.6400
sexMALE	-0.000783	0.999	0.11317	-0.00692	0.9900
drugabuseYES	0.344764	1.412	0.11777	2.927450	0.0034
tbtreatmentYES	0.318106	1.375	0.12023	2.645920	0.0290
drugadherencePoor	0.228351	1.257	0.10446	2.185940	0.0081
BMI26-30	-0.234584	0.791	0.15320	-1.53122	0.1300
BMI31-35	-0.010540	0.990	0.18871	-0.05585	0.9600
BMI36-40	0.272451	1.313	0.16431	1.658170	0.0970
BMIbelow 18	-0.102930	0.902	0.13213	-0.77899	0.4400
regimenTDF+3TC+EFV	-0.046602	0.954	0.10292	-0.45278	0.6500
cd4count201-400	-0.469141	0.626	0.21216	-2.21128	0.0270
cd4count401-600	-0.528878	0.589	0.21133	-2.50263	0.0120
cd4count601-800	-0.590686	0.554	0.21924	-2.69423	0.0071
cd4count801-1000	-0.167002	0.846	0.23390	-0.71398	0.4800
cd4countAbove 1000	-0.097146	0.907	0.23890	-0.40664	0.6800
stagingSTAGE2	-0.113678	0.893	0.14910	-0.76242	0.4500
stagingSTAGE3	0.117224	1.124	0.14392	0.81451	0.4200
stagingSTAGE4	-0.019089	0.981	0.14627	-0.13050	0.9000

Likelihood ratio test=46.6 on 18 df, p=0.00024 n= 592, number of events= 405

The appropriate Cox proportional hazard model from the output is:

$$\lambda(t) = \lambda_0(t)\exp\{0.344764DABUSE_{yes} + 0.318106TB_{yes} + 0.228351DA_{poor} - 0.469141C_1 - 0.528878C_2 - 0.590686C_3$$

5.2 Parametric Models R Output (With Known Baseline Hazards)

5.2.1 The Wiebull Model R Output:

Call:

survreg(formula = Surv(time, status) ~ age + sex + drugabuse + tbtreatment + drugadherence + BMI + regimen + cd4count + staging, data = james, dist = "weibull")

	Value	Std. Error	Z	P
(Intercept)	3.12	0.09870	31.5930	<0.01
age	0.0000914	0.00142	0.0644	0.949
sexMALE	-0.0131	0.03373	-0.3871	0.699
drugabuseYES	-0.132	0.03479	-3.8031	0.000143
tbtreatmentYES	-0.0959	0.03610	-2.6573	0.00788
drugadherencePoor	-0.0892	0.03088	-2.8875	0.00388
BMI26-30	0.0611	0.04593	1.3308	0.183
BMI31-35	-0.0101	0.05648	-0.1796	0.857
BMI36-40	-0.103	0.04940	-2.0804	0.375
BMIbelow 18	0.0121	0.03954	0.3049	0.760
regimenTDF+3TC+EFV	0.00126	0.03093	0.0407	0.968
cd4count201-400	0.177	0.06325	2.8001	0.00511
cd4count401-600	0.193	0.06322	3.0513	0.00228
cd4count601-800	0.229	0.06547	3.4948	0.000474
cd4count801-1000	0.0555	0.06980	0.7958	0.426
cd4countAbove 1000	0.0337	0.07135	0.4729	0.636
stagingSTAGE2	0.0298	0.04471	0.6674	0.504
stagingSTAGE3	-0.0391	0.04290	-0.9110	0.362
stagingSTAGE4	0.00374	0.04391	0.0851	0.932
Log(scale)	-1.21	0.03877	-31.2079	<0.001

Scale= 0.298

Weibull distribution

Loglik(model)= -1440.5 Loglik(intercept only)= -1474.1

Chisq= 67.23 on 18 degrees of freedom, p= 1.3e-07

Number of Newton-Raphson Iterations: 5

n= 592

The appropriate Weibull Proportional Hazard Model from the analysis is fitted as:

$$\lambda(t) = \lambda p(\lambda t)^{p-1}\exp\{-0.132DABUSE_{yes} - 0.0959TB_{yes} - 0.0892DA_{poor} + 0.177C_1 + 0.193C_2 + 0.229C_3\}$$

5.2.2 The Exponential Model R OutpuT:

Call:

survreg(formula = Surv(time, status) ~ age + sex + drugabuse + tbtreatment + drugadherence + BMI + regimen + cd4count + staging, data = james, dist = "exp")

	Value	Std. Error	Z	P
(Intercept)	3.0518	0.32154	9.491	<0.01
age	0.0028	0.00476	0.588	0.557
sexMALE	0.0428	0.11182	0.382	0.702
drugabuseYES	-0.1350	0.11575	-1.166	0.244
tbtreatmentYES	-0.2073	0.11902	-1.742	0.0815
drugadherencePoor	-0.0882	0.10181	-0.866	0.386
BMI26-30	0.1621	0.15276	1.061	0.289
BMI31-35	0.0511	0.18670	0.274	0.784
BMI36-40	-0.1057	0.16154	-0.654	0.513
BMIbelow 18	0.1053	0.13069	0.806	0.420
regimenTDF+3TC+EFV	0.0483	0.10123	0.477	0.633
cd4count201-400	0.1987	0.21093	0.942	0.346
cd4count401-600	0.2767	0.20990	1.318	0.188
cd4count601-800	0.2607	0.21630	1.205	0.228
cd4count801-1000	0.1026	0.23083	0.444	0.657
cd4countAbove 1000	0.0417	0.23713	0.176	0.860
stagingSTAGE2	0.0714	0.14639	0.488	0.626
stagingSTAGE3	-0.0358	0.14170	-0.252	0.801
stagingSTAGE4	0.0235	0.14416	0.163	0.871

Scale fixed at 1

Exponential distribution

Loglik(model)= -1723.2 Loglik(intercept only)= -1730

Chisq= 13.57 on 18 degrees of freedom, p= 0.76

Number of Newton-Raphson Iterations: 3

n= 592

5.2.3 The Log-Normal Model

Call:

survreg(formula = Surv(time, status) ~ age + sex + drugabuse +tbtreatment + drugadherence + BMI + regimen + cd4count + staging,data = james, dist = "lognormal")

	Value	Std. Error	Z	P
(Intercept)	2.82563	0.10304	27.4235	<0.01
age	0.00115	0.00152	0.7519	0.452
sexMALE	0.03340	0.03564	0.9371	0.349
drugabuseYES	-0.10790	0.03687	-2.9268	0.00342
tbtreatmentYES	-0.09704	0.03882	-2.4999	0.0124
drugadherencePoor	-0.06437	0.03236	-1.9893	0.00467
BMI26-30	0.06415	0.04807	1.3344	0.182
BMI31-35	0.00885	0.05968	0.1483	0.882
BMI36-40	-0.08023	0.05228	-1.5347	0.125
BMIbelow 18	0.02978	0.04141	0.7193	0.472
regimenTDF+3TC+EFV	0.00745	0.03210	0.2320	0.817
cd4count201-400	0.16293	0.06825	2.3872	0.0170
cd4count401-600	0.20330	0.06758	3.0083	0.000263
cd4count601-800	0.24353	0.06959	3.4995	0.000466
cd4count801-1000	0.06625	0.07393	0.8961	0.370
cd4countAbove 1000	0.02966	0.07636	0.3884	0.698
stagingSTAGE2	0.05255	0.04589	1.1451	0.252
stagingSTAGE3	-0.00135	0.04526	-0.0298	0.976
stagingSTAGE4	0.02831	0.04574	0.6189	0.536
Log(scale)	-1.02531	0.03587	-28.5836	<0.01

Scale= 0.359

Log Normal distribution

Loglik(model)= -1417.7 Loglik(intercept only)= -1446.2

Chisq= 56.95 on 18 degrees of freedom, p= 6.3e-06

Number of Newton-Raphson Iterations: 4

n= 592

6.0 Discussion of Results and Conclusion.

The Weibull parametric proportional hazard model (with the baseline hazard specified to follow the Weibull distribution) shows similar results with that of the Cox Proportional Hazard model. This is traceable to the fact that the Cox model depicts the shape of the distribution. However, since this happened with a higher variability, the variances (and consequently the standard errors) of the Cox model became higher than that of the Weibull model.

When the exponential model was fitted, none of the covariates was observed to have significant effect on the mortality of the HIV/AIDS infected patients. This implies that the exponential model would not be appropriate for the analysis of the data. Results from the Log-Normal model approximates that of the Weibull model. As part of the study, the Cox Proportional Hazard Model showed an increased risk of mortality among patients who had history of substance abuse (they were 1.412 times likely to die than those who do not abuse drugs and p-value = 0.0034).That is, abusers of drugs have an increased probability of 41.2% of death.

Patients that were positive to tuberculosis had a hazard rate of 1.375 times than those who were not on tuberculosis treatment (an increased chance of 37.5% of death).Thus tuberculosis treatment status was a significant predictor of mortality of HIV patients (p = 0.0081).Adherence to drugs is seen to play a part in the mortality of HIV infected patients (p = 0.0290).Subjects with poor adherence to medication tend to be 1.257 times likely to die (an increased probability of 25.7% of dying) than those with good medication habit.

From the analysis, patients with cell counts less than 801 die faster than those with T-helper cells less than 200. The result also showed that higher cell counts of above 801 proves a low probability of experiencing early death. On the other hand, it can be inferred that age, sex, regimen type, WHO clinical staging and body mass index (BMI) of patients were not significant predictors of mortality of the HIV patients.

This study therefore leaves us with two conclusions:

1. The parametric method of survival analysis will be good and probably preferred if the correct distribution is identified.
2. Drug abuse history of patients, tuberculosis status, patients' adherence to drugs as well as the patients' CD4 cell counts are prognostic factors for mortality experience of HIV/AIDS patients.

References

[1]Berkerson,J and Gage R.D(1950):Calculation of Survival rates for Cancer, *Proc Staff Meet Mayo Clin[jour*], 270-286

[2]Bernard Rosner (2010*):*Fundamentals of Biostatistics,7th Edition, Harvard university,759 – 802

[3]Cox,D. R (1972):Regression Models and Life Tables, *Journal of Royal Statistical Society*,Series B (34),187-220

[4]Crum N.F,Riffenburgh R.H,Wegner S,Agan B.K,Tasker S.A,Spooner K.M,Armstrong A.W,Fraser S and Wallace M.R (2006):Comparison of causes of death and mortality rates among HIV-infected persons:Analysis of the Pre-,Early and Late HAART eras, *Journal of Acquired Immune Deficiency Syndromes*,41(2),194 - 200

[5]Ibrahim J. G,Chen M.H and Sinha D (2001);Bayesian Survival Analysis,*Springer-Verlag*, New York

[6]Kaplan E. and Meier P. (1958):Non-Parametric Estimation from incomplete observations, *Journal of American Statistical Association* 53,457-481

[7]Kelembaum, D.G (1996):Survival Analysis:A self learning text,*Springer*,New York

[8]Krishnamorthy,K.(2006):Handbook of Statistical Distributions with Application,*University of Louisiana, Lafayette, Taylor and Francis group*, USA

[9]Lawless,J.(2003):Statistical Models and Methods for Lifetime data.*Hoboken*, John Wiley and Sons.

[10]Mocroft A,Kirk O,Barton S.E,Dietrich M,Proenca R,Colebunders R,Pradier C,Lundgren J.D,Ledergerber b and Monforte A.(2003):Anaemia is an independent predictive marker for clinical prognosis in HIV infected patients across Europe, EuroSIDA study group,AIDS,13(8),943 – 50

[11]Moore D.M,Hogg R.S,Craib K,Yib B,Wood E.and Montaner J.S.G (2006):CD4 percentage as an independent predictor of survival of patients starting ART with absolute CD4 cell counts between 200 and 350 cells/µl,*HIV medicine*

[12]Pierre D,Etard J.F,Ecochard R,Diouf A,Dieng A.B,Cilote V,Ndiaye I. and Delaporte E.(2008):Change over time of mortality predictors after HAART initiation in a Senegalese Cohort,*European Journal of Epidemiology*, Springer, 23(3)

[13]Sethi,A.K and Gange,S.J(2009),Parametric models for studying time to Antiretroviral Resistance associated with illicit drug use,*Wiscousin Medical Journal*

[14]Sibhatu B,Reda A.A and Tesfay D(2012),Predictors of mortality among HIV infected patients taking Antiretroviral treatment in Ethiopia, a retrospective cohort study, *AIDS Research and therapy*,9(15)

[15]Weiss,R.A(1993):"How does HIV cause AIDS?" *Science*, 260(5112), 1273 – 9,
url:http://www.en.wikipedia.com/hiv (cited on 9th September, 2014)

[16]Zhu H.P,Xia X,Yu C.H,Adnan A,Liu S.F and Du Y.K (2011):Application of Weibull model for survival of patients with gastric cancer,*Journal of Public Medicine*,11(1)