APPLICATION OF COX PROPORTIONAL HAZARD MODEL TO DETERMINE THE RISK FACTORS ATTRIBUTED TO HIV IN ABUJA

Anayochukwu Christie Amaka[1]* and Hamid Salihu Amoto[1]

[1]Department of Mathematics & Statistics, Kaduna Polytechnic
*Corresponding author E-mail: amaka4christie@gmail.com

ABSTRACT

HIV/AID is a global phenomenon which has remained endemic over time. Studies have identified various risk factors of this disease. However, issues of the disease trend and the underlying driving factors especially in Sub Sahara Africa and other developing nations remain an open problem. This research work aims to determine the risk factors associated to the survival time of HIV patients using the Cox-Proportional Hazard model. Descriptive statistics were used for data exploration and the Cox-proportional hazard model was fitted to the data to determine the relative risk or odd ratio. Descriptive statistics showed that there were 125 female accounting for 69.8% of cases, 54 male accounting for 30.2% of entire cases of HIV under study. It is observed that one hundred and sixty-eight (168) patients have malaria parasite accounting for 93.9% of the HIV cases under study, 11 patients showed malaria free accounting for 6.1% of the entire population. It is also observed that the survival time of one hundred and fifty-eight (158) patients are censored accounting for 88.3% of the entire cases of HIV and twenty-one (21) got event accounting for 11.7% of HIV cases. The overall statistical significance of the model by log likelihood chi-square statistic (- 2 * log likelihood =1535.833) of the model, comparing the deviance with all of the covariates specified, against the model with covariates dropped showed statistical significance. The Risk Factors showed that patient age was the major determinant or influencer of the survival time of HIV patients. Patient gender was also discovered as contributor to the time to event outcome of studying the survival of HIV patients. However, patient malaria status was observed to have a less contribution or impact on survival of HIV patient. All the model parameters displaced were statistically significant comparing their p-value with 0.05. Also, the survival function with covariates showed that the probability of surviving at time 0 is certain and afterward the probability of survival gradually depreciates. At time year twenty (20) and above, the probability of survival with the disease approaches zero. It was recommended that Integrated HIV-malaria treatment protocols should be adopted to manage co-infections effectively. Further research should explore longitudinal models and include other potential covariates like CD4 count, ART adherence, and socioeconomic factors.

Keywords: Epidemiology, mortality risk, time to event, HIV/AIDs, Cox Proportional Hazard model, Public health statistics.

1. INTRODUCTION

Survival analysis, a statistical approach used to model time-to-event data, is an indispensable tool in biomedical research, particularly in understanding the temporal dynamics of diseases such as HIV/AIDS. It enables researchers and public health experts to predict how long patients survive under various health conditions and treatment regimes, and how different covariates affect survival time (Collett, 2013). One of the most significant applications of survival analysis lies in chronic and

infectious disease modeling, where identifying factors influencing survival is crucial for improving treatment protocols and health policies. Human Immunodeficiency Virus (HIV), a leading cause of immune system failure, has claimed millions of lives since it was first discovered in the early 80s. Despite advancements in antiretroviral therapy (ART), HIV remains a pressing global health challenge, with over 39 million people living with the virus worldwide and approximately 630,000 deaths annually as of 2022 (World Health Organization [WHO], 2023). The burden is disproportionately higher in sub-Saharan Africa, where over two-thirds of global cases are concentrated. Nigeria, in particular, is home to one of the largest populations of HIV-positive individuals, with an estimated 1.8 million people living with the virus (UNAIDS, 2022).

The situation is further complicated by the prevalence of co-morbid infections, especially malaria, which is endemic in the same geographic regions heavily affected by HIV. Malaria and HIV co-infection pose a significant public health threat due to their synergistic effects on disease progression and immune suppression. Malaria has been shown to elevate HIV viral load by increasing systemic immune activation, while HIV suppresses immunity, making individuals more susceptible to frequent and severe malaria infections (Collett, 2013). These dual infections can impair treatment outcomes, increase the risk of mother-to-child transmission, and accelerate the progression to AIDS-related complications. Given these concerns, a better understanding of how malaria co-infection affects the survival of HIV patients is crucial. While numerous studies have evaluated the survival of HIV patients using clinical variables such as CD4 count, ART adherence, and tuberculosis co-infection, relatively few have examined the effect of malaria on HIV survival using robust statistical models such as the Cox Proportional Hazard (PH) model, particularly in Nigerian setting. The Cox PH model is widely acknowledged in survival analysis for its ability to handle censored data and incorporate multiple covariates without assuming a baseline hazard function (Cox, 1972). In this context, this study aims to investigate the survival time of HIV-infected individuals and to identify how malaria co-infection, along with demographic factors such as age and gender, influences mortality risk. The study applies the Cox PH model to evaluate the hazard ratios associated with each predictor, using real-world hospital data from the Federal Capital Territory, Abuja. The findings of this study are expected to fill a critical gap in the literature on HIV survival in Nigeria, where malaria remains a dominant co-infection but is rarely modeled statistically in conjunction with HIV outcomes. By leveraging survival analysis techniques, particularly the Cox PH model, this research contributes to evidence-based health planning, offering insights into integrated care for patients with HIV and malaria. The study also provides a methodological blueprint for future research seeking to model survival in the context of multiple infectious diseases.

2. LITERATURE REVIEW

2.1 Conceptual Framework

Survival analysis is defined as a branch of statistics that deals with the analysis of time-to-event data. Modelling approaches in survival analysis are non-parametric methods such as Kaplan-Meier, semi-parametric methods such as the Cox Proportional Hazard model (Collett, 2013). The Cox model, introduced by Cox (1972), does not assume a specific baseline hazard function, making it versatile for various biomedical applications.

2.2 Joint Modelling and Co-Morbidities

Survival analysis often coexists with longitudinal data analysis in clinical research. For instance, repeated measurements of viral load or CD4 counts are longitudinal in nature and can influence survival. This intersection has given rise to joint models that simultaneously assess longitudinal and survival data, often with shared random effects (Gould *et al.*, 2014; Rizopoulos, 2011). Such models have proven particularly effective in HIV studies (Billingham & Abrams, 2002).

2.3 The Cox Proportional Hazard Model

The Cox PH model estimates the hazard ratio (HR) for multiple covariates and is particularly useful in medical studies where the outcome is influenced by various patient characteristics. For instance, Guure & Ibrahim (2012) found that age, gender, and tuberculosis co-infection significantly influenced the survival of HIV patients in Ethiopia. Similarly, Ravangard, (2011) observed higher mortality rates among males and older HIV patients in Cameroon. A Cox PH model assumes proportional hazards across time, which implies that the ratio of hazards for any two individuals is constant over time. Violations of this assumption necessitate advanced modeling techniques such as stratified Cox models or time-dependent covariates (Razaei, 2014).

2.4 Malaria and HIV: A Deadly Combination

The relationship between HIV and malaria has been examined in multiple epidemiological studies. Elfaki *et al*. (2012) observed that malaria accelerates the progression of HIV by increasing viral load and lowering immunity. According to the World Health Organization (2022), co-morbid malaria infections contribute significantly to HIV-related deaths in Africa. These findings highlight the importance of modeling both conditions together for accurate survival estimations.

2.5 Gaps in Literature

Despite the availability of survival analysis techniques, relatively few studies have comprehensively modeled the co-morbidity of HIV and malaria using the Cox PH model in Nigeria. Existing literature has focused heavily on tuberculosis and hepatitis as co-morbidities (Noora & Eifaki, 2020; Tai, 2011), leaving a notable research gap in malaria-HIV interactions.

3. MATERIALS AND METHODS

3.1   Research Design:

The Data used for this research is a health data, collected from the Federal Medical Center. The data covers a period of two (2) year's record including variable such as HIV/AIDs status and covariate such as malaria status, age, years positive, and sex. The data was based on secondary method as it entails extracting patient information from the hospital records. Descriptive Statistics of dataset is performed using mean, median, mode, variance, skewness, kurtosis and percentages to provide the description of the data and its nature. Semi-Parametric survival models such as Cox-proportional hazard model, is used in the analysis

3.4.1   Cox PH Model

The Cox PH model is usually written in terms of the hazard model. This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables denoted by X. This X represents a collection (sometimes called a "vector) of predictor variables that is being modelled to predict an individual hazard. The Cox model formula says that the hazard at time t is the product of two quantities. The first of these, $h_0(t)$, is called the baseline hazard function. The second quantity is the exponential expression ℓ $\epsilon$to the linear sum of $\beta_i X_i$, where the sum is over the p explanatory X variables (David and Mitchel, 2012).

$$h(t, X) = h_0(t)\ell^{\sum_{i=1}^{p} \beta_i X_i},$$

(1)

where $X_1, P_2, ..., X_p$ explanatory/predictor variables. One important feature of the Cox PH formula is that it concerns the proportional hazards (PH) assumption, is a function of time (t), but does not involve the $X_i's$. In contrast, the exponential expression shown here, involves the $X_i's$, but does not involve t. The $X_i's$ here are called time-independent $X_i's$. However, it is possible to consider $X_i's$ which do involve t. Such $X_i's$ are called time-dependent variables. If the variables are time-dependent, then the Cox model form may still be used, but such a model do not longer satisfy the PH assumption, and is regarded as the extended Cox model. The Cox PH model does not assume any particular "survival model" but it is not truly nonparametric because it does assume that the effects of the predictor variables upon survival are constant over time and are additive in one scale. Although the Cox PH model is a robust model, the results from using the Cox model will closely approximate the results for the correct parametric model.

Properties and Assumption of the Cox PH Model

   i.   It utilizes the exponential to ensure a non-negative hazard estimate.
  ii.   When the baseline hazard is unspecified the exponential gives the estimate of the $\beta's$ to get the effect of the predictor variables.
 iii.   Hazard function $h(t, X)$ and Survival function $S(t, X)$ can be estimated even if the hazard baseline is unspecified.
  iv.   Cox PH model is preferred to logistic when survival time information is certain.

The PH assumption requires that the HR is constant over time, or equivalently, that the hazard for one individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time. That is $\hat{h}(t, X^*) = cons\tan t \times \hat{h}(t, X)$ (David and Mitchel, 2012). However, the general rule that if the hazards cross, then the PH assumption cannot be met, so that a Cox PH model is inappropriate. Remedies to this assumption is as follows;

   1.   analyse by stratifying the exposure variable; that is, do not fit any model, and, instead oblain Kaplan-Meier curves for each exposure group separately.
   2.   start the analysis at the time the hazard cross, and use a Cox PH model base on the time survivors.

3. fit Cox model for less than hazard cross and a different Cox model or greater than hazard cross to get two different hazard ratio estimates, one for each of these two time periods.
4. fit a modified Cox model that includes a time-dependent variable which measures the interaction of exposure with time. This model is called an extended Cox model.

However, specific option may lead to specific conclusion for your analysis.

3.4.2 Hazard Ratio (HR)

Hazard ratio (HR) is defined as the hazard for one individual divided by the hazard for a different individual. The two individuals being compared can be distinguished by their values for the set of predictors, that is, the $X_i's$. We can write the hazard ratio as the estimate of $h(t, X_i^*)$ divided by the estimate of $h(t, X_i)$, where X* denotes the set of predictors for one individual, and X_i denotes the set of predictors for the other individual.

To interpret $\hat{HR}$, with $\hat{HR} > 1$, i.e. If the ratio is above 1, the risk of the event happening in $\hat{h}(t, X^*) > h(t, X)$

The general rule of X_i is given by
$$\hat{HR} = \exp\left[\beta_i(X_i^* - X_i)\right]$$
(2)

Equation (2) indirectly $\hat{HR} = \exp(\beta_i) = \ell^{\beta_i}$
(3)

when $X_i^* = 1$ and $X_i = 0$

3.5 Cox Likelihood

The Cox-PH likelihood is a product of three terminology.

L = L$_1$ × L$_2$ × L$_3$

$$L = \left[\frac{h_0(t)\ell^{\beta_1}}{h_0(t)\ell^{\beta_1} + h_0(t)\ell^0 + h_0(t)\ell^0 + h_0(t)\ell^{\beta_1}}\right] \times \left[\frac{h_0(t)\ell^0}{h_0(t)\ell^0 + h_0(t)\ell^0 + h_0(t)\ell^{\beta_1}}\right] \times \left[\frac{h_0(t)\ell^{\beta_1}}{h_0(t)\ell^{\beta_1}}\right]$$

Where:
$$L_1 = \left[\frac{h_0(t)\ell^{\beta_1}}{h_0(t)\ell^{\beta_1} + h_0(t)\ell^0 + h_0(t)\ell^0 + h_0(t)\ell^{\beta_1}}\right]$$

$$L_2 = \left[\frac{h_0(t)\ell^0}{h_0(t)\ell^0 + h_0(t)\ell^0 + h_0(t)\ell^{\beta_1}}\right]$$

$$L_3 = \left[\frac{h_0(t)\ell^{\beta_1}}{h_0(t)\ell^{\beta_1}} \right]$$

Considering the Baseline hazard, $h_0(t)$ for L above

$$L = \left[\frac{\ell^{\beta_1}}{\ell^{\beta_1} + \ell^0 + \ell^0 + \ell^{\beta_1}} \right] \times \left[\frac{\ell^0}{\ell^0 + \ell^0 + \ell^{\beta_1}} \right] \times \left[\frac{\ell^{\beta_1}}{\ell^{\beta_1}} \right]$$

Thus, L does not depend on $h_0(t)$

By the General Approach to MLE

- take the likely product of k failure times

- obtain natural log of likelihood a product of K terms

- partially differentiate with parameter interest

- Solve system of equations and obtaining the Maximum likelihood estimates.

$$L = L_1 \times L_2 \times L_3 \times ... \times L_k = \prod_{f=1}^{k} L_f$$

$$\frac{\partial InL}{\partial \beta_i} = 0$$

4. RESULTS AND DISCUSSION OF FINDINGS

Table 1: Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
EDUCATION LEVEL	179	0	5	3.22	.231
MARITAL STATUS	179	1	4	2.00	.121
AGE	179	3	64	40.23	9.621
SEX	179	0	1	.30	.460
YEARS POSITIVE	179	1	37	7.92	5.265
MALARIA	179	0	1	.94	.241
HIV-STATUS	179	0	1	.91	.201
Valid N (listwise)	179				

Table 2: Descriptive of Categories

		Count	Column N %
SEX	Female	125	69.8%
	Male	54	30.2%
	Total	179	100.0%
MARITAL STATUS	Single	50	27.9%
	Married	86	48.0%
	Divorced	19	10.6%
	Widow	24	13.4%
	Total	179	100.0%
EDUCATION LEVEL	No Formal Education	21	11.3%
	Primary	41	22.9%
	Secondary	58	32.4%
	Tertiary	46	25.7%
	Post Graduate	34	19.0%
	Total	179	100.0%
MALARIA	No	11	6.1%
	Yes	168	93.9%
	Total	179	100.0%
STATUS	Censored	158	88.3%
	Event	21	11.7%
	Total	179	100.0%

Interpretations: from Table 1, it is clearly observed that they were one hundred and seventy-nine (179) cases of HIV recorded with the period under study. The covariate for the study includes patients age, sex, years positive, malaria status. It is observed that the minimum age is 3 years and the maximum is 64 years with an average age of 40.23 years implying that the cases is common among adults. Table 2 shows that there is 125 female accounting for 69.8% of cases understudy, 54 male accounting for 30.2% of entire cases of HIV under study. Fifty (50) patients are single accounting for 27.9%, 86 married accounting for 13.4% and 19 patients are divorced accounting for 10.6% of the sample. Twenty-one (21) patients reported to have no formal education, 41 patients reported to attain primary education accounting for 22.9% and 558 patients reported to have secondary education accounting for 32.4% of the sample. One hundred and sixty-eight (168) patient shows to have malaria parasite accounting for 93.9% of the HIV cases under study 11 patient were malaria free accounting for 6.1% of the entire population. One hundred and fifty-eight (158) patients are censored accounting for 88.3% of the entire cases of HIV and 21 even accounting for 11.7% of HIV cases.

4.2   Application of Cox PH Model

Table 3: Case Processing Summary

		N	Percent
Cases available in analysis	Event[a]	21	11.7%
	Censored	158	88.3%
	Total	179	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		179	100.0%

a. Dependent Variable: YEARS POSITIVE

Table 4: Categorical Variable Codings[a,d]

		Frequency	(1)[c]
SEX[b]	0=Female	125	1
	1=Male	54	0
MARITAL STATUS	1=Single	50	1
	2=Married	86	2
	3=Divorced	19	3
	4=Widow	24	4
EDUCATION LEVEL	1=No Formal Education	21	1
	3=Primary	41	2
	4=Secondary	58	3
	5=Tertiary	46	4
	6=Post Graduate	34	5
MALARIA[b]	0=No	11	1
	1=Yes	168	0

a. Category variable: SEX
b. Indicator Parameter Coding
c. The (0,1) variable has been recoded, so its coefficients will not be the same as for indicator (0,1) coding.
d. Category variable: MALARIA

Table 5: Omnibus Tests of Model Coefficients

-2 Log Likelihood
1542.421

Interpretation: A test of the overall statistical significance of the model is given, here the Log likelihood chi-square statistic is calculated by comparing the deviance (- 2 * log likelihood =1542.421) of the model, with all of the covariates that have been specified, against the model with all covariates dropped

Test of significance of the model

H_0: the Cox-ph model is statistically insignificant
H_1: the Cox-ph model is statistically significant

Table 6: 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.
1537.142	6.991	4	.0061	6.991	4	.0058	6.891	4	.0054

a. Beginning Block Number 1. Method = Enter

Conclusion:
Since p-value = 0.0061is less than 0.05 we therefore reject H_0 and conclude that the model is statistically significant at 5% level of significance.

Table 7: Correlation Matrix of Regression Coefficients

	AGE	SEX
SEX	.247	
MALARIA	.332	.176

Interpretation: Table 7 shows that there is no correlation between the covariate indicating the absence of collinearity among the independent variable. Hence justify the non-violation of Cox Proportional Hazard model assumptions.

Test of significance for β_i
H_0: $\beta_1 = \beta_2 = \beta_3 = 0$
H_1: $\beta_1 \neq \beta_2 \neq \beta_3 \neq 0$

Table 8: Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
AGE	-.020	.009	5.175	1	.023	.981	.964	.997
SEX	-.298	.170	3.070	1	.048	.743	.532	1.036
EDUCATION LEVEL	-.185	.101	4.215	4	.041	.687	.412	.921
MARITAL STATUS	-401	.092	8.124	3	.021	.941	.918	.998
MALARIA	-.304	.333	.835	1	.031	.738	.384	1.417

Conclusion

For β_1 (coefficient of AGE) since p-value = 0.023 which is less than 0.05, we therefore reject H_0 and conclude that β_1 is statistically significant at 5% level of significance.

For β_2 (coefficient of SEX) since p-value = 0.048 which is less than 0.05 we therefore reject H_0 and conclude that β_2 which is significant statistically at 5% level of significant.

For β_3 (coefficient of EDUCATION LEVEL) since p-value = 0.041 which is less than 0.05 we therefore reject H_0 and conclude that β_3 which is significant statistically at 5% level of significant.

For β_4 (coefficient of MARITAL STATUS) since p-value = 0.021 which is less than 0.05 we therefore reject H_0 and conclude that β_4 which is significant statistically at 5% level of significant.

For β_5 (coefficient of MALARIA) since p-value = 0.031 which is less than 0.05 we therefore reject H_0 and conclude that β_5 which is significant statistically at 5% level of significance.

These conclusions is an evidence of the statistic as provided by the Wald -test. The Wald test determines the significance of each model parameters, in this case the Cox-PH model has three slope parameters which all shoes to be statistically significant.

The Expect (B)/Risk Function shows that patient age is the major determinant or influencer of the survival time of HIV patients. Patient gender was also discovered as good contributor to the survival of HIV patients. However, patient malaria status was observed to have a less contribution or impact on survival of HIV patient.
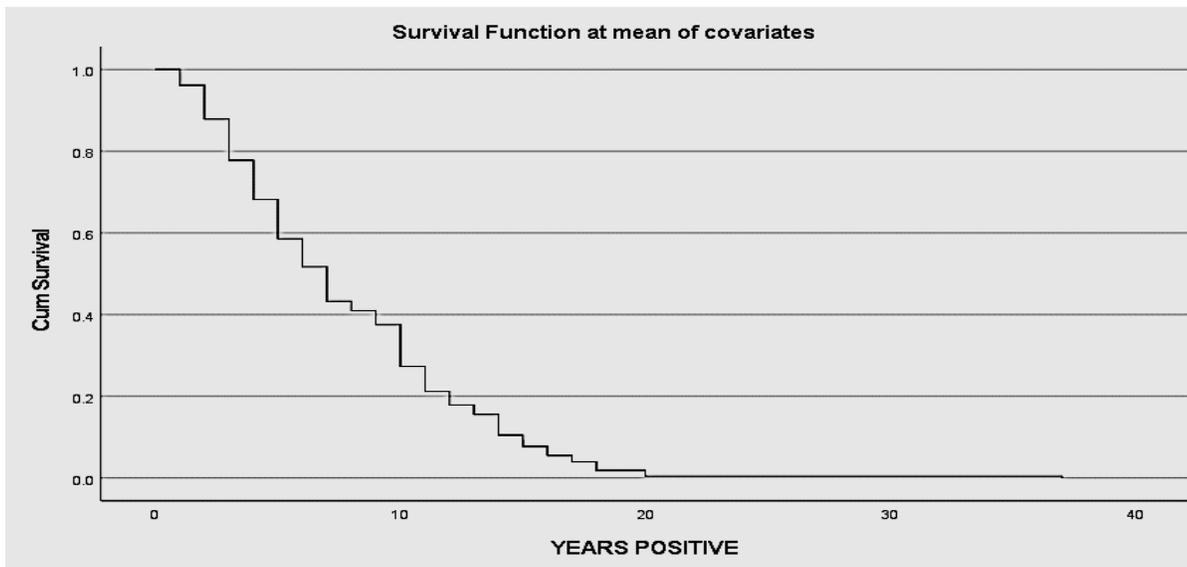
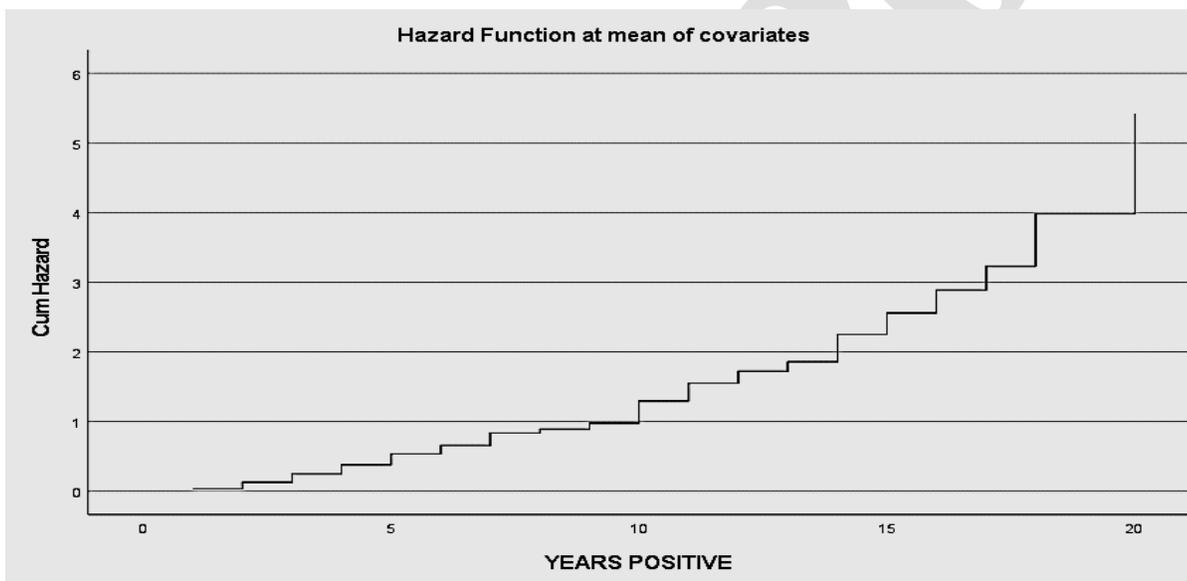**Figure 1: shows the survival function at the mean and covariates**

Figure 2: shows the hazard function at the mean and covariates

Interpretation: the survival function with covariates shows that the probability of surviving at time 0 is certain and afterward the probability of survival gradually depreciated. At time 20 years and above the probability of survival with the disease is nearly zero. This implies that the certainty of survival at time 20 years above is not assured. The hazard function with covariates shows that the probability of death (event) due to HIV/AIDs at time when the disease is not contaminated is impossible. However, the probability of getting an event as the time progress (after getting the disease) is gradually increasing which will amount to a probability of 1 certainty of the death eventually. The sum of probability of event gives the cumulative probability represented in the figure above.

5. CONCLUSION AND RECOMMENDATIONS

Based on the analysis carried out, it is concluded that Cox PH Model is statistically fit in the prediction of HIV dataset. Furthermore, based on the objectives of the study, and its significance, we make the following conclusions; under the assumptions of model, the model fitted is reliable and can be used to predict future survival times of HIV patients. The patient age shows to be more significant, followed by malaria status and then patient sex. Malaria-status is found to be a major player in survival times of HIV patients

Based on the analysis carried out and the various test carried out on the parameters and some assumptions of the model and the conclusions and the findings, it is therefore recommended that;

  i.   Early intervention programs should prioritize older patients to mitigate their higher risk of mortality.

  ii. Gender-sensitive policies in HIV treatment should address the unique challenges faced by females.

  iii. Integrated HIV-malaria treatment protocols should be adopted to manage co-infections effectively.

  iv. Further research should explore longitudinal models and include other potential covariates like CD4 count, ART adherence, and socioeconomic factors.

REFERENCES

Billingham, L. J., & Abrams, K. R. (2002). Simultaneous analysis of quality of life and survival data. *Statistics in Medicine*, *21*(19), 2791–2810. https://doi.org/10.1191/0962280202sm269

Collett, D. (2013). *Modelling Survival Data in Medical Research* (3rd ed.). Chapman & Hall/CRC.

Cox, D. R. (1972). Regression models and life- tables. *Journal of the Royal Statistical Society: Series B*, *34*(2), 187–220.

Elfaki, T. R., Bishai, W. R., Singer, J., & Willett, J. (2012). Mortality among malaria patients on treatment in Singapore. *International Journal of Malaria Parasite Disease*, *13*, 328–334.

Gould, C., Froese, T., Barrett, A. B., Ward, J., & Seth, A. K. (2014). An extended case study on the phenomenology of spatial form synaesthesia. *Frontiers in Human Neuroscience*, *8*, 433. https://doi.org/10.3389/fnhum.2014.00433

Guure, J. L., & Ibrahim, N.A., (2012). Mortality and associated risk factors in a cohort of HIV/AIDS and tuberculosis patients. *BMC Infectious Diseases*, *11*(127), pg. 45. https://doi.org/10.1186/1471-2334-11-127

Noora, A. L., & Eifaki, H. K. (2020). Survival times of patients with Advanced Acute Myelogenous Leukemia. *CBN Journal of Applied Statistics*, *8*(2), 45–56.

Ravangard, I., (2011). Determinants of survival in AIDS patients on antiretroviral therapy. *Tropical Medicine and International Health*, *14*(1), 36–43.

Razaei, K. E. (2014). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *Journal of the American Statistical Association*, *83*, 204–212.

Rizopoulos, D. (2011). Dynamic predictions in joint models for longitudinal and time-to-event data. *Biometrics*, *67*(3), 819–829.

Tai, K. K., (2011). Survival of patients with HIV/AIDS and associated factors. *Clinical Infectious Diseases*, *34*(6), 752–759.

World Health Organization (2023). *HIV statistics and data*. Retrieved from https://www.who.int/data/gho/data/themes/hiv-aids