COX PROPORTIONAL HAZARD MODEL TO THE SURVIVAL STATUS OF HIV INFECTED PATIENTS

M. A. Ijomah and P. K. Ofoka

Department of Mathematics/Statistics
University of Port Harcourt, Choba, Rivers State.
maxwell.ijomah@uniport.edu.ng, priskenny1@yahoo.com

Received: 21-03-19
Accepted: 12-04-19

ABSTRACT

The paper is on the application of Cox proportional hazards model with a view to identify the factors that affect the survival and influence death status of HIV infected patients. A secondary data from a retrospective cohort study based on test from HIV infected patients on antiretroviral treatment in the University of Port Harcourt Teaching Hospital was used. The result of the analysis showed that out of 330 registered HIV patients, 56 (34.2%) died during the study period and 274 (65.8%) were censored. Cox hazard regression analysis showed that the covariate HB and CD4 count were significant risk factors associated with survival in HIV infected patients.

Keywords: Cox regression model, Hazard ratio, Non-proportional hazards, Kaplan-Meier Estimator, Log-Rank Test, Survival Function.

INTRODUCTION

Survival analysis is generally defined as a set of statistical methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest (Klembaum, 1996). It examines and models the time it takes for event to occur and therefore called time to event analysis. The event can be occurrence of a disease, death, marriage, etc. The time to event or survival time can be measured in hours, days, weeks, months or years. Survival analysis focuses on the distribution of survival times. Although there are well known methods for estimating unconditional survival distributions, most interesting survival modeling examines the relationship between survival and one or more predictors, usually termed covariates in the survival-analysis literature (Cox, 1972). The idea of the model is to define a hazard level as a dependent variable which is explained by the time-related component (so-called baseline hazard) and the covariates-related component. The model is based on several restrictive assumptions one of which is the assumption of proportional hazard.

Cox proportional hazard model is one of the most common methods used in the analysis of time to event data. The idea of the model is to define a hazard level as a dependent variable which is explained by the time-related component (so-called baseline hazard) and the covariates-related component. The model is defined as follows:

$$\lambda(t,x) = \lambda_0(t)\exp(\beta x) \quad (1)$$

where:

Scientia Africana, Vol. 18 (No. 1), June, 2019. Pp 82-108
© Faculty of Science, University of Port Harcourt, Printed in Nigeria ISSN 1118 – 1931

(t, x) is hazard function that depends on timepoint t and vector of covariates x,

λ_0 (t) is baseline hazard function that depends on time only,

exp(βx) is covariates-related component.

Cox model is based on several restrictive assumptions. One of them is the assumption of proportional hazard that the name of the model refers to and which results directly from the model formula as follows:

$$HR = \frac{\lambda(t, x_1)}{\lambda(t, x_2)} = \frac{\lambda_0(t)\exp(\beta x_1)}{\lambda_0(t)\exp(\beta x_2)} = \frac{\exp(\beta x_1)}{\exp(\beta x_2)} = \exp[\beta(x_1 - x_2)] \quad (2)$$

where: HR is hazard ratio,

x_1 is vector of covariates of subject I,

x_2 is vector of covariates of subject II.

The assumption states that the hazard ratio for two subjects who are characterized by different sets of covariates depends only on the values of these covariates and does not depend on time. In other words, the hazard ratio is constant over time which means that the effect of a given covariate on a hazard level is the same at all timepoints.

In the hazard ratio model, the resulting value is no longer time-dependent so that the ratio of the two objects remains at all times proportional hazard. The proportional Cox regression model emphasizes the fulfillment of the proportional hazard assumption which means that the ratio between one individual hazard functions and another individual hazard functions is constant.

Several works have been done by researchers on Cox proportional hazard model (Song etal., 2001; Spotswood etal., 2004; Emel, 2007; Ray etal., 2009; Royston and Parmar, 2011;Gurpit etal., 2013;Tolosie and Sharma, 2014; Folorunso etal., 2015; Mustapha and Lawal, 2015; Chatree etal.,2016;Fan, 2017;Chandra and Akansha, 2017). However, biomedical researchers tend to choose semi-parametric methods to model time to event data. In particular, Sethi et Al. (2009) analyzed data from a prospective cohort study of 195 adults receiving HIV/AIDS care and highly active antiretroviral therapy in Baltimore; they were followed for 1188 visits between February 2000 and December 2001 using Kaplan-Meier estimation, Cox and Weibull regressions. Results showed that illicit drug users experienced a greater hazard of clinically significant antiretroviral resistance as compared to non-users. Also, in a study by Gran et al. (2010), they argued that when estimating the effect of treatment on HIV using longitudinal data, standard methods may produce biased estimates due to the presence of time-dependent confounders. Such confounding can be present when a covariate, affected by past exposure, is both a predictor of the future exposure and the outcome. They gave an example of CD4 cell count, being a marker for disease progression for HIV patients, but also a marker for treatment initiation and influenced by treatment. Fitting a marginal structural model (MSM) using inverse probability weights is one way to give appropriate adjustment for this type of confounding.

Based on these reviews, the event of interest in this study is death from HIV infected patients who have started treatment and followed up for the outcomes of either the event (death) or censored which may result as a loss of follow up or withdrawal from treatment. The aim is therefore to use Cox proportional hazard model in a cohort of HIV infected patients to determine survival status of patients based on the age, sex, blood test (i.e. level of hemoglobin

(HB),Cluster of differentiation 4 (CD4) cells and white blood cell (WBC).

This paper outlines in section 2 a brief literature on Kaplan Maier estimation, Section 3 has materials and method such as survival analysis, proportional hazard model, estimated Cox proportional hazard model and data collection. In Section 4, we illustrate the performance of our models using HIV/AIDS data set from HIV infected patients on antiretroviral treatment in the University of Port Harcourt Teaching Hospital. Section 5 concludes the paper with a summary and some possible areas of extension.

Kaplan-Meier Estimator

The Kaplan-Meier estimators of the survival function (due to Kaplan and Meier, 1958) is a major step in the development of suitable models for medical data that are related to time till incident. Most assessments are made conditional on the knowledge of the present situation of the patients at the time of the analysis, and subject to changes over time. Usually, as the population under study is changing, we only consider the individual risk to get recurrent stroke for those who are had the first one. The Kaplan-Meier method (KM), also called Product-limit method, is a nonparametric model function (Kaplan and Meier, 1958). It is the most widely used model in estimating survival functions. When there is no censored data, the KM estimate is actually the sample proportion of subjects surviving longer than the time. Censoring in survival analyses occurs when we have some information about individual survival time, but we do not know the survival time exactly. This happens because of the possibility that a person does not experience the event before the study ends or he has been lost to follow-up or withdraws during the study period.

Then the Kaplan-Meier estimator of the survival function at time t is given by:

$$\hat{S}(t) = \prod_{j-1}^{r}\left(\frac{\pi_j - d_j}{n_j}\right) t(k) \le t(k+1)$$

$$j = 1,2,.....,r\ \hat{S}(t) = 1 \quad t < t_1 \quad \text{for} \qquad (3)$$

where, n_j= the number of individuals who are at risk of dying at time t_j and $d_j$is the number of individuals who failed (died) at time t_j. The variance of Kaplan-Meier survival estimator is estimated using Greenwood's formula (Collett, 2003) given as

$$Var(\hat{S}(t) = (\hat{S}(t))^2 \sum_{j=1}^{r} \frac{d_j}{n_j(n_j - d_j)} \qquad (4)$$

Cox proportional hazards model:

The Cox proportional hazards (PH) model is a semi-parametric method of analyzing the effects of different covariates on the hazard function. It is the most widely used for the analysis of survival data in the presence of covariates or prognostic factors due to its simplicity and not being based on any assumptions about the survival distribution. The model is given by:

$$h_i(t/x) = h_0(t)\exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}) = h_0(t)\exp(\beta_i x_i) \qquad (5)$$

where,

$h_i(t/x)$is the hazard function of individual i given x at time t.

$h_0(t)$is called the baseline hazard function.

$x = (x_{1i}, x_{2i}, \dots, x_{ki})$is the values of the vector of explanatory variables for a particular individual.

$\beta = (\beta_1, \beta_2, \dots, \beta_k)$is a vector of regression coefficients which can be estimated using Cox's partial likelihood estimator without specifying and estimating the baseline hazard $h_0(t)$.

Scientia Africana, Vol. 18 (No. 1), June, 2019. Pp 82-108
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

The Survival Function

The object of primary interest is the survival function, conventionally denoted S, which is defined $S(t) = \Pr(T \geq t)$ (6)

where t is some time, T is a random variable denoting the time of death, and Pr refers to probability. That is, the survival function is the probability that the time to death is later than some specified time t. The lifetime distribution function, conventionally denoted F, is defined as the complement of the survival function,

$$F(t) = \Pr(T \geq t) = 1 - S(t) \quad (7)$$

If F is differentiable then the derivative, which is the density function of the lifetime distribution, is conventionally denoted f

$$f(t) = F'(t) = \frac{\partial}{\partial t} F(t)$$

The survival function gives the probability of surviving or being event-free beyond time t. Because $S(t)$ is a probability, it is positive and range from 0 to 1. It is defined as $S(0) = 1$ and as t approaches ∞, $S(t)$ approaches 0.

The hazard function, is defined as the event rate at time t conditional on survival until time t or later,

$$\lambda(t) = \lim_{\partial t \to 0} \frac{\Pr(t \leq T < t + \partial t)}{\partial t . S(t)} = \frac{F(t)}{S(t)} = -\frac{S'(t)}{S(t)} \quad (8)$$

It can alternatively be defined as;

$$\lambda(t) = -\log S(t)$$

Future lifetime at a given time t_0 is the time remaining until death, given survival to age t_0. Thus, it is $T - t_0$ in the present notation. The expected future lifetime is the expected value of future lifetime. The probability of death at or before age $t + t_0$, given survival until age t_0, is;

$$\Pr(T_{i,1} \leq t_0 + \tfrac{t}{T} > t_0) = \frac{\Pr(t_0 < T \leq t_0 + t)}{\Pr(T > t_0)} = \frac{F(t + t_0) - F(t_0)}{S(t_0)} \quad (9)$$

Therefore, the probability density of future lifetime is;

$$\frac{\partial}{\partial t} = \frac{F(t + t_0) - F(t_0)}{S(t_0)} = \frac{f(t + t_0)}{S(t_0)} \quad (10)$$

And the expected future lifetime is given as;

$$\frac{1}{S(t_0)} \int_0^\infty t f(t + t_0) \partial t = \frac{1}{S(t_0)} \int_0^\infty S(t) \partial t \quad (11)$$

MATERIALS AND METHODS

A retrospective cohort study based on test from HIV infected patients on antiretroviral treatment in the University of Port Harcourt Teaching Hospital (UPTH) and was followed up till the outcome of the event or censored. The follow up was for four visits.

Variables in the Study

The response variable in this study is the survival time (measured in months) from the date the treatment starts until the occurrence of an event (death, denoted 1) or alive or censored (denoted 0). The predictor variables also called covariates in this study are those variables that are assumed to influence the survival of HIV infected patients. The covariates are gender, age, Cluster of differentiation 4 (CD4) cells,

levels of hemoglobin (HB), and white blood cells (WBC). The response variable in this study is the survival time (measured in months) from the date the treatment starts until the occurrence of an event (death, denoted 1) or alive or censored (denoted 0).The predictor variables also called covariates in this study are those variables (test) that are assumed to influence the survival of HIV infected patients. The covariates are gender, age, Cluster of differentiation 4(CD4) cells, levels of hemoglobin (HB), and white blood cells (WBC).

RESULTS

The formation of Cox proportional hazard model was conducted to determine the relationship between survival time and the variables suspected to influence survival time. Estimation of Cox proportional hazard model parameters was done using Breslow method approach. .

The study indicated that from among 330 HIV infected patients, 56 died while 274 were censored. The percentage of censored among the patients was 83.03%. This is as shown in table 1 below.

Table 1: Summary of the Number of Event and Censored Values

Total	**Event**	**Censored**	**Percentage censored**
330	56	274	83.03

Table 2 presents tests of the Global Null Hypothesis that all the mod are equal to zero. The P-value of the Likelihood Ratio, Wald, and score chi-square statistics at the bottom of the output are asymptotically equivalent tests of the omnibus null hypothesis that all of the β's are zero. In this instance, the test statistics are in close agreement, and the hypothesis is soundly rejected.

Table 2: Testing Global Null Hypothesis: BETA=0

Test	**Chi square**	**DF**	**P- value**
Likelihood Ratio	63.4126	15	< 0.0001
Score	65.6891	15	< 0.0001
Wald	40.6174	15	0.0004

In determining whether there is significant difference among different groups of the covariates, we employed the log rank test of equality as shown in Table 3. The test indicates that HB and CD4 count show a significant difference ofsurvival among the patients. However, covariates including Sex, Age and WBC are insignificant.

Table 3: Test of equality using log rank

Wald Effect	**DF**	**Chi-Square**	**P-value**
Sex	1	2.1447	0.1431
Age	3	3.6956	0.2963

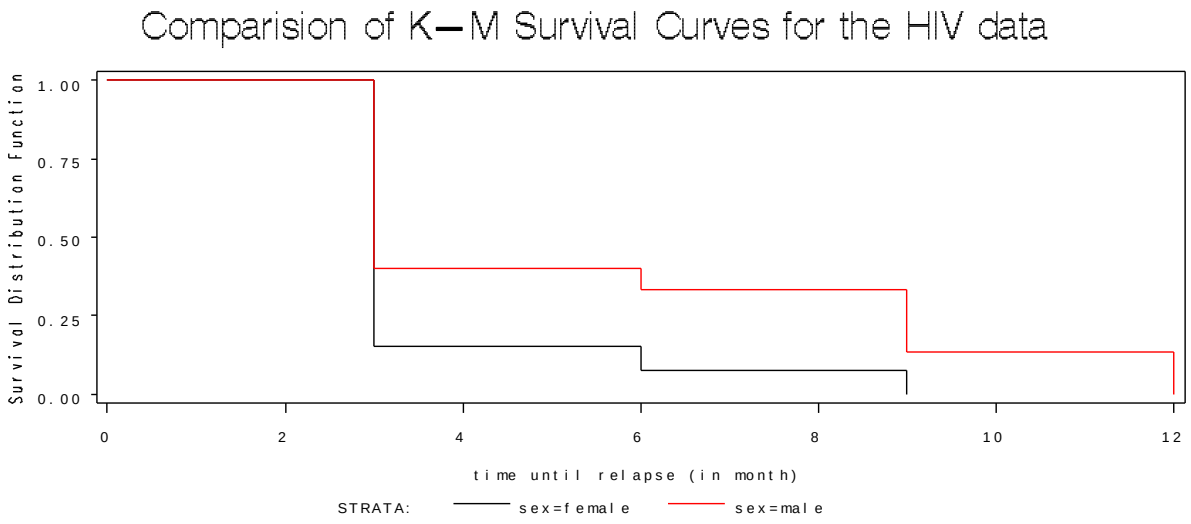
Scientia Africana, Vol. 18 (No. 1), June, 2019. Pp 82-108
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

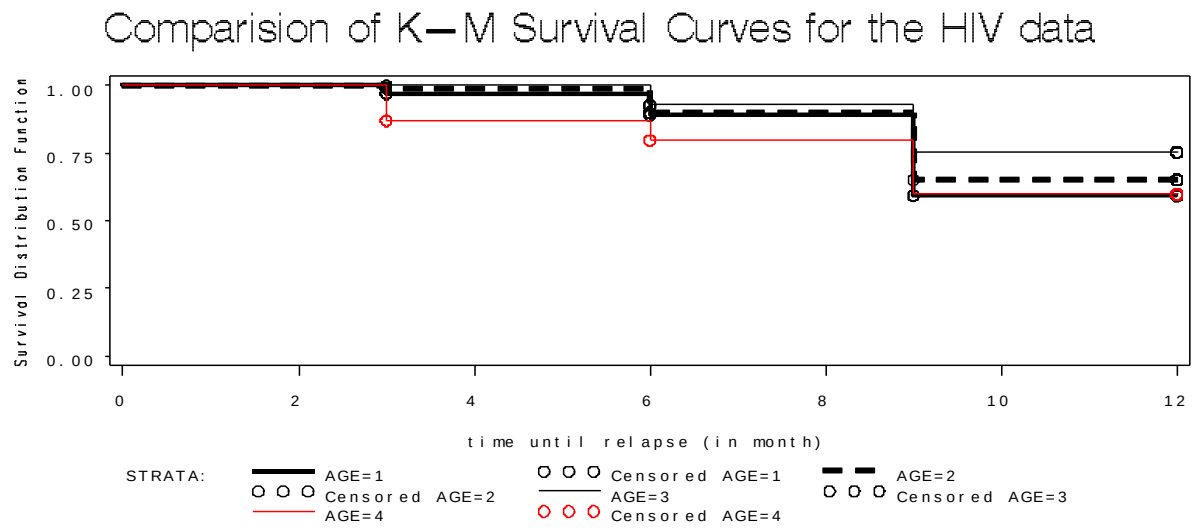
HB	4	18.3940	0.0010
CD4	4	17.9699	0.0013
WBC	3	1.6402	0.6503

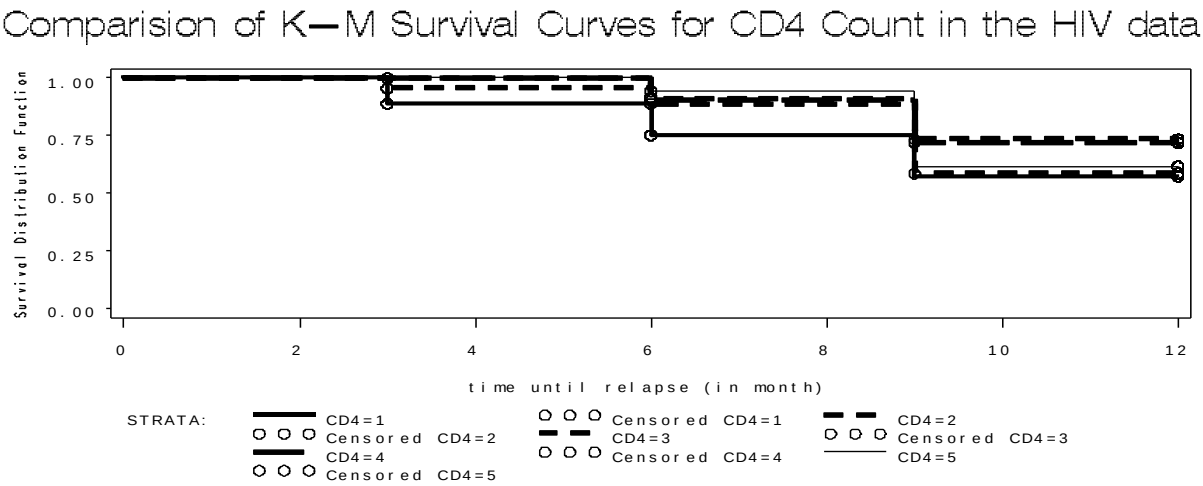
The Cox PH model is fitted to the data and the results are summarized in table 4 below, giving the estimators of hazard ratios for each covariates and their confidence intervals and its p-value from the likelihood ratio test. Each covariate is classified in such a way to capture the most sensitive group of each covariate.

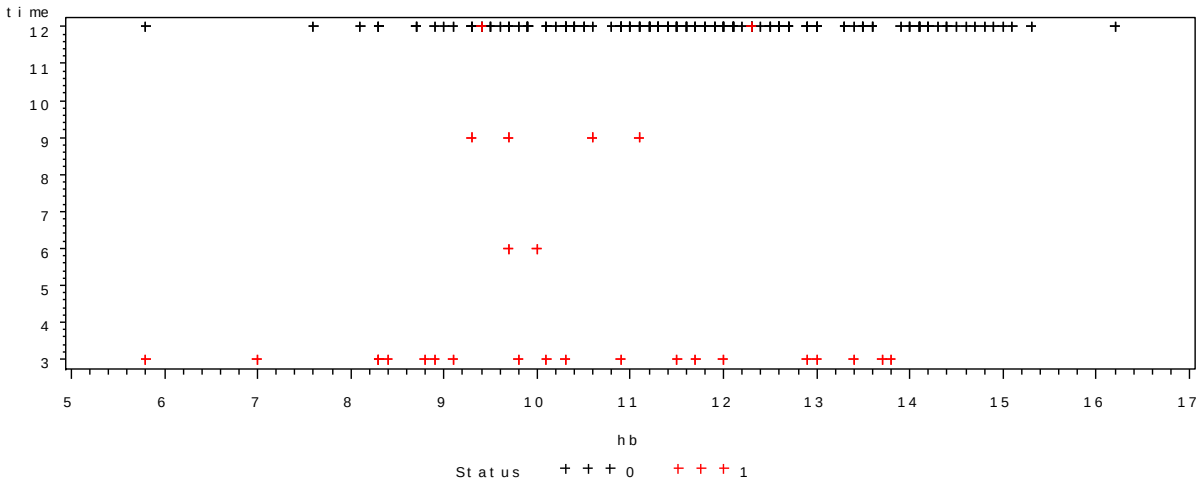
Table 4: Parameter Estimation Results Cox Proportional Hazard Model of HIV/AIDS Status.

Parameter	Group	Df	Parameter Estimate	Standard Error	Ch-Square	P-value	Hazard Ratio
Sex	Male	1	-0.45604	0.31140	2.1447	0.1431	0.634
Age	18 – 27	1	-0.95915	0.57027	2.8288	0.0926	0.383
Age	28 – 37	1	-0.60140	0.43411	1.9192	0.1659	0.548
Age	38 +	1	-0.35181	0.41051	0.7345	0.3914	0.703
HB	5.1 -7.0	1	1.05774	0.59607	3.1489	0.0760	2.880
HB	7.1 – 9.0	1	0.50775	0.40396	1.5798	0.2088	1.662
HB	9.1 – 13.0	1	-0.84503	0.35605	5.6329	0.0170	0.430
HB	13.1+	1	-1.09521	0.47981	5.2101	0.0225	0.334
CD4	5 – 100	1	-1.11657	0.64104	3.0339	0.0815	0.327
CD4	101 – 300	1	-0.93979	0.42108	4.9813	0.0256	0.301
CD4	301 – 500	1	-2.18399	0.52298	17.4396	< 0.001	0.113
CD4	501+	1	-16.9672	99.4163	0.0003	0.9864	0.001
WBC	1.0 – 3.9	1	14.81525	208.80	0.0001	0.9943	0.0001
WBC	4.0 – 6.9	1	15.24223	2088	0.0001	0.9942	0.7080
WBC	7.0+	1	-0.34461	5486	0.0000	0.9999	1.0416




Comparision of K–M Survival Curves for the HIV data
Survival Distribution Function
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10
12
time until relapse (in month)
STRATA:
AGE=1
Censored AGE=2
AGE=4
Censored AGE=1
AGE=3
Censored AGE=4
AGE=2
Censored AGE=3


Comparision of K–M Survival Curves for CD4 Count in the HIV data
Survival Distribution Function
1.00
0.75
0.50
0.25
0.00
0
2
4
6
8
10
12
time until relapse (in month)
STRATA:
CD4=1
Censored CD4=2
CD4=4
Censored CD4=5
Censored CD4=1
CD4=3
Censored CD4=4
CD4=2
Censored CD4=3
CD4=5


time
12
11
10
9
8
7
6
5
4
3
5
6
7
8
9
10
11
12
13
14
15
16
17
hb
Status
0
1

Scientia Africana, Vol. 18 (No. 1), June, 2019. Pp 82-108
© Faculty of Science, University of Port Harcourt, Printed in Nigeria ISSN 1118 – 1931

DISCUSSION

As shown in the result above, female has decrease hazards (36%) as compared with males. The model assumes that the estimated hazard ratio of 0.634 is the same at each and every point during follow up and all combinations of the other covariates. HB and CD4 counts are the covariates that show statistically significant impact on the survival status of HIV patients at the level of 5%, though age between 18 – 27 years is also significant at 10% level. It is estimated that individual aged between 18 – 27 years at diagnosis experience 0.383 times higher risk of death due to HIV/AIDS than individuals 0-17 years at diagnosis. That is increase in age between 18-27 years, has a decrease hazard of 61% decreases. Similarly, individuals aged 28 – 37 at diagnosis have an estimated has a hazard decrease of approximately 30%, 38 – 47 also has hazard decrease of 45%. In all, hazard risk increases with increase in age. For HB, HB level between 5.1-7.0 indicates that patients within this group has 74% chance of survival status to HIV/AIDS treatment. Also for HIV patients with HB level between 7.1– 9.0, the above hazard ratio of 1.662 reveals that sampled patients has 66.2% risk of death increase every year. For HIV patients with HB level between 9.1 – 13.0 with hazard ratio 0.430 indicates that AIDS progression hazard decreased as HB increases while HB level above 13 has the least hazard ratio of 0.334, which implies that it has 67% hazard decrease.The estimated hazards ratio (HR) for a 5 up to 100 cells/mm^3 increase in the baseline CD4 cell count was 0.327, between 101 up to300 cells/mm^3has hazard ratio of 0.301, between 301 up to 500 cells/mm^3 has hazard ratio of 0.113 while from 500 above has hazard ratio of 0.001 showing that the hazard rate for patients, whose CD4 cell count was higher by 500 cells/mm^3 was about 99% lower than for patients in the next lower category.

The estimated ratio for WBC between 1.0 - 3.9, 4.0 -6.9 and 7.0 and above is (0.0001, 0.7080,1.0416) respectively. This result reveals patients with WBC within 1.0 and 3.9 are above 99% lower than for patients in the next lower category. Also, WBC between 4.0 – 6.9 reveals a hazard decrease of approximately 30%. For WBC from 7 and above, the hazard ratio is 1.0416 indicating that sampled patients have 4.16% risk of death increase every year.

The study is on the application of Cox proportional hazards model with a view to identify the factors that affect the survival and influence death status of HIV infected patients. A secondary data from a retrospective cohort study based on test from HIV infected patients on antiretroviral treatment in University of Port Harcourt Teaching Hospital was used. The result of the analysis showed that out of 330 registered HIV patients, 56 (34.2%) died during the study period and 274 (65.8%) were censored. Univariate Cox Proportional Hazards regression models were developed to assess the relation between each covariate survival status and their selected variables. The result of Cox proportional hazards regression model showed that Patients who had been in age group 18 up to 27 years lived longer than patients who had been in age above 27 years. CD4 cell count was also a significant predicator of survival time in HIV/AIDS patients. The increment of patient CD4 cell count had a significant effect on favorable survival time.The results of Kaplan-Meier and log-rank test showed that patients who had: baseline CD4 count 300 cells/μL and above had better survival time compared with reference groups For the analysis of level of HB, lower HB less than 7.0 has higher hazard risk compared to

Hb level above 13.In the other hand, it was found that factors which had no significant impact on the survival of HIV patients were gender and white blood cell.

The researchers are grateful to the anti-retroviral (ART) clinic and Department of Chemical Pathology University of Port Harcourt Teaching Hospital for helping us with the data.

REFERENCES

Ajay K.Seth, Stephen .J.Gange (2009). Wiscosin medical journal, parametric models for studying time to antiretroviral resistance associated with illicit drug use.

Anderson Per Kragh and Keiding Niels (2014). Survival analysis overview.Chatree Chai- Adisaksopha, Alfonso Lorio, Christopher Hillis, Wendy Lim and Mark Crowther (2016). A systematic review of using and reporting survival analysis in acute lymphoblastic leukemia literature. BMC hematology, vol. 16, 2016. pmc 4898350

Emel Basar (2007). Non-proportional hazards with application to kidney transplant data. Commun.fac.sci.univ.Ank.series Al. volume 55,2, pg. 55-63. ISSN:1303-5901.

Gulprit Grover, V. Sree Nivas, Sudeep Khanna, Divya Seth (2013). Estimation of survival of

liver cirrhosis patients, in the presence of prognostic factors using accelerated failure timemodel as an alternative to proportional hazard model. International journal of statistics and applications 2013, volume 3, No 4, pages 113-122.

Gran et al Song et.al (2001), Spotswood et.al (2004),, Ray et.al (2009), Royston and Parmar (2011),Tolosie and Sharma (2014), Folorunso et.al (2015), Mustapha and Lawal (2015), Chatree et.al (2016),Fan (2017),Chandra and Akansha (2017), Sethi, et. al, (2009),