



APPLICATION OF COX PROPORTIONAL HAZARDS MODEL IN TIME TO EVENT ANALYSIS OF HIV/AIDS PATIENTS

*¹Apeagee, B. B., ²Agada, P. O., ³Dzaar, D. A. and ¹Ede, A. A.

¹Benue State School of Nursing, Makurdi

²Department of Mathematics/Statistics/Computer science, Federal University of Agriculture Makurdi

³Department of Statistics, Fidei Polytechnic, Gboko

*Corresponding Author's Email: paixkristi@gmail.com

ABSTRACT

The Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS) remains a public health crisis that has contributed to the majority of deaths recorded in the past decade, affecting Nigeria and other countries of the world as it has become drug resistance in some patients. This study was aimed at estimating the effects of covariates on the survival time for HIV/AIDS patients using the Cox PH model. The KM results indicated that 91 patients were males, out of which 31 experienced the event of interest, and 60 (68.9%) were censored, 209 were females, 65 died due to AIDS, and 144 were censored (68.9%) respectively. The results of the Cox PHM indicated that sex, age, and health of patients are positively associated with death due to AIDS with the associated negative length of survival for HIV/AIDS patients with HR (1.149, 1.235, 1.887, and 1.306) respectively. The study concluded that CD4 cell counts are the only variable or covariate that showed a lower risk of death due to AIDS. The results further stated that patients with high CD4 cell counts have lower risks of death due to AIDS but an increase in survival time considering other factors. The study, therefore recommends that survival analysis should be used to assess the various risk factors and the confounding effects associated with them stressing that a patient's lifestyle should be improved to live healthy as they continue to age older.

Keywords: HIV/Aids, survival, analysis, cox, proportional, hazards, model.

INTRODUCTION

The Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS) remains one of the greatest public health challenges still facing Nigeria and the entire world as it has become drug-resistant in some patients (Agada, *et al.*, 2019). Although, the joint United Nations Programme on HIV/AIDS (UNAIDS, 2014) stated that an estimated 36.9 million people were living with HIV/AIDS worldwide of which 3.2 million were estimated to be in Nigeria, the second largest burden in the world after South Africa.

However, HIV/AIDS reduces the rate of growth particularly in countries seriously affected by the disease (Agada *et al.*, 2019). Hence, the need to model the time until a patient dies from AIDS or HIV-related death after a confirmed positive result is paramount to both policymakers, Government, medical researchers, and public health. This is because the only available therapeutic option for the treatment of HIV is antiretroviral therapy (ART).

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is the time until an event occurs (Kleinbaum and Klein, 2012). Thus, the time variable in survival analysis is usually referred to as the survival time because it gives them the time that an individual has survived over some follow-up period after contracting the virus. The event is usually referred to as a failure because the event of interest is either death, disease, discharge, relapse, incidence, or other negative individual experiences (Kleinbaum *et al.*, 2012). The modeling of time to event data is an interesting area with many applications in diverse areas including but not

limited to medicine, sociology, marketing, and economics (Emmert-Streib and Dehmer, 2019).

For researchers to analyse time to event data, different survival techniques such as Kaplan-Meier, Log-rank test and Cox proportional hazards (PH) model, and parametric models such as exponential, Weibull model, etc. are needed (Khanal *et al.*, 2019). However, Cox PH is one of the most widely used regression models for the analysis of time to event data since the baseline hazard function embedded in this model does not require to follow any known probability distribution (Khanal *et al.*, 2019). This implies that the model does not require to specify probability distribution to quantify the effects of independent variables. On the contrary, the application of the Cox PH to various epidemiological studies frequently involves heterogeneous and time-varying exposures, many years of follow up and confounding by numerous measured and unmeasured risk factors (Moolgavkar, *et al.*, 2018).

The hallmark of the HIV infection is the progressive depletion of a class of lymphocytes called CD4⁺ which plays a very important regulatory role in the immune response to infections and tumours (Dessie, 2014). However, since the immune system of an HIV patient depends on the amount of CD4 cell count present in his/her bloodstream which in turn provides a marker for classifying the clinical stages of HIV patients. As strategies to reduce the burden of HIV-associated tuberculosis (TB) by most researchers both intensified case finding (ICF) and antiretroviral therapy were applied to HIV positive individuals with *CD4 Counts* > 350 *cells/μL* in South Africa (Kufa *et al.*, 2016). The results showed that 80.9% were female, 57.9% were

on ART with the median CD4 count of 562 cells/ μ L. The study concluded that TB incidence was high and associated with low body mass index (BMI). Intensified case finding for TB should be strengthened for all HIV positive individuals regardless of their CD4 counts or ART status.

In describing the various effects of ART on the risk of severe bacterial infections in people with high CD4 Cell counts, Lancet (2017) stated that immediate ART reduces the risk of severe bacterial infections in HIV-positive people with high CD4 cell counts.

MATERIALS AND METHODS

Data: The data collected for the study was a retrospective cohort study. The secondary data utilized were collected from the patients' follow-up records. A total of 300 HIV/AIDS patients receiving treatment every six months at the HIV Counselling and Testing (HCT) unit of Federal Medical Centre (FMC) Makurdi were followed from January 2013 –December 2018.

Variables: The primary outcome variable of interest was survival time after a confirmed diagnosis of HIV. However, some subjects entered the study at different times over the study period. The maximum follow-up time is different for each study participant. The response variable is time. The follow-up time is the number of months between the entry date and the end date, age of subjects at the start of follow-up (in years) and Censor which indicates vital status at the end of the study (1=Death due to AIDS, and 0=Lost to follow-up or alive).

The predictor variables which were assumed to influence the survival of HIV patients in the model are sex, age, CD4 count, and patients' health state.

Methods of Data Analysis

Survival analysis: Since the focus of this study was time to event (time to death due to AIDS), the appropriate method for the study was survival analysis. The study employed both Kaplan Meier (KM) estimator and Cox Proportional Hazards (PH) model for the analysis. The Log-rank test was used to compare survival functions. In this study, KM was used to study the survival pattern of HIV/AIDS patients using KM plots to indicate the shape of the survival distribution. The survival function $S(t)$ gives the probability that a patient survives longer than some specified time t . that is, $S(t)$ gives the probability that the random variable T exceeds the specified time t (Kleinbaum and Klein, 2012). The mathematical computation is represented in the equation (1) below.

$$S(t) = P(T > t) \quad 1$$

Kleinbaum and Klein (2012) further stated that survival function is very fundamental to survival analysis because obtaining survival probabilities for different values of t provided unique summary information from survival data.

The Kaplan Meier (KM) Estimator

The KM method according to Stevenson (2007) is based on individual survival times and assumes that censoring is independent of survival time (that is, the reason an observation is censored is unrelated to the cause of failure). The KM estimator of survival at time t is shown in equation 2. In the equation, $t_{j,j=1,2,\dots,n}$ is the total set of failure times recorded

(with t^+ as the maximum failure time), d_j is the number of failures at time t_j , and r_j is the number of individuals at risk at time t_j .

$$\hat{S}(t) = \prod_{j:t_j \leq t} \frac{(r_j - d_j)}{r_j}, \text{ for } 0 \leq t \leq t^+ \quad 2$$

Log-Rank Test

To describe how to evaluate whether or not KM curves for two or more groups are statistically equivalent, a log-rank test was employed. The log-rank test is a chi-square test that provides an overall comparison of the KM curves being compared (Kleinbaum and Klein, 2012). They further stated that the log-rank test makes use of observed versus expected cell counts over categories of outcomes.

The Cox Proportional Hazards (CPH) Model

The PH model is considered as one of the most popular models used for analysing survival data. It is a semi-parametric model that makes a parametric assumption about the effect of the predictors on the hazard function but makes no assumption regarding the nature of the hazard function $h(t)$ itself (Harrell, 2001). The Cox PH model assumes that predictors act multiplicatively on the hazard function but do not assume that the hazard function is constant (Harrell, 2001; Stel, et al., 2011). Hence, the regression portion of the model is fully parametric. That is, regressors are linearly related to log hazard or log cumulative hazard.

However, in many situations, the form of the true hazard function is either unknown or it is complex, which in turn makes the Cox PH model an advantageous model. Also, one is usually more interested in the effect of the predictors than in the shape of $h(t)$, and the Cox PH model allows the researcher to essentially ignore $h(t)$, which is often not of primary interest. The Cox PH model is usually stated in terms of the hazards function as shown below

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad 3$$

Where $h(t)$ is the expected hazard at time t , $\beta_1, \beta_2, \dots, \beta_p$ are regression coefficients, $h_0(t)$ is the baseline hazard and represents the hazard when all the predictors (independent variables $X_1 X_2 \dots X_p$) are equal to zero. Note that, the predicted hazard $h(t)$ or the rate of suffering the event of interest in the next instant is the product of the baseline hazard $h_0(t)$ and the exponential function of the linear combination of the predictors. Thus, the predictors have a multiplicative or proportional effect on the predicted hazard. Thus, the formula for the Cox PH model can simply be written as;

$$h(t, X) = h_0(t) \exp(\beta_i X_i) \quad 4$$

Where β_i represents the regression coefficient

RESULTS AND DISCUSSION

The statistical package for social sciences (SPSS version 23) was used for the data analysis.

Summary Statistics: Out of the total 300 registered HIV/AIDS

patients, 91 are males (30.3%) and 209 are females (69.7%) respectively (Table 1a). Also, the age group 38 years and above showed a greater percentage with 202 (67.3%) compared to other age groups as shown in Table 1b.

Table 1a: Sex of Patients

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	91	30.3	30.3	30.3
	Female	209	69.7	69.7	100.0
	Total	300	100.0	100.0	

Table 1b: Age of patients in years

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	23-27	7	2.3	2.3	2.3
	28-32	28	9.3	9.3	11.7
	33-37	63	21.0	21.0	32.7
	38 years above	202	67.3	67.3	100.0
	Total	300	100.0	100.0	

KM descriptive survival analysis

Table 2a showed that out of the total 300 registered HIV/AIDS patients receiving treatment at the HCT unit of FMC Makurdi, 91 are males (31 experienced the event of interest and 60 were censored accounting for 65.9%) and 209 were females (65 died due to AIDS and 144 were censored with percentage censored of 68.9%). The overall results showed that 96 HIV patients

experienced the event of interest (Dead due to AIDS) accounting for 32% and 204 (68%) were censored during the period under review. The age group 38 years and above showed a greater percentage associated with death due to AIDS with 69 deaths against 65.8 (censored) more than any other age group (Table 2b).

Table 2a: Case Processing Summary

Sex of Patients	Total N	N of Events	Censored	
			N	Percent
Male	91	31	60	65.9%
Female	209	65	144	68.9%
Overall	300	96	204	68.0%

Table 2b: Case Processing Summary

Age of patients in years	Total N	N of Events	Censored	
			N	Percent
23-27	7	1	6	85.7%
28-32	28	8	20	71.4%
33-37	63	18	45	71.4%
38 years above	202	69	133	65.8%
Overall	300	96	204	68.0%

The mean survival time for male patients was 60.953 (Months) and the median survival time was 62.000 (Months). Similarly, the

mean and median survival time for female patients stood at 62.231 and 64.00 (Months) respectively. Hence, the results further showed that female patients have higher survival time compared to their male counterparts (Table 2c).

Table 2c: Means and Medians for Survival Time

Sex of Patients	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Male	60.953	.540	59.893	62.012	62.000	.779	60.474	63.526
Female	62.231	.365	61.514	62.947	64.000	.817	62.399	65.601
Overall	61.971	.315	61.353	62.589	63.000	.647	61.731	64.269

a. Estimation is limited to the largest survival time if it is censored.

Based on the results in Table 2d, the Log-rank test indicates that there is a statistically significant difference between the sexes of patients in the survival time during the period under review.

Table 2d: Overall Comparisons

	Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)	1.503	1	.220

Test of equality of survival distributions for the different levels of Sex of Patients.

The Figures (1 and 2) showed that there were differences among survival curves of age and sex respectively.

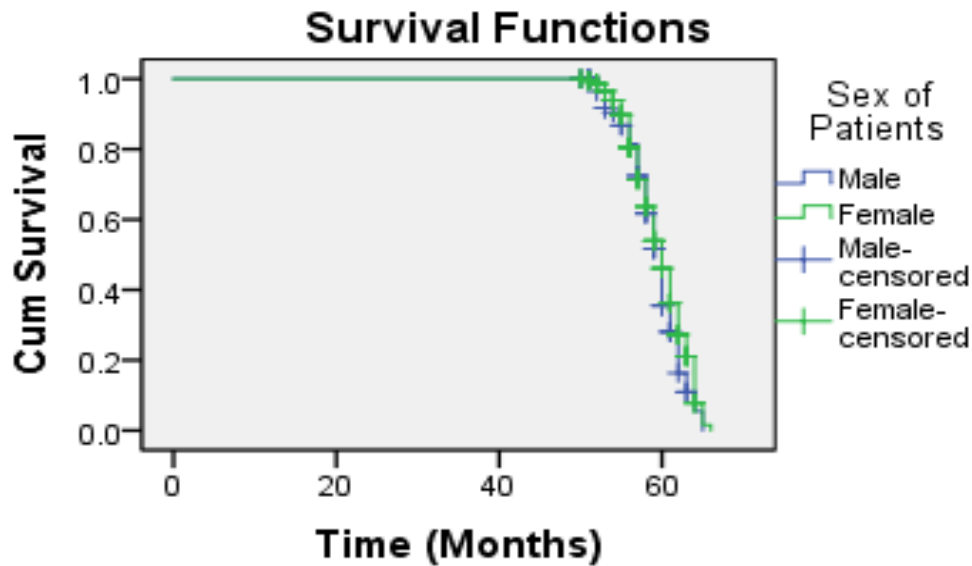


Figure 1 Survival curve of sex

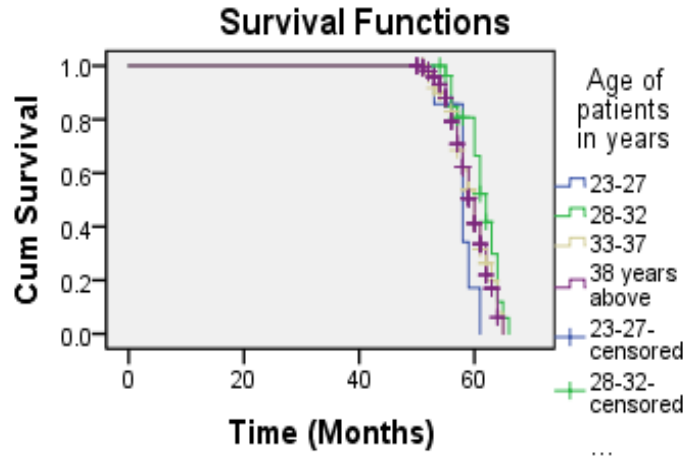


Figure 2 Survival curve of age

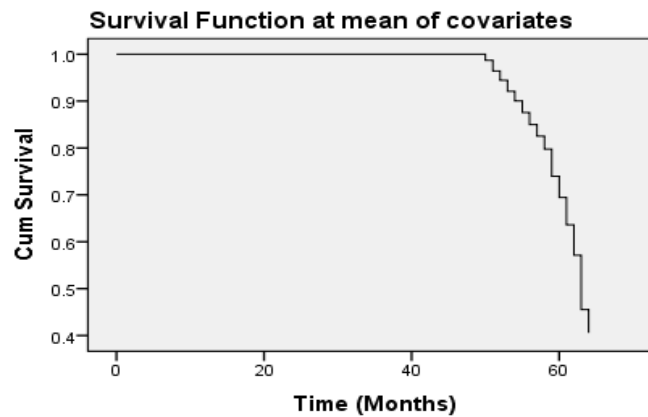


Figure 3: Cumulative Survival function at mean of covariates

Results of the Cox proportional hazards model (CPHM)

To estimate a Cox proportional hazards model and also relates covariates of sex, age, and CD4 cell count of patients to the time of death due to AIDS, the parameters are shown in Table (3a-3c) which exhibited the summary of variables used in the analysis. The Omnibus test of Model coefficients showed a statistically significant model with a chi-square value of 3.126 ($p = 0.081$) and 5 degrees of freedom (Table 3b). This implied that at least one of the variables in the model is statistically significant.

Table 3a: Case Processing Summary

		N	Percent
Cases available in analysis	Event ^a	96	32.0%
	Censored	204	68.0%
	Total	300	100.0%
Cases dropped	Cases with missing values	0	0.0%
	Cases with negative time	0	0.0%
	Censored cases before the earliest event in a stratum	0	0.0%
	Total	0	0.0%
Total		300	100.0%

a. Dependent Variable: Time (Months)

Table 3b: Omnibus Tests of Model Coefficients^a

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
963.555	4.722	5	.451	4.907	5	.427	4.907	5	.427

a. Beginning Block Number 1. Method = Enter

The regression coefficients: The results of the model are shown in Table3c which is also interpreted in terms of hazard ratios (HR). From the results, CD4 cell counts of patients have a negative regression coefficient (-0.001) which showed a reduction in the hazards (lower risks of death). All other covariates have a positive regression coefficient which showed that hazards (risk of death) due to AIDS are higher.

Table 3c: Variables in the Equation

	B	SE	Wald	Df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Sex of Patients	.139	.237	.342	1	.559	1.149	.722	1.828
CD4 Counts of Patients	-.001	.001	1.206	1	.272	.999	.998	1.001
Age of patients in years	.211	.157	1.809	1	.179	1.235	.908	1.680
State of Patients			1.410	2	.494			
State of Patients(1)	.635	.559	1.290	1	.256	1.887	.631	5.648
State of Patients(2)	.267	.385	.481	1	.488	1.306	.614	2.776

Hazards ratios: The exponential coefficient $\text{Exp}(\beta)$ are interpreted here in terms of hazard ratios which gives the effect size of covariates. From the results, sex of patients, age, and health states are positively associated with death due to AIDS (increase in hazards) with the associated negative length of survival for HIV/AIDS patients with HR (1.149, 1.235, 1.887, and 1.306) respectively. However, CD4 cell counts indicated a reduction in the hazard (HR= 0.999). This implies that patients with higher CD4 cell count have lower risks of death due to AIDS but an increased in the length of survival of HIV/AIDS patients. The cumulative survival function at the mean of covariates is shown in Figure 3.

Confidence intervals of the hazard ratios: The 95% confidence interval (C.I) of the hazard ratios showed that sex of patients has a lower and upper confidence intervals of (0.799 and 1.969); (0.998 and 1.001) for CD4 cell count, and (0.984 and 1.024) for the age of patients respectively. All the variables were statistically significant for the study.

CONCLUSION

The study concluded that CD4 cell count is the only variable or covariate that showed a lower risk of death due to AIDS. Sex, age, and state of patients had high hazards (risks of death) due to AIDS. Age, sex, and patient's health state were positively associated with death due to AIDS with an associated negative length of survival time for HIV/AIDS patients. Patients with high CD4 cell counts have lower risks of death due to AIDS but an increase in the survival time considering other factors. The

KM showed that 96 patients experienced the event of interest (death due to AIDS) accounting for 32.0% and 204 (68.0%) were censored.

RECOMMENDATIONS

The study recommends that survival analysis should be used to assess the various risk factors and the confounding effects associated with them. Thus, the patient's lifestyle should be improved to live healthy as they continue to age older.

REFERENCES

- Agada, P.O., Apeagee, B.B., & Ogwuche, I.O. (2019). A stochastic model for assessing the efficacy of antiretroviral therapy (ART) on farmers living with HIV/AIDS. *Nigerian Annals of pure and Applied Science*, 2: 247-258.
- Dessie, Z.G. (2014). Modelling of HIV/AIDS dynamic evolution using non-homogeneous semi-markov process. SpringerPlus, 3:537 Retrieved from <http://www.springerplus.com/content/3/1/537>
- Emmert-Streib, F., & Dehmer, M. (2019). Introduction to survival analysis in practice. *Machine learning and Knowledge Extraction*, 1: 1013-1038
- Harrell, F.E. (2001). Regression modelling strategies. Springer science, New York: 465-466
- Khanal, S.P., Sreenivas, V., & Acharya, S.K. (2019). Comparison of Cox proportional hazards model and Log-normal

- accelerated failure time model: Application in time to event analysis of acute liver failure patients in India. *Nepalese Journal of Statistics*, 3: 21-40
- Kleinbaum, D.G., & Klein, M. (2012). *Survival analysis: A self-learning text*, (3rd Ed), Springer New York
- Kufa, T., Chihota, V., Mngomezulu, V., Charalambous, S., Verver, S., Churchyard, G., & Borgdoff, M. (2016). The incidence of tuberculosis among HIV-positive individuals with high CD4 counts: Implications for policy. *BMC Infectious Disease*, 16: 1-7
- Lancet HIV (2017). Effect of immediate initiation of antiretroviral therapy on risk of severe bacterial infections in HIV-positive people with CD4 cell counts of more than 500 cells per μL . Secondary outcome results from a randomized controlled trial: e105-e112. [http://dx.doi.org/101016/sz352-3018\(16\)30216-8](http://dx.doi.org/101016/sz352-3018(16)30216-8)
- Moolgavkar, S.H., Chang, E.T., Watson, H.N., & Lau, E.C. (2018). An assessment of the Cox proportional hazards regression model for Epidemiologic studies. *Risk Analysis*, 38(4), 777-794
- Stel, V.S., Dekker, F.O., Tripepi, G., Zoccali, C., & Jager, K.J. (2011). Survival analysis II: Cox regression. *Nephron Clinical Practice*, 119: c255-c260
- Stevenson, M. (2007). An introduction to survival analysis. EpiCentre, IVABS, Massey University: 1-31
- UNAIDS (2014). Global AIDS response progress reporting.



©2020 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via <https://creativecommons.org/licenses/by/4.0/> which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.