Scientific Research Publishing

Open Journal of Statistics, 2023, 13, 595-619
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Exploring the Impact of Factors Affecting the Lifespan of HIVs/AIDS Patient's Survival: An Investigation Using Advanced Statistical Techniques

Christiana I. Ezeilo[1], Edith U. Umeh[1], Daniel C. Osuagwu[2], Chrisogonus K. Onyekwere[1*]

[1]Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria
[2]College of Business, Missouri State University, Springfield, USA
Email: *chrisogonusjohnson@gmail.com

How to cite this paper: Ezeilo, C.I., Umeh, E.U., Osuagwu, D.C. and Onyekwere, C.K. (2023) Exploring the Impact of Factors Affecting the Lifespan of HIVs/AIDS Patient's Survival: An Investigation Using Advanced Statistical Techniques. *Open Journal of Statistics*, **13**, 595-619.
https://doi.org/10.4236/ojs.2023.134029

Received: July 28, 2023
Accepted: August 20, 2023
Published: August 23, 2023

Abstract

This study investigates the impact of various factors on the lifespan and diagnostic time of HIV/AIDS patients using advanced statistical techniques. The Power Chris-Jerry (PCJ) distribution is applied to model CD4 counts of patients, and the goodness-of-fit test confirms a strong fit with a p-value of 0.6196. The PCJ distribution is found to be the best fit based on information criteria (AIC and BIC) with the smallest negative log-likelihood, AIC, and BIC values. The study uses datasets from St. Luke hospital Uyo, Nigeria, containing HIV/AIDS diagnosis date, age, CD4 count, gender, and opportunistic infection dates. Multiple linear regression is employed to analyze the relationship between these variables and HIV/AIDS diagnostic time. The results indicate that age, CD4 count, and opportunistic infection significantly impact the diagnostic time, while gender shows a nonsignificant relationship. The F-test confirms the model's overall significance, indicating the factors are good predictors of HIV/AIDS diagnostic time. The R-squared value of approximately 72% suggests that administering antiretroviral therapy (ART) can improve diagnostic time by suppressing the virus and protecting the immune system. Cox proportional hazard modeling is used to examine the effects of predictor variables on patient survival time. Age and CD4 count are not significant factors in the hazard of HIV/AIDS diagnostic time, while opportunistic infection is a significant predictor with a decreasing effect on the hazard rate. Gender shows a strong but nonsignificant relationship with decreased risk of death. To address the violation of the assumption of proportional hazard, the study employs an assumption-free alternative, Aalen's model. In the Aalen model, all predictor variables except age and gender are

*Corresponding author.

statistically significant in relation to HIV/AIDS diagnostic time. The findings provide valuable insights into the factors influencing diagnostic time and survival of HIV/AIDS patients, which can inform interventions aimed at reducing transmission and improving early diagnosis and treatment. The Power Chris-Jerry distribution proves to be a suitable fit for modeling CD4 counts, while multiple linear regression and survival analysis techniques provide insights into the relationships between predictor variables and diagnostic time. These results contribute to the understanding of HIV/AIDS patient outcomes and can guide public health interventions to enhance early detection, treatment, and care.

Keywords

Chris-Jerry Distribution, Power Chris-Jerry Distribution, Cox Proportional Hazard, Aalen's Model, Factors Affecting HIV/AIDS Patients, CD4 Counts of HIV/AIDS Patients

1. Introduction

HIV/AIDS continues to be a significant global health concern, affecting millions of individuals worldwide. The lifespan of HIV/AIDS patients can vary considerably, with several factors influencing disease progression and survival rates. Understanding these factors is crucial for healthcare professionals and policymakers to develop effective strategies for managing and improving the quality of life for HIV/AIDS patients. This article aims to analyze the impact of diagnosis date, gender, age, and CD4 counts on the lifespan of HIV/AIDS patients, utilizing multiple linear regression and as well to model the CD4 counts data of patients from St. Luke's hospital Uyo, Nigeria using Power Chris-Jerry distribution, a new model proposed by the authors for modeling CD4 counts of patients with HIV/AIDS. Before now, many studies exist that have written much on the factors that affect the time to diagnosis of HIV/AIDS infection. For instance, the time of HIV/AIDS diagnosis plays a critical role in patient outcomes. Earlier diagnosis allows for early initiation of antiretroviral therapy (ART), leading to improved immune function and prolonged survival. Mark *et al.* [1] demonstrated that delayed diagnosis is associated with a higher risk of disease progression and mortality. Mugavero *et al.* [2] in their studies showed that delayed diagnosis often leads to a delayed start of ART, resulting in lower CD4 counts and higher viral loads, both of which are strong predictors of disease progression. Therefore, interventions to promote early HIV testing and diagnosis are crucial for improving patient outcomes and prolonging lifespan. Gender as well has been recognized as a significant factor influencing the lifespan of HIV/AIDS patients. Numerous studies have reported differences in survival rates between men and women, with conflicting findings. Some studies have suggested that men have a higher risk of disease progression and mortality compared to women, May *et al.*

[3], while others have found no significant gender-based differences. Sobrino-Vegas *et al.* [4]. However, they stated that factors such as access to healthcare, engagement in treatment, adherence to ART, and underlying biological differences may contribute to the observed gender disparities. Palella *et al.* [5] argued that age at the time of HIV/AIDS diagnosis has consistently shown to impact patient's survival. Young individuals tend to experience slower disease progression and longer lifespans compared to older adults. However, Pathai *et al.* [6] in his studies discovered that old age is associated with a higher incidence of comorbidities, reduced immune response, and increased vulnerability to opportunistic infections, all of which contribute to poorer outcomes. Furthermore, age-related factors such as social support, socioeconomic status, and overall health status may influence disease progression and survival rates. Effective management strategies tailored to the unique needs of older HIV/AIDS patients are crucial for optimizing outcomes. Studies have also shown that CD4 counts are a key immunological marker used to assess the progression of HIV/AIDS and predict patient outcomes. Lawn *et al.* [7] has shown that lower CD4 counts reflect more advanced immunosuppression and are associated with an increased risk of opportunistic infections and mortality. Studies by Kuller *et al.* [8] has proven that utilizing multiple linear regression have consistently identified CD4 counts as a strong predictor of survival in HIV/AIDS patients. Early initiation of ART and regular monitoring of CD4 counts are essential for maintaining immune function and extending the lifespan of patients with HIV/AIDS. Nacher *et al.* [9], also claim that Human Immune Deficiency Virus (HIV) is an infection that attacks the body's immune system, specifically the white blood cells called CD4 cells. They're also called CD4 T lymphocytes or "helper T cells." That's because they help fight infection by triggering your immune system to destroy viruses, bacteria, and other germs that may make one sick. The loss of CD4 T-lymphocytes will result in the inability to have a proper immune response HIV destroys these CD4 cells, weakening a person's immunity against opportunistic infections, such as tuberculosis and fungal infections, severe bacterial infections and some cancers. A CD4 count is a blood test that measures the number of CD4 cells in a sample of the blood. It is a type of white blood cell. If you have AIDS, your CD4 count is so low that you may develop serious infection from virus, bacteria or fungi that usually don't cause problems in healthy people. These are called "opportunistic infections," and they can become life-threatening and can lead to death. There is no cure of AIDS but there is certain medicine which are used to slow down the virus for the HIV patient to stay healthier for a long time Coffin [10]. Udofia *et al.* [11] conducted a study in South-South Nigeria to investigate and model the survival rates of patients undergoing Antiretroviral Therapy (ART), based on stages of immune suppression and opportunistic infections. The research utilized data from 221 Human Immune Deficiency Virus (HIV) patients treated at St. Luke's Hospital Uyo, Nigeria, spanning the period from 2008 to 2017. To analyze the survival data, four different parametric

models (extreme, lognormal, logistic, and log-logistic distributions) and the nonparametric Kaplan-Meier method were considered. These models were employed to assess the survival rates and experiences of patients. Additionally, the goodness of fit for each model was evaluated using Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC). The results of the analysis indicated that the extreme distribution model exhibited the lowest AIC and BIC values, suggesting it were the most suitable parametric model for representing the survival of HIV patients at the hospital. Furthermore, the Kaplan-Meier method revealed that female patients had a more favorable survival experience compared to male patients. Jin *et al.* [12] studied the factors influencing survival status of HIV/AIDS after administration of HAART in Huzhou City, Eastern China. The data on patients' sociodemographic characteristics, treatment information, and follow-up results from the Information Management System of the Chinese Center for Disease Control and Prevention were obtained and bivariate and stepwise multivariate Cox proportional hazards regression model analyses were performed. The results indicated that respondents who were elderly and who had heterosexual transmission and whose current WHO clinical stage was stage III or IV were more likely to have died; respondents whose baseline $CD4^+$ T-lymphocyte count was equal to or more than 200 cells/μL were unlikely to have died.

In the field of literature, there is a significant focus on developing new parametric distributions to accurately represent real-world phenomena. This area of research aims to enhance the flexibility of existing distributions by introducing additional scale or shape parameters. The objective is to improve the applicability of statistical distributions in various practical domains. As a result, numerous extensions and modifications have been proposed to enhance the capabilities of existing distributions.

Shanker [13] introduced a lifetime distribution named Shanker distribution which contains a two-component mixture of an exponential distribution (with scale parameter θ) and a gamma distribution (with shape parameter 2 and scale parameter θ), with mixing proportion $p=\frac{\theta^2}{\theta^2+1}$. Ghitany *et al.* [14] proposed Power Lindley distribution (PLD), a convex combination of Weibull (α, θ) and a generalized gamma (2, α, θ) distribution with mixing proportion $\frac{\theta}{\theta+1}$, when α = 1, the PLD reduces to Lindley distribution introduced by Lindley [15], a convex combination of exponential (θ) and gamma (2, θ) distributions with their mixing proportion $\frac{\theta}{\theta+1}$. Assuming power transformation $X=Y^{\frac{1}{\alpha}}$, Shanker *et al.* [16] proposed a power Shanker distribution (PSD) a two-component mixture of Weibull distribution (with shape parameter α and scale parameter θ), and a generalized gamma distribution (with shape parameters 2, α and scale parameter θ) with mixing proportion $\frac{\theta^2}{\theta^2+1}$. Shukla [17] introduced Prakaamy distribu-

tion, a two-component mixture of an exponential distribution with scale parameter θ and a gamma distribution with shape parameter 6 and scale parameter θ, with mixing proportion $\frac{\theta^5}{\theta^5+120}$. The power transformation of Prakaamy distribution resulted to a new distribution called Power Prakaamy distribution (PPKD), Clearly PPKD is also a two-component mixture of WD with parameters (θ, α) and a generalized gamma distribution (GGD) with parameters (θ, α, 6) suggested by Shukla *et al.* [18] with their mixing proportion $p=\frac{\theta^5}{\theta^5+120}$. Using power transformation on Akshaya distribution proposed by Shanker R. [19], Ramadan *et al.* suggested a new generalized power Akshaya distribution with probability distribution. Rosa *et al.* [20] proposed Alpha Power Transformed Power Lindley Distribution, a generalization of the Power Lindley Distribution. Onyekwere and Obuleze [21] proposed a new one parameter lifetime distribution with a mixture of exponential distribution with scale parameter θ and Gamma distribution with shape and scale parameter 3 and θ respectively. The mixture is of the form $f_{CJ}(x,\theta)=pg_1(x,\theta)+(1-p)g_2(x,3,\theta)$, where $p=\frac{\theta}{\theta+2}$ having the probability distribution function (pdf):

$$f_{CJ}(x,\theta)=\frac{\theta^2}{\theta+2}\left(1+\theta x^2\right)\ell^{-\theta x},\ x>0,\theta>0 \tag{1}$$

and Cumulative density function

$$F_{CJ}(x,\theta)=1-\left(1+\frac{\theta x(\theta x+2)}{\theta+2}\right)\ell^{-\theta x},\ x>0,\theta>0 \tag{2}$$

One of the drawbacks of most of parametric, semi-parametric and nonparametric distributions already proposed by different authors is their applicability in different areas. It common to notice that what is obtainable in one country may not be obtainable in another country. Different factors may contribute to this. For instance, changing weather, climatic factors, experimental units, availability of equipment, others. These factors are considered are considered when manufacturing vaccines for treatment of diseases. These makes it impracticable for one model applied for describing an event in one place to be used in another place. Hence, the need to develop another model that can best describe the event under study while considering the environmental factors affecting the treatments.

In this paper, we proposed Power Chris-Jerry distribution, a new distribution for modelling the CD4 counts of HIV/AIDS patients at St. Luke's hospital Uyo, Nigeria. The fitting of CD4 count data using Power Chris-Jerry model provides a systematic and quantitative framework for analyzing and understanding the dynamics of HIV/AIDS progression, aiding in patient care and clinical research. Also, the model can be used to forecast future CD4 counts for individual patients or groups. This information is valuable for treatment planning, determining optimal timing for interventions such as starting antiretroviral therapy (ART), and monitoring disease progression. In the analysis of the data obtained

for studying how age, gender, CD4 count, and opportunistic infections affect the diagnostic date of HIV/AIDS patients at St. Luke Hospital in Uyo, Nigeria, we used multiple regression analysis, Cox proportional hazard modeling, and Aalen's additive model.

Multiple regression analysis was used to assess the relationship between the dependent variable (diagnostic date) and multiple independent variables (age, gender, CD4 count, opportunistic infections) simultaneously. It helps identify the strength and significance of the associations between these variables and the diagnostic date. The analysis provides regression coefficients that indicate the direction and magnitude of the relationships, as well as statistical significance tests.

Cox proportional hazard model help to evaluate the impact of covariates (age, gender, CD4 count, opportunistic infections) on the time to an event (diagnostic date). It allows for the estimation of hazard ratios, which measure the relative risk associated with each covariate. This modeling technique takes into account censoring, which occurs when the event of interest (diagnostic date) has not yet occurred for some individuals in the study. While Aalen's additive model is another approach in survival analysis that focuses on estimating the cumulative hazard function. It provides a nonparametric estimation of the hazard rate over time, considering the effects of covariates (age, gender, CD4 count, opportunistic infections). Aalen's model is useful for examining the time-varying effects of these covariates on the diagnostic date and can help identify how the hazard changes with different values of the covariates.

By employing multiple regression analysis, Cox proportional hazard modeling, and Aalen's additive model, researchers can gain a comprehensive understanding of how age, gender, CD4 count, and opportunistic infections influence the diagnostic date of HIV/AIDS patients at St. Luke Hospital in Uyo, Nigeria. These statistical techniques allow for the identification of significant predictors, estimation of hazard ratios, and assessment of time-varying effects, facilitating a more nuanced analysis of the data.

2. Power Chris-Jerry Distribution

The distribution was formed by employing power transformation method. If we let $X = Y^{\frac{1}{\alpha}}$, and performing the transformations involved, we get the pdf of the power Chris-Jerry distribution as given in Equation (1). For more information on power transformation, readers may consider reading the articles by Abebe *et al.* [22], Shanker *et al.* [23], others.

$$f\left(x,\alpha,\theta\right)=\frac{\alpha\theta^2}{\theta+2}\left(1+\theta x^{2\alpha}\right)x^{\alpha-1}\mathrm{e}^{-\theta x^{\alpha}};\ x>0,\ \alpha>0,\ \theta>0 \tag{3}$$

and the corresponding cumulative density function is

$$F\left(x,\alpha,\theta\right)=1-\left(1+\frac{\theta x^{\alpha}\left(\theta x^{\alpha}+2\right)}{\theta+2}\right)\mathrm{e}^{-\theta x^{\alpha}};\ x>0,\ \alpha>0,\ \theta>0 \tag{4}$$

The plot of the pdf and cdf of the new distribution for CD4 counts of patients are provided in **Figures 1(a)-(d)** for some values of the parameter θ and α. The graph was plotted using R software version 4.0.3.

As α and θ assume different values, the shapes changes. In **Figure 1(a)** and **Figure 1(b)**, increasing the value of α makes the distribution assume leptokurtic, also, as the values of α and θ go close to zero, the curve moves towards becoming mesokurtic. That is, it tends to assume the shape of a normal distribution. The cdf as we can see in **Figure 1(c)** and **Figure 1(d)** has an increasing function for varying values of α and θ. **Figure 2** below shows more on the visualization of pdf and the original data from CD4 count cells of patients.

Equation (5) below shows the rth moment $E\left(X^r\right)$ of PCJ distribution

$$E\left(X^r\right)=\mu_r'=\frac{\theta\Gamma\left(\frac{r}{\alpha}+1\right)+\Gamma\left(\frac{r}{\alpha}+3\right)}{(\theta+2)\theta^{\frac{r}{\alpha}}} \tag{5}$$

For r = 1, 2, 3, and 4, we obtain the first four crude moments of the PCJ distribution. The rth moment helps describe the central tendency, spread, shape, and behavior of the event under study. Moments provide important statistical measures and serve as building blocks for various statistical techniques, such as estimation, hypothesis testing, and distribution characterization. **Figure 2** shows the coefficient of skewness and Kurtosis of the Power Chris-Jerry distribution.

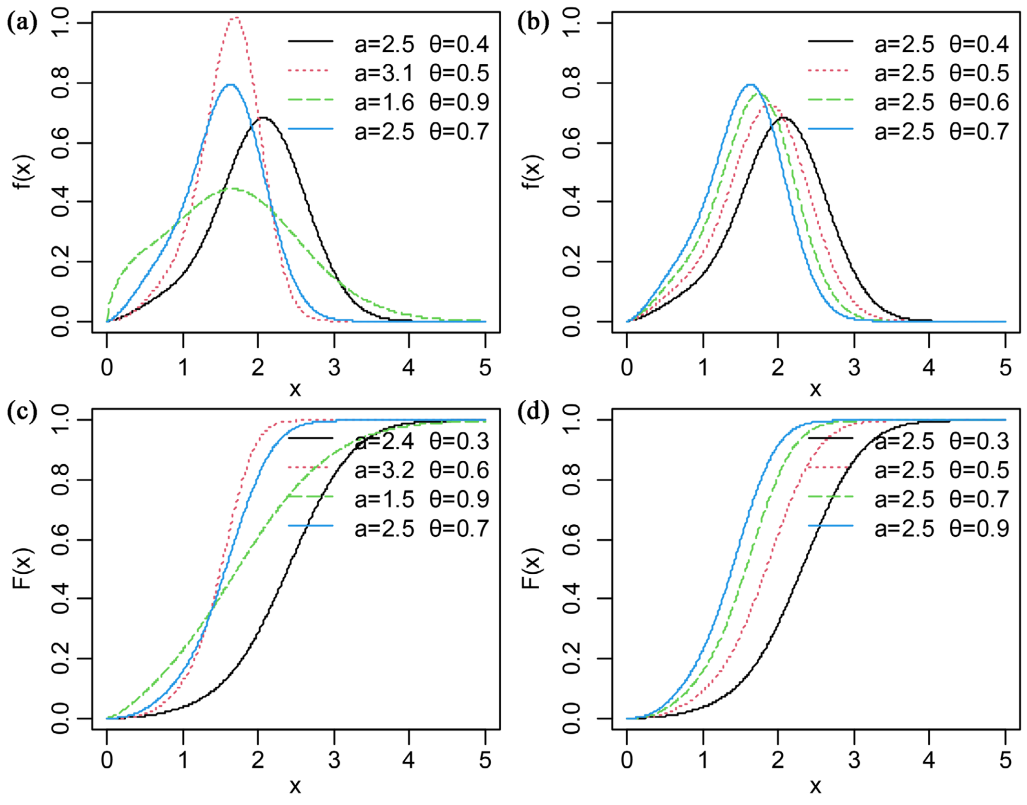

Figure 1. (a) pdf of PC-J distribution; (b) pdf of PC-J distribution; (c) cdf of PC-J distribution; (d) cdf of PC-J distribution.

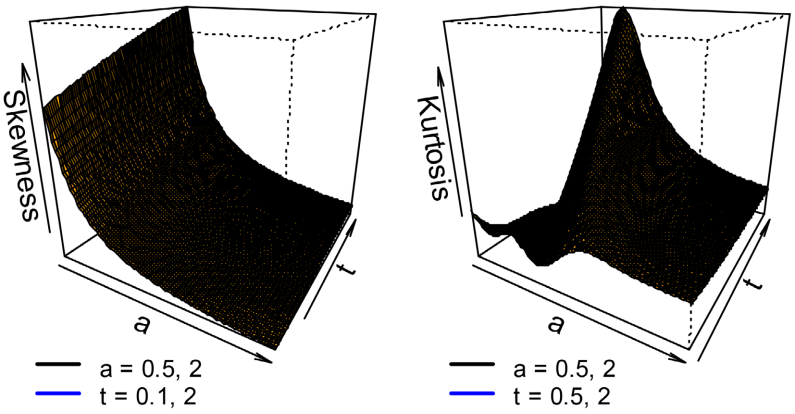

Figure 2. Skewness and kurtosis of power Chris-jerry distribution.

2.1. Maximum Likelihood Method

Let $x_1, x_2, \cdots, x_n$ constitute a random sample of size$n$ from PCJ (α, θ) distribution. Then, the likelihood function is defined as

$$\begin{aligned} L = L(\alpha,\theta \mid x) &= \prod_{i=1}^{n} \frac{\alpha\theta^2}{\theta+2}\left(1+\theta x^{2\alpha}\right) x^{\alpha-1} \mathrm{e}^{-\theta x^{\alpha}} \\ &= \left(\frac{\alpha\theta^2}{\theta+2}\right)^{n} \sum_{i=1}^{n}\left(1+\theta x^{2\alpha}\right) \sum_{i=1}^{n} x^{\alpha-1} \mathrm{e}^{-\theta\sum_{i=1}^{n} x^{\alpha}} \end{aligned} \tag{6}$$

Taking the natural logarithm of the above gives the following

$$\begin{aligned} \ln L = {} & n\ln\alpha + 2n\ln\theta - n\ln(\theta+2) + (\alpha-1)\sum_{i=1}^{n}\ln x \\ & + \sum_{i=1}^{n}\ln\left(1+\theta x^{2\alpha}\right) - \theta\sum_{i=1}^{n} x^{\alpha} \end{aligned}$$

The maximum likelihood estimators of $\hat{\alpha}_{MLE}$ and $\hat{\theta}_{MLE}$ for the parameters α and θ can be obtained numerically by maximizing, with respect to α and θ the log-likelihood function. In this case, the log-likelihood function is maximized by solving in α and θ, the non-linear equations are:

$$\frac{\partial LL}{\partial\theta} = \frac{2n}{\theta} - \frac{n}{\theta+2} - \sum_{i=1}^{n} x^{\alpha} + \sum_{i=1}^{n} \frac{x^{2\alpha}}{1+\theta x^{2\alpha}} \tag{7}$$

$$\frac{\partial LL}{\partial\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\ln x + \sum_{i=1}^{n} \frac{2\theta x^{2\alpha}\ln x}{1+\theta x^{2\alpha}} - \theta\sum_{i=1}^{n} x^{\alpha}\ln x \tag{8}$$

The solution of $\frac{\partial \ln L}{\partial\theta} = 0$ and $\frac{\partial \ln L}{\partial\alpha} = 0$ gives the maximum likelihood estimates of the parameters

2.2. Application of the Proposed Distribution to CD4 Count Data of HIV/AIDS Patients

In order to illustrate the real-life application of Power Chris-Jerry distribution, a data set on CD4 count cells of HIV/AIDS patients from St. Luke hospital Uyo, found in Udofia *et al.* [11], was fitted to the new distribution, and compared together with Power Prakaamy distribution (PP), Power Garima distribution (PG), Chris-Jerry distribution (CJ), and Lindley distribution (LD). Table 1 show the

Table 1. MLE, LL, AIC, BIC, KS and P-values of PCJ (CD4 Count).

Model	Parameters	S.E	LL	AIC	BIC	KS	P
PCJ	$\theta = 0.07630469$	0.01309	−1391.94	2787.88	2794.677	0.05002	0.6196
	$\alpha = 0.6976365$	0.0309866					
PP	$\theta = 0.5818447$	0.0694722	−1392.132	2788.265	2795.061	0.06693	0.2635
	$\alpha = 0.4539251$	0.0219133					
PG	$\theta = 0.0041663$	0.0006301	−1392.509	2789.018	2795.814	4.809	3.11E−15
	$\alpha = 1.09783782$	0.0271936					
CJ	$\theta = 0.0144119$	0.0005608	−1430.526	2863.052	2866.45	0.17723	0.000001
LD	$\theta = 0.009644$	0.0004537	−1397.698	2797.395	2800.793	0.10394	0.01564

results of the estimates, that is, MLE, LL, AIC, BIC, KS and P-values, obtained when analyzed with R software. See appendix for the data used in this study.

The test of goodness of fit conducted using Kolmogorov Smirnov test indicates that the CD4 count cell data fitted the proposed distribution very well since the value of KS for the distribution (PCJ) is 0.05002 with p value 0.6196, which is above 5% level of significance. To discriminate among the models, the information criteria, that is, AIC and BIC were used. The distribution with the largest negative log-likelihood, least AIC and BIC, is considered the best. In Table 1, it is obvious that PCJ distribution possesses the largest negative log-likelihood with value −1391.94, with AIC and BIC respectively given by 2787.88 and 2794.677. These values are smaller than that of the competing distributions, hence, power Chris-Jerry distribution is chosen in preference to them. Hence, the PCJ model provided a suitable fit with the smallest AIC and BIC values.

3. Multiple Linear Regression Model

The Multiple linear regression model aid to explain the relationship between the response variable, HIV/AIDS date of diagnosis and the regressor variables. That is, age, CD4 count of patients, opportunistic infections and gender. The relationship is given by the model

$$\text{HIV/AIDS} = \beta_0 + \beta_1 \text{AGE} + \beta_2 \text{CD4\_C} + \beta_3 \text{OI\_T} + \beta_4 \text{Gender} + \zeta \tag{9}$$

The aim is to examine the impact of age, CD4 count of patients, opportunistic infection dates and gender on the HIV/AIDS diagnosis time of patients at St. Luke hospital Uyo, Nigeria.

The relationship between HIV date of diagnosis and age, CD4 count, opportunistic infection, and gender can be analyzed using a multiple linear regression model. In this type of model, we can estimate the effect of each independent variable (age, CD4 count, opportunistic infection, and gender) on the dependent variable (HIV date of diagnosis), while controlling for the other independent variables in the model.

The coefficients of the independent variables in the multiple linear regression

model can help us understand the direction and magnitude of the relationship between each independent variable and the dependent variable. For example, if the coefficient for age is negative, it suggests that older patients tend to have a shorter time from HIV infection to diagnosis, while a positive coefficient suggests that younger patients tend to have a shorter time to diagnosis.

If the coefficient for CD4 count is negative, it suggests that patients with higher CD4 counts tend to have a longer time from HIV infection to diagnosis, while a positive coefficient suggests that patients with lower CD4 counts tend to have a shorter time to diagnosis.

Also, if the coefficient for opportunistic infection is positive, it suggests that patients who have had an opportunistic infection tend to have a shorter time from HIV infection to diagnosis, while a negative coefficient suggests that patients who have not had an opportunistic infection tend to have a longer time to diagnosis.

For gender, if the coefficient for gender is positive, it suggests that female patients tend to have a shorter time from HIV infection to diagnosis compared to male patients, while a negative coefficient suggests that male patients tend to have a shorter time to diagnosis compared to female patients. Table 2 shows the regression coefficients and the corresponding p values. After performing the regression analysis using R software, we observe that the value of β_0 is about 13, which indicates the average diagnostic time of the HIV/AIDS patients in the hospital when other factors are kept constant. The value of the coefficient of age, β_1 provided in Equation (10) is 0.02434. This suggests that younger patients tend to have a shorter time from time of contracting HIV infection to diagnosis time than the older ones.

Also, the value of β_2 is −0.02254. Since the coefficient for CD4 count is negative, it suggests that patients with higher CD4 counts tend to have a longer time from HIV infection to diagnosis. While controlling for the effects of age, CD4 count, the coefficient of opportunistic infection β_3 is 0.77155. This suggests that patients who have had an opportunistic infection tend to have a shorter time from HIV infection to diagnosis.

Gender with coefficient β_4 and value −2.29265 suggest that male patients in the hospital tend to have a shorter time to diagnosis compared to female patients.

The estimated regression model for the diagnostic time of HIV/AIDS patients

Table 2. Regression coefficients.

Coefficients	Estimates	S.E	t-value	p value
Intercept	12.72536	6.00300	2.120	0.035161
AGE	0.02434	0.09509	0.256	0.798181
CD4_C	−0.02254	0.00631	−3.572	0.000436
OI_T	0.77155	0.03423	22.539	0.0001
Gender	−2.29265	2.32511	−0.986	0.325216

in the hospital is given by

$$\text{HIV\_AIDS} = 12.73 + 0.02\text{AGE} - 0.02\text{CD4\_C} + 0.77\text{OI\_T} - 2.29\text{Gender} + \zeta \quad (10)$$

To assess the significance of the model formulated, we implement a statistical technique called hypothesis testing. To perform the hypothesis testing, we adopt the t-test. The null hypothesis generally denoted by H_0 and alternative hypothesis represented by H_1 can be formulated as $H_0 : \beta_i = 0$ Vs $H_1 : \beta_i \neq 0$. A p value less than 0.05 shows significance impact. From Table 2, the t-value for age is 0.256 with p value 0.7981. This shows that, although younger patients tend to have a shorter time of HIV infection to diagnosis time than the older ones, the time is not statistically significant. Also, apart from Gender whose p value is above 0.05, other factors such as CD4 count and opportunistic infection diagnostic time have significant impact on the HIV/AIDS diagnostic time. Figure 3

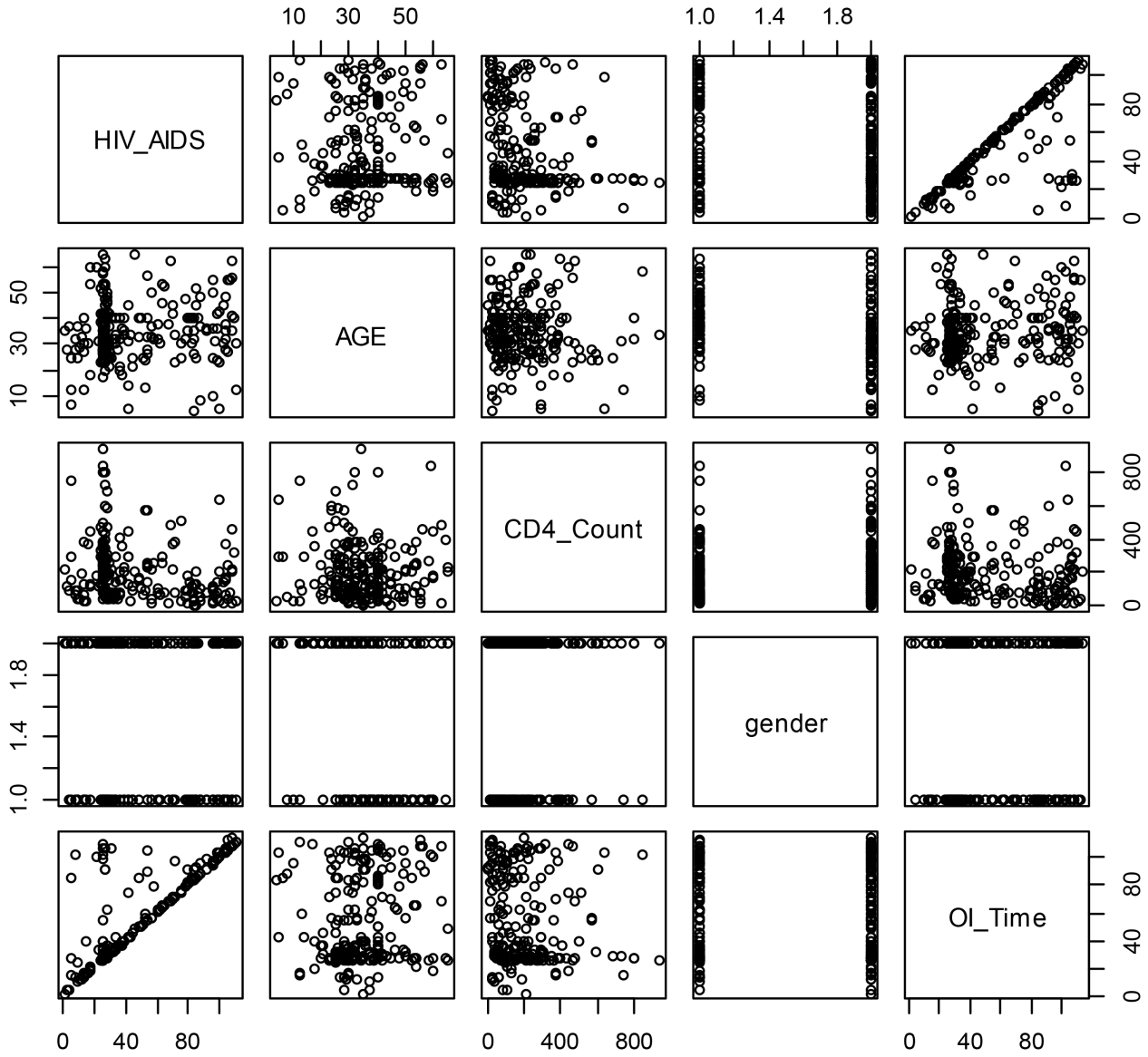

Figure 3. Relationship between the HIV/AIDS predictor variables.

shows a visual display (graphical illustration) of the relationship between HIV/AIDS predictor variables used in this research.

Figure 3 shows the visualization of the data sets. Residuals represent the deviation of an observed value of an element and its theoretical value. In regression analysis, the residual is the difference between any data point and the regression line. Sometimes they are also known as an error. The residual is the error that is not explained by the regression line.

Figure 4 shows the behavior of the residuals. From the graphical visualization of the residuals in 4, the following were observed:

1) The red line in the residual vs. fitted plot lies closer to the residual value of 0. Based on the plot, it can be seen that the residuals of the model are weakly linearly related. Linearity means that the predicted variable in the regression model has a straight-line relationship with *Y*.

2) Homoscedasticity is a fundamental assumption of linear regression models. If this assumption is violated, the problem of heteroscedasticity arises. The scale-location plot shows the fact that the residuals follow the homoscedasticity property.

3) In regression analysis, an observation whose deletion from the data has a significant effect on the estimates of the model parameters is called influential

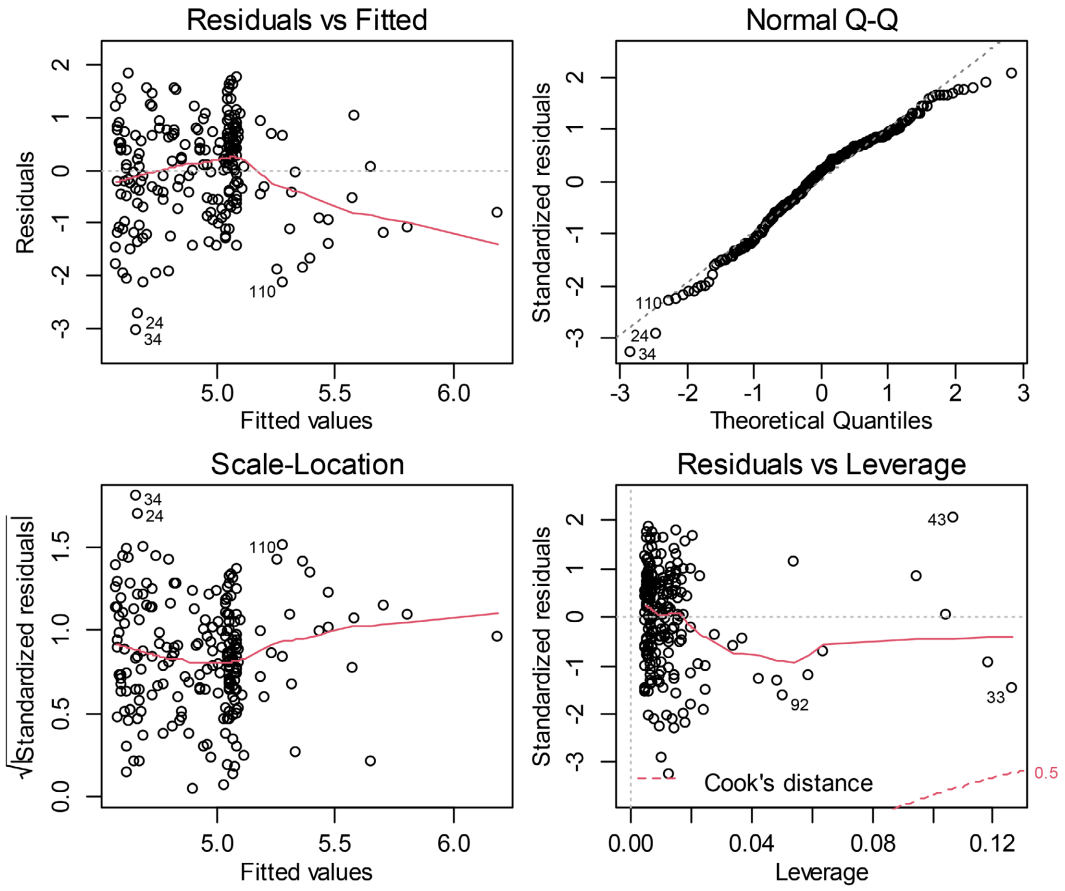

Figure 4. Visual display of residual and fitted values, normal q-q, scale location and leverages.

observation. Residual vs. leverage plot shows that there are some influential observations.

4) Quantile-quantile (Q-Q) function is a visual approach to check the normality. The Q-Q function of the residuals follows normal distribution.

F-Test

Here, we implement another powerful statistical test (called F-test) to check the overall significant of the model. If the value of the F-statistic is far from zero, then it is indicating a positive impact of age, CD4 count, opportunistic infection time and gender on HIV/AIDS diagnostic time. As given in Table 3, the value of the F-statistic is 140.5 with p value 2.20×10^{-16}. This indicates that the factors considered are better predictor of the HIV/AIDS diagnostic time model. To examine the percentage contribution of the HIV/AIDS patients predictor variable or goodness of fit of the model, R^2 is used. For this study, the R^2 is about 0.72 (72%). This indicates that the HIV/AIDS diagnostic time to infection of the patients can be improved to about 72%. This can be done by administering antiretroviral therapy (ART) to HIV patients which tends to suppress the replication of the virus and prevent further damage to immune system.

4. Application of HIV/AIDS Patient Data to Cox Proportional Hazard Model

Cox proportional hazards regression aid to describe the effects of the predictor variables on survival time of patients. Table 4 & Table 5 below show the results

Table 3. Model summary.

R-Square	Adj. R-Sq	F-value	p value
0.7223	0.7172	140.5	2.20E−16

Table 4. Cox Proportional hazard model.

	coef	exp(coef)	se(coef)	z	p
AGE	−0.00257	0.9974377	0.007601	−0.338	0.736
CD4_Count	−0.00041	0.9995867	0.000635	−0.651	0.515
OI_Time	−0.20507	0.8145931	0.017953	−11.422	2e−16
Gender	−0.22447	0.7989408	0.181366	−1.238	0.216

Table 5. Cox Proportional hazard model

	exp(coef)	exp(coef)	lower.95	upper.95
AGE	0.9974	1.003	0.9827	1.0124
CD4_Count	0.9996	1.0000	0.9983	1.0008
OI_Time	0.8146	1.228	0.7864	0.8438
Gender	0.7989	1.252	0.5599	1.14

obtained using R software, for Cox proportional hazard model.

The output shows that the model includes four predictor variables, age, CD4 count, opportunistic infection date of diagnosis (time) and gender. The P value for all the tests (likelihood ratio test, Wald test and Score) are significant, indicating that the Cox hazard model is significant. These tests evaluate the omnibus null hypothesis that all the (β_i) are 0. From the results, the test statistics are in close agreement, and the null hypothesis is rejected. In the multivariate Cox analysis, the covariates age, CD4 count of patients and gender are not significant ($p > 0.05$). However, the covariate opportunistic infection is the only factor that appears to be significant ($p = 2 \times 10^{-16}$). The coefficient estimates indicate the change in the log hazard ratio associated with a one-unit increase in the predictor variable, holding the other predictor variables constant. Based on the p-values, age has p value 0.736 with $HR = 0.99$ which is not significant at 0.05. This indicates that the association between patient's age and decreased risk of death is not significant. For CD4_Count, the coefficient is −0.00041, with p value 0.515 and $HR = \exp(\text{coef}) = 0.99$. This suggest that CD4 count is not significantly associated with the hazard of HIV/AIDS diagnostic time after adjusting for the other covariates included in the model. Specifically, for one unit increase CD4 count, the hazard of the HIV/AIDS diagnostic time decreased by a factor of 0.99, although the relationship is not significant. Opportunistic infection has coefficient of −0.2051, $p = 2 \times 10^{-16}$ and $HR = 0.81$. This means that opportunistic infection is a significant predictor of HIV/AIDS diagnostic time. That is, it significantly associated with the hazard of HIV/AIDS diagnostic time. For a unit increase in opportunistic infection, the hazard of HIV/AIDS diagnostic time decreased by a factor 0.81.

However, gender has coefficient of −0.224, p values 0.216 and $HR = \exp(\text{coef}) = 0.79$ respectively. For the gender, it indicates a strong and nonsignificant relationship between the patient's gender and decreased risk of death. This means that the male gender has higher hazard rate than the female.

4.1. Test of Assumption for Proportional Hazard

We test the assumption using the graph of scaled Schoenfeld residuals (shown in Figure 5) against the transformed time.

From the graphical inspection, there is pattern with time. However, the assumption of proportional hazard appears to have been violated. The Global Schoenfeld value is 37.07 with p value of 1.7×10^{-7}. This is an indication of the violation of the assumption of Cox proportional hazard model. Since the assumption has been violated, in order to obtain the right estimate of the coefficient of the model, it is pertinent to resort to an assumption free model, a nonparametric alternative to Cox model. That is Aalen Model. Figure 6 shows the graph of the survival function

The survival curve for a Cox model represents the probability of survival over time for a group of individuals, based on their characteristics or covariates.

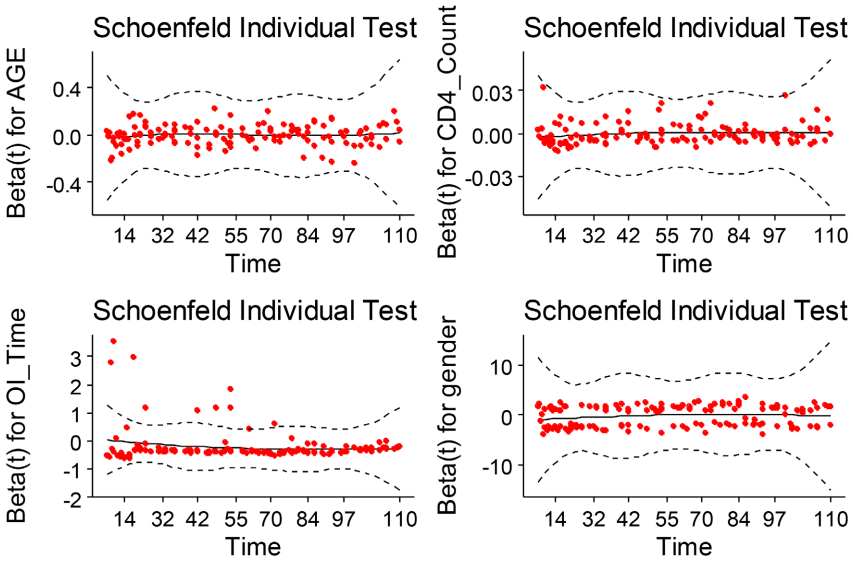

Figure 5. Graph of Schoenfeld residuals.

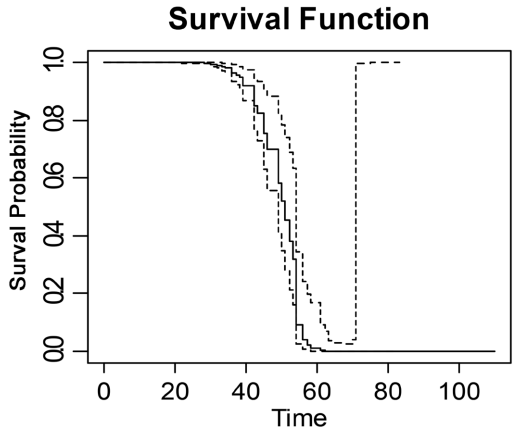

Figure 6. Plot of the survival function for Cox proportional Model.

When the survival curve is increasing, it means that the probability of survival is increasing over time. Conversely, when the survival curve is decreasing, it means that the probability of survival is decreasing for the HIV/AIDS patients. Figure 7 also shows the hazard function

Figure 6 and Figure 7 show the survival and hazard rate of the Cox proportional model. From Figure 6, it can be seen that the probability of survival decreased and increased over time. Figure 6 revealed that age, CD4 count, opportunistic infection and gender had a negative effect on the HIV/AIDS diagnostic time for some time (for area with decreasing function), at a point, they covariates began to produce positive effects (for areas with increasing function). Also, the hazard plot in Figure 7, shows that the hazard rate is highest at the beginning of the study and later stationary with time.

4.2. The Aalen Model

The results obtained by R software for Aalen model is given in Table 6 below.

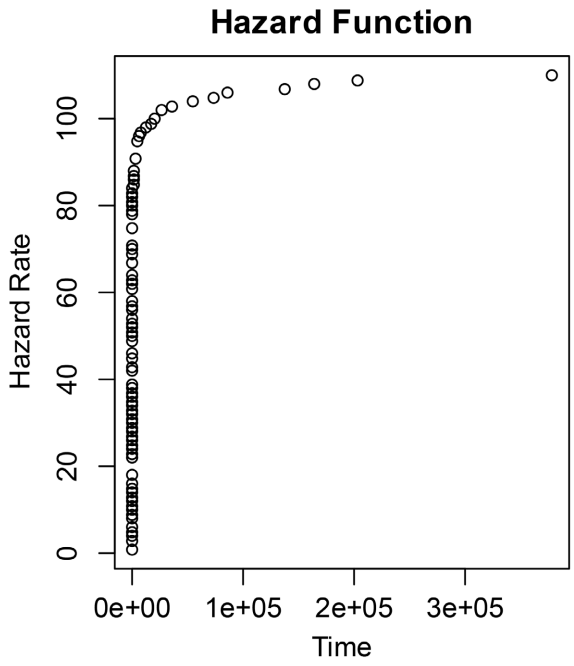

Figure 7. Plot of the hazard rate for Cox proportional model.

Table 6. Aalen Model Estimates of the coefficients.

	slope	coef	se(coef)	z	p
INTERCEPT	0.214	0.0303	0.00369	8.2	0.0001
AGE	0.000154	0.0000328	0.0000462	0.711	0.477
CD4_Count	0.0000569	0.00000847	3.84E−06	2.21	0.0274
OI_Time	−0.00249	−0.000255	0.0000234	−10.9	0.0001
Gender	0.00456	0.00083	0.00116	0.713	0.476

$$\chi^2 = 132.11 \quad \text{on 4 df}, \quad p \le 2.0 \times 10^{-16}$$

From Table 6, all the predictor variables except for age and gender are statistically significant. This means that they are likely to have a significant effect on HIV/AIDS diagnostic time. On the general note, chi-square value of 132.11 with p value of 2.0×10^{-16} indicates that the Aalen's model is significant.

The baseline hazard rate at time 0 is exp (0.214) = 0.58, which means that the hazard rate of the event of interest at time 0 is 0.58.

For every one-year increase in age, the hazard of being diagnosed with HIV/AIDS increases by 0.000154 units, holding other variables constant. However, since the p value for age is not significant, we cannot conclude that this effect is real.

For every one-unit increase in CD4 count, the hazard of being diagnosed with HIV/AIDS increases by 0.0000569 while keeping other variables constant, and this is statistically significant.

For individuals with opportunistic infections, the hazard of being diagnosed

with HIV/AIDS decreases by 0.00249 units, holding other variables constant. However, since the p value is 0.0001, we can conclude that the effect is statistically significant.

For females compared to males, the hazard of being diagnosed with HIV/AIDS increases by 0.00456 units, holding all other variables constant. However, the effect of gender on HIV/AIDS is not statistically significant since the p value is 0.476. These results provide important insights into the factors that influence the time to HIV/AIDS diagnosis, and can help inform public health interventions aimed at reducing HIV/AIDS transmission and improving early diagnosis and treatment. Figure 8 demonstrate the probability plots of the Alen's model used in modeling the HIV/AIDS diagnostic time of the patients in St. Lukes hospital Uyo. It can be seen from the plot that intercept and CD4 counts have an increasing effect on the hazard of HIV/AIDS time over time. The covariates become more influential as time progresses. Age and gender remain fairly constant and after some time, became more beneficial in the study. After some sometime in the study, growing older and a male becomes an advantage in the study.

However, from the graph, opportunistic infection has a decreasing effect on the hazard of HIV/AIDS diagnostic time.

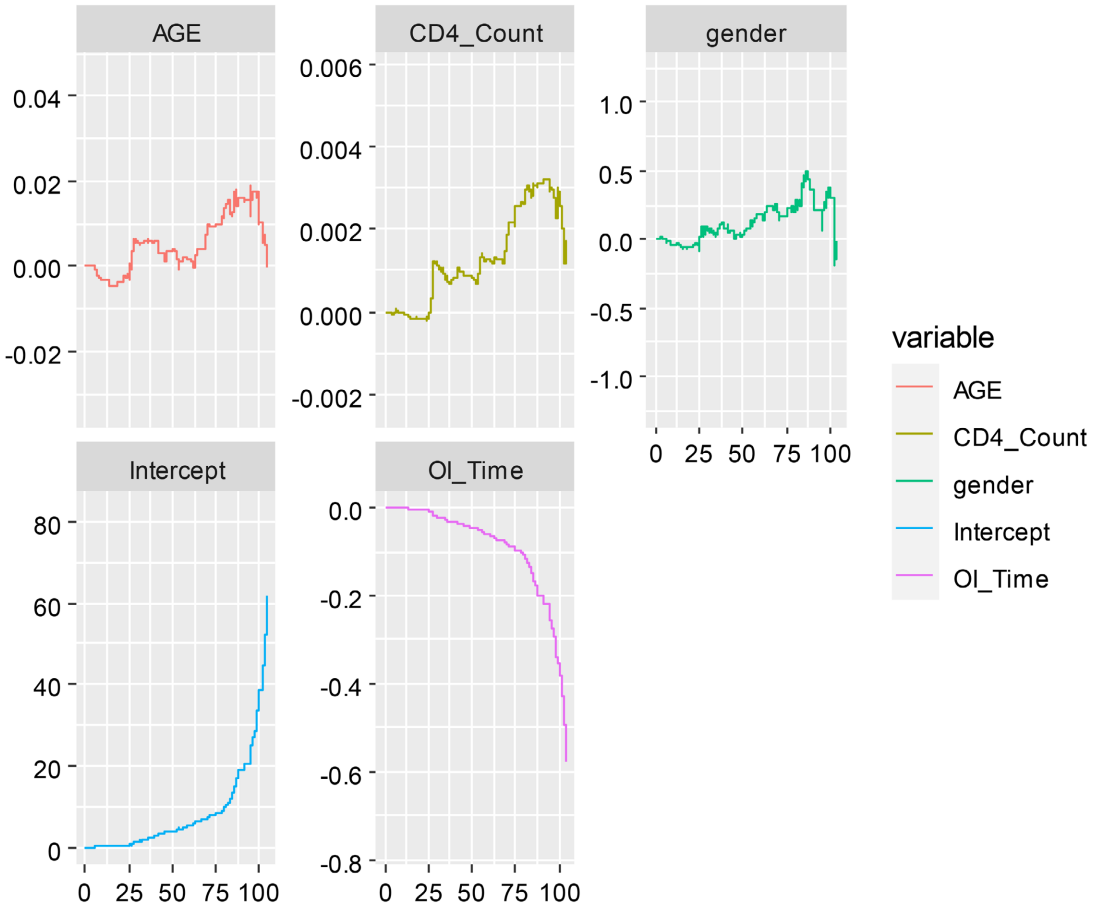

Figure 8. Probability plot of the effects of the covariates on the HIV/AIDS diagnostic time.

5. Conclusions

In conclusion, this study aimed to explore the impact of various factors on the diagnostic time and survival of HIV/AIDS patients using advanced statistical techniques. The analysis involved Power Chris-Jerry distribution, multiple linear regression, Cox proportional hazards regression, and Aalen's additive hazard model. In this research, a new distribution known as "Power Chris-Jerry distribution" was introduced. It was used to model the CD4 counts of patients. In order to measure its performance and superiority over other models, the log-likelihood (LL), Akaike information criteria (AIC), Bayesian information criteria (BIC) and K-S statistics were used. After comparing the results with other distributions as shown in Table 1, PCJ distribution appear to perform better than the other distributions.

The multiple linear regression model revealed that age, CD4 count, opportunistic infection, and gender were significant predictors of the diagnostic time of HIV/AIDS patients. Younger patients tended to have a shorter time from HIV infection to diagnosis, while patients with higher CD4 counts had a longer time to diagnosis. Patients with opportunistic infections were diagnosed more quickly, and male patients tended to have a shorter time to diagnosis compared to female patients. However, the significance of age and gender was not statistically supported.

The Cox proportional hazards regression model focused on survival time and identified opportunistic infection as the only significant predictor. CD4 count showed a non-significant association with the hazard of HIV/AIDS diagnostic time after adjusting for other covariates. Gender demonstrated a nonsignificant relationship with decreased risk of death. The assumption of proportional hazards was violated, necessitating the use of Aalen's additive hazard model as a nonparametric alternative.

The Aalen's model confirmed that age, CD4 count, and opportunistic infection significantly influenced HIV/AIDS diagnostic time, while gender did not. The hazard of being diagnosed with HIV/AIDS increased with age and CD4 count, while opportunistic infections decreased the hazard. These findings provide valuable insights into the factors affecting the time to HIV/AIDS diagnosis and can contribute to public health interventions aimed at reducing transmission and improving early detection and treatment. It is vital to note that this study has certain limitations, such as being based on data from a specific hospital and focusing on a particular region. Therefore, the generalizability of the results to other populations and settings may be limited. Further research incorporating a broader sample and considering additional variables would enhance our understanding of the factors influencing the survival of HIV/AIDS patients. In general, the study highlights the significance of advanced statistical techniques in exploring and identifying factors that affect the lifespan and diagnostic time of HIV/AIDS patients, providing valuable information for improving patient care and public health interventions.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Marks, G., Crepaz, N. and Janssen, R.S. (2006) Estimating Sexual Transmission of HIV from Persons Aware and Unaware That They Are Infected with the Virus in the USA. *AIDS*, **20**, 1447-1450. https://doi.org/10.1097/01.aids.0000233579.79714.8d

[2] Mugavero, M.J., Amico, K.R., Horn, T. and Thompson, M.A. (2013) The State of Engagement in HIV Care in the United States: From Cascade to Continuum to Control. *Clinical Infectious Diseases*, **57**, 1164-1171. https://doi.org/10.1093/cid/cit420

[3] May, M., Gompels, M. and Delpech, V. (2011) Impact of Late Diagnosis and Treatment on Life Expectancy in People with HIV-1: UK Collaborative HIV Cohort (UK CHIC) Study. *The BMJ*, **343**, d6016. https://doi.org/10.1136/bmj.d6016

[4] Sobrino-Vegas, P., Rodríguez-Urrego, J. and Berenguer, J. (2001) Delayed Diagnosis of HIV Infection in Spain: Missed Opportunities for Optimal Intervention. *Journal of Acquired Immune Deficiency Syndromes*, **26**, 513-520.

[5] Palella Jr, F.J., Deloria-Knoll, M. and Chmiel, J.S. (2003) Survival Benefit of Initiating Antiretroviral Therapy in HIV-Infected Persons in Different CD4+ Cell Strata. *Annals of Internal Medicine*, **138**, 620-626. https://doi.org/10.7326/0003-4819-138-8-200304150-00007

[6] Pathai, S., Bajillan, H., Landay, A.L. and High, K.P. (2014) Is HIV a Model of Accelerated or Accentuated Aging? *The Journals of Gerontology*: *Series A*, **69**, 833-842. https://doi.org/10.1093/gerona/glt168

[7] Lawn, S.D., Butera, S.T. and Folks, T.M. (2001) Contribution of Immune Activation to the Pathogenesis and Transmission of Human Immunodeficiency Virus Type 1 Infection. *Clinical Microbiology Reviews*, **14**, 753-777. https://doi.org/10.1128/CMR.14.4.753-777.2001

[8] Kuller, L.H., Tracy, R. and Belloso, W. (2008) Inflammatory and Coagulation Biomarkers and Mortality in Patients with HIV Infection. *PLOS Medicine*, **5**, e203. https://doi.org/10.1371/journal.pmed.0050203

[9] Nacher, M., Huber, F., Adriouch, L., Djossou, F., Adenis, A. and Couppie, P. (2018) Temporal Trend of the Proportion of Patients Presenting with Advanced HIV in French Guiana: Stuck on the Asymptote? *BMC Research Notes*, **11**, Article No. 831.

[10] Coffin, J.M. (1999) Molecular Biology of HIV. In: Crandall, K.A., Ed., *The Evolution of HIV*, Johns Hopkins University Press, Baltimore, 3-40.

[11] Udofia, E.M., Umeh, E.U. and Onyekwere, C.K. (2021) Modeling of Survival of HIV Patients by Stages of Immune Suppression and Opportunisic Infections. *American Journal of Theoretical and Applied Statistics*, **10**, 233-242. https://doi.org/10.11648/j.ajtas.20211006.12

[12] Jin, M., Yang, Z., Li, J., Liu, X. and Wu, Z. (2022) Factors Influencing Survival Status of HIV/AIDS after HAART in Huzhou City, Eastern China. *Canadian Journal of Infectious Diseases and Medical Microbiology*, **2022**, Article ID: 2787731. https://doi.org/10.1155/2022/2787731

[13] Shanker, R. (2015) Shanker Distribution and Its Applications. *International Journal*

of Statistics and Applications, **5**, 338-348.
https://doi.org/10.15406/bbij.2016.03.00075

[14] Ghitany, M.E., Al-Mutairi, D.K., Balakrishnan, N. and Al-Enezi, L.J. (2013) Power Lindley Distribution and Associated Inference. *Computational Statistics & Data Analysis*, **64**, 20-33. https://doi.org/10.1016/j.csda.2013.02.026

[15] Lindley, D.V. (1958) Fiducial Distributions and Bayes' Theorem. *Journal of the Royal Statistical Society*: *Series B* (*Methodological*), **20**, 102-107.
https://doi.org/10.1111/j.2517-6161.1958.tb00278.x

[16] Shanker, R. and Shukl, K.K. (2017) Power Shanker Distribution and Its Application. *Turkiye Klinikleri Journal of Biostatistics*, **9**, 175-187.
https://doi.org/10.5336/biostatic.2017-56601

[17] Shukla, K.K. (2018) Prakaamy Distribution with Properties and Applications. *Journal of Applied Quantitative Methods*, **13**, 30-38.

[18] Shukla, K.K. and Shanker, R. (2020) Power Prakaamy Distribution and Its Applications. *Journal of Computational & Theoretical Statistics*, **7**, 25-36.

[19] Shanker, R. (2017) Akshaya Distribution and Its Application. *American Journal of Mathematics and Statistics*, **7**, 51-59.

[20] Rosa, L.A., Nurrohmah, S. and Fithriani, I. (2019) α Power Transformed Lindley Distribution. *ICSA-International Conference on Statistics and Analytics* 2019, Bogor, 2-3 August 2019, 130-142. https://doi.org/10.29244/icsa.2019.pp130-142

[21] Onyekwere, C.K. and Obulezi, O.J. (2022) Chris-Jerry Distribution and Its Applications. *Asian Journal of Probability and Statistics*, **20**, 16-30.
https://doi.org/10.9734/ajpas/2022/v20i130480

[22] Abebe, B., Tesfay, M., Eyob, T. and Shanker, R. (2019) A Two-Parameter Power Rama Distribution with Properties and Applications. *Biometrics and Biostatistics International Journal*, **8**, 6-11. https://doi.org/10.15406/bbij.2019.08.00262

[23] Shanker, R. and Shukla, K.K. (2017) Power Shanker Distribution and Its Application. *Turkiye Klinikleri Journal of Biostatistics*, **9**, 175-187.
https://doi.org/10.5336/biostatic.2017-56601

Appendix A.

R Codes for the Analysis

```
#load the survival packages
library(survival)
library(ggplot2)
library(dplyr)
#create a survival object
surv<-with(HIV_DATA,Surv(time,status))
#Fit a Cox proportional hazards model
cox_model<-coxph(surv~AGE+CD4_Count+OI_Time+gender, data = HIV_DATA)
#View model summary
summary(cox_model)
================================================================================
### Test of assumption for cox model
test.ph<-cox.zph(cox_model)
test.ph
================================================================================
####Plot graph of diagnostic
library(survminer)
ggcoxzph(test.ph)
####Plot of the survival function
sfit<-survfit(cox_model)
plot(sfit,main="Survival Function", xlab = "Time", ylab = "Survival Probability")
###Plot of the Hazard rate
hfit<-basehaz(cox_model)
plot(hfit,main="Hazard Function", xlab = "Time", ylab = "Hazard Rate")
================================================================================
# Fit the Aalen model
model <- aareg(Surv(time) ~ AGE+CD4_Count+OI_Time+gender, data = HIV_DATA)
# Print the model summary
summary(model)
autoplot(model)
================================================================================
  # Plot the estimated cumulative hazard function
plot(model, type = "hazard", main = "Estimated Cumulative Hazard Function")

# Plot the covariate effects
plot(model, type = "covariate", main = "Covariate Effects")

# Plot the estimated survival function
plot(model, type = "survival", main = "Estimated Survival Function")

# Plot the cumulative incidence function
```

```
plot(model, type = "cuminc", main = "Cumulative Incidence Function")
================================================================================
#Plot of the Skewness of PCJ distribution
windows()
par(mfrow=c(1,2))
# Required library
library(plot3D)
# Define the power Chris-Jerry distribution PDF
pdf_power_chris_jerry <- function(x, a, t) {
  (1/(t+2))*(a*t^2)*(x^(a-1))*(1+(t*x^(2*a)))*(exp(-t*x^a))
}
# Define the skewness function for power Chris-Jerry distribution
skewness_power_chris_jerry <- function(a, t) {
  third_moment <- integrate(function(x) x^3 * pdf_power_chris_jerry(x, a, t), lower = 0, upper = Inf)$value
  variance <- integrate(function(x) x^2 * pdf_power_chris_jerry(x, a, t), lower = 0, upper = Inf)$value
  standard_deviation <- sqrt(variance)
  (third_moment / standard_deviation^3)}
# Parameter values
a <- seq(0.5, 2, length.out = 100)    # Range of 'a' values
t <- seq(0.1, 2, length.out = 100)    # Range of 't' values
# Compute skewness for each combination of 'a' and 't'
skewness <- outer(a, t, Vectorize(function(a, t) skewness_power_chris_jerry(a, t)))
# Create the 3D plot using persp()
persp(a, t, skewness, theta = 30, phi = 20, col = "orange",
        xlab = "a", ylab = "t", zlab = "Skewness")
# Add legend for the parameter values
legend("bottomleft", legend = c("a = 0.5, 2", "t = 0.1, 2"), col = c("black", "blue"), lty = c(1, 1),
          bty = "n", lwd = 2, cex = 0.8)

================================================================================
#Plot of the Kurtosis of PCJ distribution
  # Required library
  library(plot3D)
# Define the power Chris-Jerry distribution PDF
pdf_power_chris_jerry <- function(x, a, t) {
  (1/(t+2))*(a*t^2)*(x^(a-1))*(1+(t*x^(2*a)))*(exp(-t*x^a))
}
# Define the kurtosis function for power Chris-Jerry distribution
kurtosis_power_chris_jerry <- function(a, t) {
  mean <- integrate(function(x) x * pdf_power_chris_jerry(x, a, t), lower = 0, upper = Inf)$value
  fourth_central_moment <- integrate(function(x) {
    (x - mean)^4 * pdf_power_chris_jerry(x, a, t)
  }, lower = 0, upper = 10)$value    # Adjust the upper limit for integration
```

```
  variance <- integrate(function(x) {
    (x - mean)^2 * pdf_power_chris_jerry(x, a, t)
  }, lower = 0, upper = 10)$value    # Adjust the upper limit for integration
  (fourth_central_moment / variance^2) - 3
}
# Parameter values
a <- seq(0.5, 2, length.out = 100)    # Range of 'a' values
t <- seq(0.5, 2, length.out = 100)    # Range of 't' values
# Compute kurtosis for each combination of 'a' and 't'
kurtosis <- outer(a, t, Vectorize(kurtosis_power_chris_jerry))
# Create the 3D plot using persp()
persp(a, t, kurtosis, theta = 30, phi = 20, col = "orange", xlab = "a", ylab = "t", zlab = "Kurtosis")
# Add legend for the parameter values
legend("bottomleft", legend = c("a = 0.5, 2", "t = 0.5, 2"), col = c("black", "blue"), lty = c(1, 1),
       bty = "n", lwd = 2, cex = 0.8)
```

==========

Data source

The data sets below represent the age, CD4 counts, opportunistic infection, gender and HIV/AIDS date of diagnosis of patients at St. Luke Hospital Uyo, Nigeria. This data sets can also be found in Udofia *et al.* (2021)

Age of Patients (in years)

21, 37,40,52,30,5,39,32,43,40,36,34,48,27, 37, 25,
25, 47, 40, 18, 32, 24, 32, 35,
49, 45, 12, 30, 31, 35, 49, 30,
4, 45, 8, 13, 33, 33, 21, 23,
32, 40, 5, 23, 33, 42, 60, 12,
30, 27, 40, 35, 40, 43, 50, 55,
29, 25, 28, 6, 35, 33, 36, 55,
35, 25, 35, 32, 36, 40, 27, 40,
63, 42, 40, 65, 28, 34, 32, 32,
14, 60, 28, 45, 18, 45, 30, 38,
40, 25, 30, 10, 36, 26, 25, 60,
40, 55, 40, 35, 37, 26, 37, 32,
50, 31, 28, 46, 36, 40, 25, 12,
40, 38, 32, 28, 30, 57, 40, 53,
49, 64, 40, 20, 42, 35, 32, 24,
28, 30, 63, 34, 25, 40, 42, 34,
39, 24, 54, 37, 20, 31, 30, 41,
35, 38, 54, 56, 28, 25, 42, 21,
30, 30, 50, 30, 29, 40, 25, 26,
25, 46, 28, 25, 45, 34, 53, 27,
40, 32, 38, 31, 26, 23, 35, 30,
40, 48, 32, 37, 35, 40, 28, 35,

31, 24, 26, 54, 27, 27, 17, 40,
30, 27, 40, 59, 39, 29, 23, 30,
35, 30, 40, 50, 31, 37, 40, 32,
58, 42, 54, 33, 28, 28, 26, 65,
30, 52, 23, 34, 25

CD4 Counts of patients

156, 252, 83, 200, 14, 290, 400, 138, 44, 298, 152,
380, 53, 66, 90, 92, 116, 176, 406, 361, 82,
146, 96, 7, 32, 373, 30, 22, 61, 13, 62,
376, 30, 5, 56, 91, 18, 261, 85, 80, 253,
201, 640, 207, 194, 79, 444, 134, 32, 72, 111,
80, 72, 32, 222, 13, 33,
60, 59, 296, 215, 162, 110, 79, 205, 84, 184,
92, 80, 85, 240, 319, 16, 218, 236, 210, 370,
56, 35, 150, 292, 161, 38, 134, 130, 143, 82,
161, 207, 59, 73, 23, 37, 130, 129, 176, 27,
33, 146, 161, 35, 248, 100, 37, 103, 77, 113,
290, 29, 23, 54, 748, 105, 77, 104, 573, 67,
144, 157, 77, 50, 391, 803, 243, 340, 200, 801,
601, 199, 103, 483, 32, 515, 51, 299, 228, 125,
572, 112, 41, 60, 443, 379, 219, 204, 16, 254,
461, 497, 153, 152, 44, 187, 301, 201, 472, 190,
42, 282, 587, 309, 65, 36, 686, 191, 80, 241,
275, 292, 345, 277, 731, 251, 77, 293, 101, 436,
106, 287, 290, 231, 198, 240, 161, 117, 177, 240,
245, 222, 298, 445, 86, 92, 237, 212, 841, 266,
360, 300, 363, 351, 128, 65, 345, 96, 156, 253,
387, 267, 118, 381, 388, 278, 84, 149, 235, 207,
337, 471, 938, 366

HIV/AIDS diagnostic time (in months)

36, 30, 31, 85, 102, 42, 27, 28, 33, 37, 71, 78,
13, 4, 9, 45, 28, 36, 38, 36, 98, 78,
84, 88, 70, 110, 110, 97, 79, 103, 16, 83,
85, 87, 52, 57, 54, 45, 71, 58, 39, 99,
99, 12, 98, 18, 13, 75, 30, 82, 91, 86,
100, 56, 95, 11, 30, 8, 5, 1, 49,
96, 106, 39, 95, 80, 8, 91, 80, 105,
109, 108, 70, 54, 46, 104, 33, 32, 103,
42, 27, 25, 42, 18, 102, 33, 104, 36,
51, 82, 95, 98, 84, 85, 22, 84, 105,
81, 104, 56, 54, 88, 83, 95, 25, 3,
63, 15, 14, 31, 6, 50, 61, 61, 54,

62, 54, 49, 64, 28, 27, 26, 27, 27,
27, 27, 27, 27, 27, 69, 43, 75, 27,
24, 39, 24, 53, 18, 10, 36, 42, 14,
107, 107, 67, 63, 107, 25, 6, 29, 29,
29, 28, 28, 28, 28, 28, 28, 28, 28,
28, 35, 28, 27, 27, 27, 27, 27, 27,
27, 27, 27, 27, 26, 27, 27, 27, 27,
27, 27, 26, 26, 26, 23, 26, 26, 27,
26, 26, 26, 26, 25, 25, 31, 26, 25,
26, 26, 26, 25, 25, 34, 25, 25, 25,
25, 25, 25, 25, 25, 25, 25, 25, 25,
25, 25, 24, 24, 25, 25

Gender (male = 1, female = 0)

1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,1,
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0

Opportunistic infection diagnostic date (months)

36,30,35,85,103,42,28,35,29,34,37,97, 79,14,4,24,45,63,37,39,37,99,79
91,89,70,111,111,98,82,105,17,84,93,88,60,57,55,46,72,79,40,104,104,12,
103,18,16,76,30,85,99,87,101,56,95,12,30,13,85,2,85,97,108,39,95,92,102,
91,80,105,110,108,71,105,49,104,33,32,104,43,27,29,43,19,103,35,105,37,
52,84,96,99,85,86,100,84,112,82,107,57,54,96,83,96,26,4,64,40,14,32,15,
50,61,62,54,63,90,50,65,34,28,27,28,91,27,28,92,27,28,69,43,75,28,25,40,
25,56,21,10,37,75,15,108,114,68,65,108,39,27,30,30,30,32,32,28,36,29,29,
32,32,33,36,29,28,28,28,28,30,29,29,29,28,28,107,28,7,28,107,32,28,29,27,
39,34,34, 28, 26, 28,99,27,109,28,26,29,107,102,26,26,26,54,29,26,42,29,26,
26,33,27,25,29,27,27,25,30,35,25,25,26,25,26,26