European Journal of Statistics and Probability
Vol.8, No.2, pp, 14-24, September 2020
Published by *ECRTD-UK*
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

COMPARISON ON PERFORMANCE OF THE LOGNORMAL, LOG LOGISTIC AND WEIBULL DISTRIBUTION ON SURVIVAL OF HIV PATIENTS WITH OPPORTUNISTIC INFECTIONS IN ANAMBRA STATE, NIGERIA

Okpala, S.T.[1] and Okoli, C. N.[2]

[1, 2] Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria

ABSTRACT: *This paper examined performance of the Lognormal, Log logistic and Weibull distributions on survival of HIV patients with opportunistic infections with the following specific objectives: to fit model for the survival of HIV patients with opportunistic infections for the log-normal distribution, log-logistic distribution and Weibull distribution and to determine the best distribution model for estimating the survival of HIV patients with opportunistic infections. The R-software is employed for data analysis. The secondary data used was fitted to the Weibull model, lognormal model and the log logistic model for selecting the best model for estimating the survival of HIV patients with opportunistic infections in Anambra State from 2010 to 2018. The Akaike information criteria (AIC) and the sum of squares error (SSE) were adopted for assessing the performance of the models. The findings of the study showed that the Weibull model recorded the least value of AIC of 3090.021, the log logistic model was found to be the second with value of AIC = 3178.795 and the log normal model showed the highest value of 3203.729.Furthermore, the Weibull model equally showed the least SSE of 1333891, following log normal model that showed SSE = 1458851 and the log logistic model showed the highest sum of square error value of 1719035. From the results of the findings, the Weibull model is selected as the best parametric model for estimating the survival of HIV patients with opportunistic infections in Anambra State.*

KEYWORDS: lognormal model, loglogistic model, weibull model, opportunistic infections

INTRODUCTION

The human immunodeficiency virus (HIV) has already remained a major public health problem worldwide. The final and most serious stage of HIV infection is the acquired immunodeficiency syndrome (AIDS), which severely damages the body's immune system. Since the start of the epidemic, around 78 million people have been infected with the HIV virus and around 35 million people have died from AIDS-related illnesses. As reported by the World Health Organization (WHO), at the end of 2015, approximately 36.7 million people were living with HIV / AIDS worldwide (WHO, 2016).

Presently, there is no functional remedy available for HIV infection. However, antiretroviral therapy (ART) can effectively control the progression of the HIV virus and help patients return to

a relatively healthy and productive life. Although the life expectancy of patients with the human immunodeficiency virus (HIV) has greatly improved with the introduction of antiviral therapy (ART) and highly active antiviral therapy (HAART), HIV and Acquired immunodeficiency diseases (AIDS) continue to be a major health problem threatening people around the world. Current statistics indicate that as of 2013, there were 35 million HIV-positive people worldwide, of which 31.8 million were adults and 3.2 million were children under the age of 5. In addition, global statistics show that 2.1 million people were newly infected with HIV in 2013 (Kedir et al., 2014).

In addition, several prognostic factors, including chronic pathologies associated with immunodeficiency, chronic viral and bacterial infections can complicate treatment. There are evidence which shows that lifespan can be extended and quality of life can be improved significantly by suppressing HIV levels and keeping CD4 count high (over 200). The risk of opportunistic infections is a potentially fatal problem for HIV-infected patients. HIV patients developing infection such as dementia/Encephalitis, tuberculosis, oral thrush, fever and cough has been identified as one of the leading cause of HIV-related death. Hence, the aim of this study is to examine the performance of the Log-normal, Log-logistic and Weibull distributions on survival of HIV patients with opportunistic infections in Anambra state, Nigeria.

REVIEW OF RELATED LITERATURE

Hamidi et al. (2017) focused on identifying the prognostic factors influencing disease progression in patients with human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) in Iran. The study employed the parametric multi-state model to take into account the intermediate event in the analysis. The outcomes of interest in the study were the transition times from HIV diagnosis to AIDS and AIDS to death. The effect of several prognostic factors on both transitions was investigated. The result obtained from the study revealed that using the parametric model that AIDS progression was significantly associated with an increase in age, low education, and a decreased CD4 cell count. Further findings showed that AIDS-related death was significantly associated with male sex, tuberculosis co-infection, antiretroviral therapy and a decreased CD4 cell count.

Erango et al. (2017) demonstrated joint modelling of longitudinal observation of CD4 counts and time-to-death using AFT models under Bayesian settings. They analyzed two of the data sets with various models and found out interesting results on how covariates and shared frailty affect survival outcome of the patients.

Erangoand Goshu (2017) compared the survival of HIV/AIDS patients under ART follow-up in three different hospitals in Ethiopia. Three parametric accelerated failure time distributions: lognormal, loglogistic and Weibull were used to analyze, predict and compare survival probabilities of the patients. The results indicate that the empirical hazard rates of the three data

sets reveal maximal peaks. The patients from Arba Minch hospital were found to have highest event intensity. The AFT loglogistic model was selected as the best fit to each of the data sets. Different covariates except TB infection status were found to affect patients' survival at each of the hospitals. Patients with TB infection at baseline tend to have shorter survival time as compare to one with no TB infection, with significant differences of survive time between the two groups. Patients under follow-up at Shashemene hospital tend have consistently highest survival probabilities in both TB positive and negative groups. Patients from Bale Robe hospital tend to have longest survival time, while those from Arba Minch hospital have shortest survival time. Patients with bedridden status have the shortest survival time. The AFT-loglogistic was recommended to be best for modelling time-to-event data considered in this study.

Handiso et al. (2019) modelled the factors that affect survival time of HIV infected patients by using Cox proportional hazard and parametric survival regression models. The Kaplan-Meier and Log Rank Test were used to estimate descriptive analysis. Cox's regression model was employed to identify the covariates that have statistical significant effect on the survival time of HIV infected patients and exponential, weibull, log logistic and log- normal survival regression models were applied to compare efficiency of the models. The overall mean estimated survival time of patients was 51.5 months. The Cox Proportional Hazards Regression Model result revealed that baseline weight, ART adherence, baseline CD4 count, WHO clinical stage, level of education, substance use and TB co-infection of patients are the major factors that affect significantly survival time of HIV infected patients. Among the parametric regression models, based on model Comparison methods, the Weibull regression model is better fit. The Weibull regression model revealed a baseline weight<50 kg, low CD4 count at baseline, no education, WHO stages III and IV, poor ART adherence, co-infection with TB and substance abuse are the categories that reduce the survival probability of HIV infected patients.

Ji et al. (2018) in their study noted that TB infection still places a heavy burden on people infected with HIV in China and other developing countries. They argued that knowledge of the survival of HIV-infected patients with pulmonary tuberculosis (PTB) would provide important information for the clinical management of this population. The results of the study after examining 4,914 patients admitted with HIV infection identified 359 cases of PTB. At the time of diagnosis of PTB, the median CD4 + count of patients was 51 / mm3 and 27.30% of patients were on combined antiretroviral therapy (CART). For the 333 cases included in the survival analysis, the overall mortality was found to be 15.92% during a median follow-up of 27 months. Risk factors such as age older than 60, complications from bacterial pneumonia, delayed diagnosis, CD4 + count below 50 / mm3 and pulmonary atelectasis were found to be responsible for the poor survival of patients. In addition, in patients without CART prior to TB treatment, late initiation of CART (more than 8 weeks after initiation of TB treatment) has been shown to increase mortality while initiation of CART in 4 to 8 weeks after starting anti-TB treatment was associated with the fewest deaths. Consequently, it was found that the timely diagnosis of PTB, the prevention of secondary bacterial

pneumonia by prophylactic management and the optimization of the time of initiation of CART could have significant impacts on the reduction of mortality among populations co-infected with HIV / PTB.

Mkenda et al. (2019) examined statistical approach to the implementation of national directives and policies on HIV / AIDS in health facilities located in the Kilimanjaro region of Tanzania, in particular those related to the prevention of mother-to-child transmission (PMTCT) and HIV Care and Treatment Center (CTC). The study adopted an exploratory analysis of data from 196 HIV-positive women, children and youth who received PMTCT and CTC services in nine (9) health facilities located in the Kilimanjaro region. Exploration and distributional adjustment of the collected data sets indicated that the age at confirmation of HIV for women follows a logarithmic logistic distribution while for children and young people follow a logistic distribution. The average age of HIV infection for the women in the study was 27 years. Most of the children and youths claimed to have acquired HIV infection through vertical transmission. The study revealed that 50% of the children and adolescents targeted by this study had lost at least one parent and 29% had lost both parents at the time of the study. In addition, the majority of women mentioned that the reluctance of their male partners to test for HIV infection was the biggest obstacle to eradicating HIV / AIDS. Similarly, caregivers cited the poor attitude toward HIV testing and retesting among pregnant women as an ongoing challenge in managing HIV / AIDS. However, the study found an encouraging improvement in PMTCT services, which enabled 69% of babies born by HIV-positive women in this study to be HIV-free.

MATERIAL AND METHOD

Method of Data collection

The source of data for this study was secondary collected from three hospitals comprising of Nnamdi Azikiwe University Teaching Hospital, Nnewi, ChukwuemekaOdumeguOjukwu University Teaching Hospital (COOUTH) Awka, and General Hospital Onitsha. The data consist of patient's Age, CD4 count, WHO Stage, Number of Opportunistic Infections (OIs), Gender, and OIs Date of Diagnosis and status. Where death is the event of interest, when there is a death, the status is 1 and when there is lost to follow-up (censored) the status is 0.

Method of Data Analysis

Survival models are important statistical methods to describe and analyze the time-to-death data of HIV/AIDS patients. An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group. In summarizing survival data, the two common functions applied are the survivor function and the hazard rate functions. The basic quantity employed to describe time-to-event process is the survival function, the probability of an individual surviving beyond time t. Moreover, the distribution of survival time is characterized by three functions: the probability density function, the survivorship function, and the hazard function (Klein and Moeschberger, 2003).

Survival data often contain incomplete information which is referred to as censored subjects. Such cases may arise due to no follow up of subjects on the part of the observer or the subject not being able to participate to the end of the study. Censoring and abnormality distribution of survival data raises specific methodological and statistical techniques required for an adequate analysis. Censoring of subjects may be on the right or left. Right censoring: when the patient/subject of interest does not experience event during the study observation period. There are three types of right censoring: right censoring- type 1, right censoring-type 11 and right censoring-type 11 (random censoring). Left censoring: when the patient/subject of interest experiences the event in question before the beginning of the study observation period. Definition of function such as survival and hazard function is important to understanding survival distributions.

Survivor function involves the cumulative survival probability that a subject will have event of interest. If S(t) denotes the survival function of T, then

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(T)\, dT \quad (1)$$

Conversely, the hazard function describes the risk of event in an interval time after time t, assuming that the individual has event of interest to the beginning of the interval. It also known as the instantaneous failure rate, hazard rate or force of mortality; that is, the probability that an individual experiences the event of interest at a time point given that the event has not yet occurred. It can be shown that the hazard function is given by the instantaneous probability of failure at time t divided by the probability of surviving up to time t.

$$h(t) = \lim_{\Delta t \to 0} \frac{pr\left[\left(t \leq T \leq t + \Delta t\right) / T \geq t\right]}{\Delta t} = \frac{f(t)}{S(t)} \quad (2)$$

Where, $pr\left[\left(t \leq T \leq t + \Delta t\right) / T \geq t\right]$ is the probability that a subject fail to experience event of interest in the time interval $(t, t + \Delta t)$ given that subject has survived to time point t. The hazard function or log hazard is usually employed while modeling survival Analysis because it is the instantaneous rate of a subject not likely to experience event of interest at time t given that the subject would have experience the event at any earlier time.

In survival analysis, an accelerated failure time (AFT) model is a parametric model that provides an alternative to the commonly used proportional hazards models for the analysis of survival time data. Under AFT models interest is on measuring the direct effect of the explanatory variables on the survival time instead of the hazard (Erango et al., 2017).

As a general approach to the analysis of time to event data is to plot the hazard function for the observed data and determine whether or not it is consistent with a parametric distribution. If the data follows a parametric distribution, parametric methods are preferred to non-parametric methods for describing and quantifying factors that influence time to event.

The parametric model indicates that the result is assumed to follow a family of distributions of similar shape with unknown parameters. It is only when the value of the parameters is known that the exact distribution is fully specified. For parametric regression models; the data are generally used to estimate the values of the parameters that fully specify this distribution. A parametric survival model is a model in which the survival time (the result) is assumed to follow a known distribution.

Suppose $T_i = \min(t_i, c_i)$ is the observed time for the i^{th} subject, where t_i is the time-to-event and c_i represents the censoring time which is assumed independent of t_i. Hence, $\delta_i = 1$, if event is observed and $\delta_i = 0$ otherwise.

The parametric survival models considered in this study are: (a) Log-Normal, (b) Log-Logistic and (c) Weibull distributions.

(a) The Log-Normal Distribution
The lognormal distribution is expressed as

$$f(t) = \frac{\sqrt{\tau}}{\sqrt{2\pi t}} \ell^{\left(-\frac{\tau}{2}\{\ln(t)-\ln(\mu)\}^2\right)} \quad (3)$$

The survival function for the lognormal distribution is

$$S(t) = 1 - \phi\left(\frac{\ln(t)-\mu}{1/\sqrt{T}}\right) \quad (4)$$

The hazard function can be obtained by diving (3) by (4) given the definition in (2). Hence, this is expressed as

$$h(t) = \frac{f(t)}{S(t)} \quad (5)$$

(b) The log-logistic distribution is expressed as

$$f(t) = \frac{\lambda\ \ell t^{\rho-1}}{\left(1+\ \lambda\ t^{\rho}\right)^2} \quad (6)$$

The survival function for the log-logistic distribution is

$$S(t) = \frac{1}{1+\ \lambda\ t^{\rho}} \quad (7)$$

The hazard function can be obtained by diving (6) by (7) given the definition in (2). Hence, this is expressed as

$$h(t) = \frac{\lambda\ \ell t^{\rho-1}}{1+\ \lambda\ t^{\rho}} \quad (8)$$

Loglogistic and lognormal distributions were found to have hazard rate functions that are non-monotonic; thereby increasing to reach a peak and then declining over time (Erango et al., 2017).

(c) The Weibul distribution is expressed as

$$f(t)=\lambda\ \rho t^{\rho-1}\ell^{(-\lambda t^{\rho})} \quad (9)$$

The survival function for the Weibul distribution is

$$S(t)=\lambda\ \rho t^{\rho-1} \quad (10)$$

The hazard function can be obtained by diving (9) by (10) given the definition in (2). Hence, this is expressed as

$$h(t)=\ell^{(-\lambda t^{\rho})} \quad (11)$$

Tools for assessing the performance of the Models

To determine the performance of the models in this research, the Akaike information criteria and the sum of squares error will be used for selection of the best model.

The Akaike information criteria

The Akaike information criteria (AIC) are an estimator of the out-of-sample prediction error and therefore help in providing the relative quality of the statistical models for a given set of data. AIC estimates the relative amount of information lost by a given model. It should be noted that the less information a model loses, the higher the quality of that model. In estimating the amount of information lost by a model, AIC dealswith both the risk of over-fitting and the risk of under-fitting in the model.

In this study, the AIC used to compare the parametric models is defined as:

$$AIC= -2LL + 2p \quad (12)$$

Where,

LL is the log-likelihood; p is the number of parameters in the model. Smaller value of AIC suggests a better model for data adaptation (Burnham and Anderson, 2002).

The sum of squared errors (SSE)

The sum of squared errors (SSE) is the sum of the squares of the residuals (expected deviations from the actual empirical values of the data). It is a measure of the gap between the data and an estimation model. A small SSE indicates a tight fit of the model to the data. It is used as an optimality criterion in theselection of parameters and the selection of models.

In this study, the SSE used to compare the parametric models is defined as:

$$\mathrm{SSE}=\sum_{i=1}^{n}\left(y_i-\widehat{y}_i\right)^2 \quad (13)$$

Where,

y_i is the ith value of the variable to be predicted

$\hat{y}_i$ is the predicted value of y_i

RESULTS OF DATA ANALYSIS

Table 1: Goodness-of-fit criteria for parametric models

Models	Log likehood	AIC	SSE
Weibull	-1539	3090.021	2533891
Lognormal	-1595.9	3203.729	1458851
Log logistic	-1583.4	3178.795	1719035

Table 2: Performance of the model based on AIC

Models	Log likehood	AIC	Rank
Weibull	-1539	3090.021	1
Lognormal	-1595.9	3203.729	3
Log logistic	-1583.4	3178.795	2

The result obtained in table 2 showed that the Weibull model has least AIC = 3090.021, the log logistic model was found to be the secondwith AIC = 3178.795, and the log normal model with highest value of AIC = 3203.729.

Table 3: Performance of the model based on SSE

Models	Log likehood	SSE	Rank
Weibull	-1539	1333891	1
Lognormal	-1595.9	1458851	2
Log logistic	-1583.4	1719035	3

The result obtained in table 3 showed that the Weibull model has least value of SSE = 1333891, following the log normal model with SSE= 1458851. The log logistic model showed the highest SSE value of 1719035.

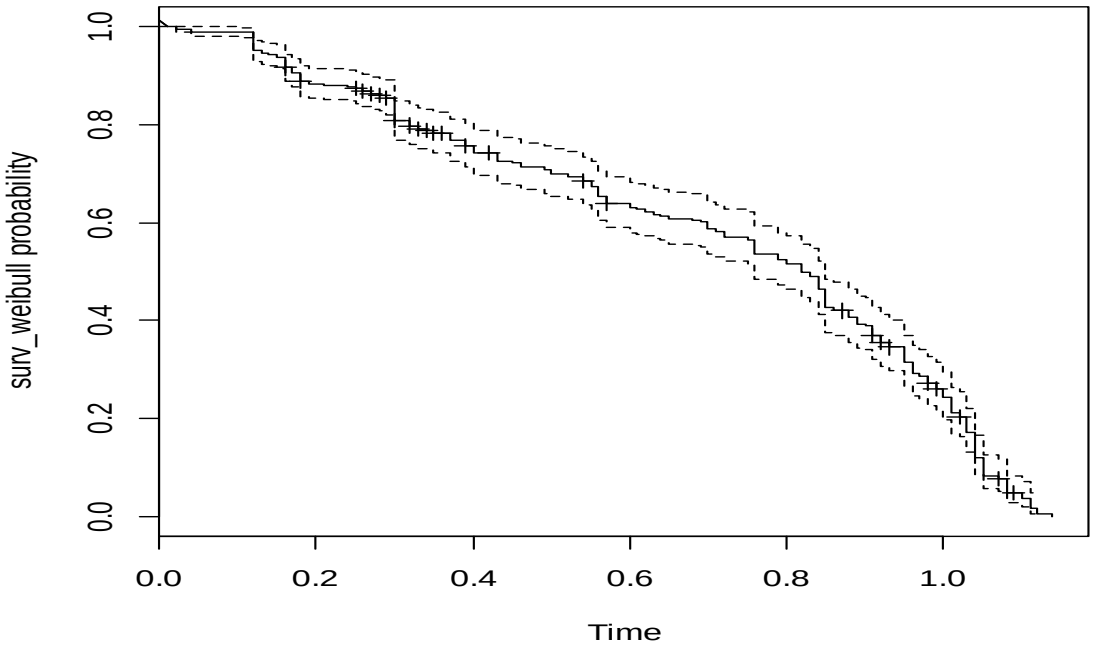

Figure 1: Survival Plot of HIV patients with opportunistic infections in Anambra State for the Weibull Distribution

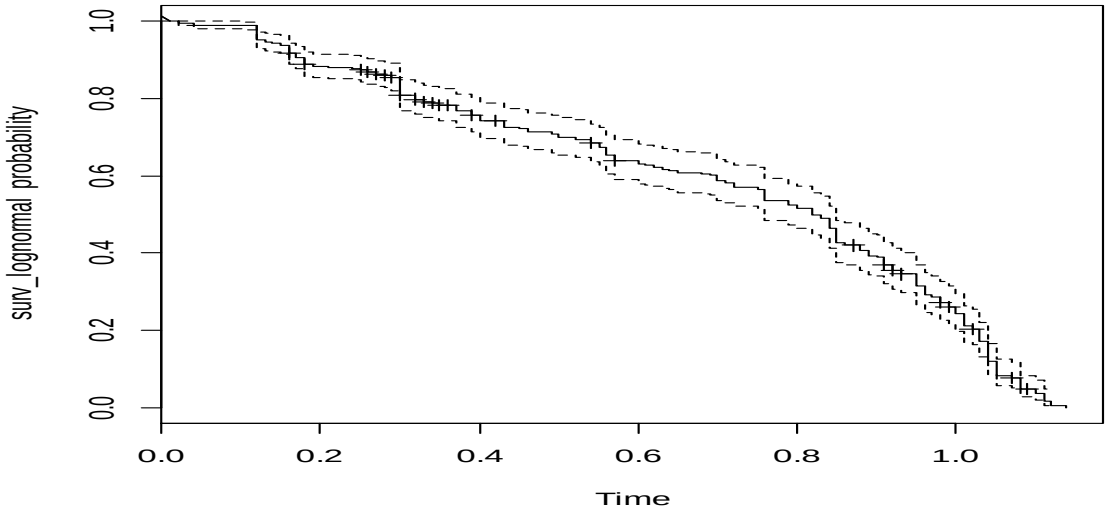

Figure 2: Survival Plot of HIV patients with opportunistic infections in Anambra State for the Lognormal Distribution

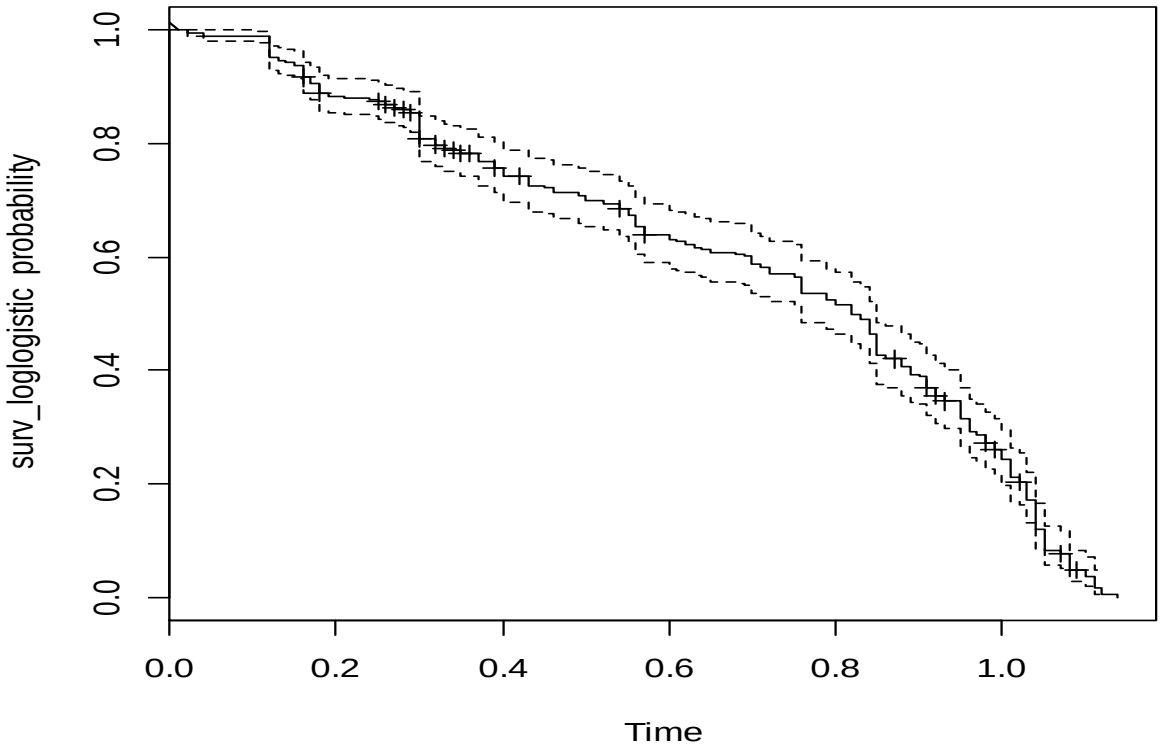

Figure 3: Survival Plot of HIV patients with opportunistic infections in Anambra State for the Loglogistic Distribution

CONCLUSION

This paper examined performance of theLognormal, Loglogistic and Weibull distributions on survival of HIV patients with opportunistic infections in Anambra State. The secondary data collected was fitted to the Weibull model, the lognormal model and the log logistic model with for selecting the best model for determining the survival of HIV patients with opportunistic infections in Anambra State from 2010 to 2018. The AIC and the SSE were adopted for assessing the performance of the models. The findings of the study showed that the Weibull model with least values of AIC and SSE is selected as the best parametric model for determining the survival of HIV patients with opportunistic infections in Anambra State.

Based on the findings of the study, we recommend the use of the Weibull model for estimating the survival of HIV patients with opportunistic infections in Anambra State. Also, health care workers and medical practitioners should anticipate and inform patients about the possible related risk factors of death through early diagnosis and appropriate intervention.

References

Erango, M. A. and Goshu, A. T. (2017). Prediction of Survival of HIV/AIDS Patients from Various Sources of Data Using AFT Models.*Science Journal of Applied Mathematics and Statistics*, 5(4): 127-133. doi: 10.11648/j.sjams.20170504.11

Erango, M. A., Goshu, A. T., Buta, G. B. and Dessiso, A. H. (2017). Bayesian Joint Modeling of Survival of HIV/AIDS Patients using Accelerated Failure Time Data and Longitudinal CD4 Cell Counts. *British Journal of Medicine & Medical Research*, 20(6), 1-12. DOI: 10.9734/BJMMR/2017/32123.

Hamidi, O., Poorolajal, J. and Tapak, L. (2017). Identifying predictors of progression to AIDS and mortality post-HIV infection using parametric multistate model.*Epidemiology Biostatistics and Public Health*, 14(2):1-9. DOI: 10.2427/12438.

Handiso, A. A., Negash, Y., and Mekiso, G. T. (2019).Modeling Time to Death of HIV Infected Patients on Antiretroviral Therapy in case of Hossana Queen Elleni Mohammad Memorial Hospital, South Ethiopia.*International Journal of Public Health and Epidemiology*, 5(1): 071-082.

Ji, Y., Liang, P., Shen, J., Sun, J., Yang, J., Chen, J., Qi, T., Wang, Z., Song, W., Tang, Y., Liu, L., Zhang, R., Shen, Y. and Lu, H. (2018). Risk factors affecting the mortality of HIV-infected patients with pulmonary tuberculosis in the cART era: a retrospective cohort study in China. Infectious Diseases of Poverty, 7(25): 1-9. Doi.org/10.1186/s40249-018-0405-8

Kedir, A. A., Desta, A., and Fesseha, G. (2014).Factors affecting survival of HIV positive children taking antiretroviral therapy at Adama Referral Hospital and Medical College, Ethiopia.*J AIDS Clin Res*.,5(3):1-6.

Klein, J. P. and Moeschberger, M. L. (2003).*Survival Analysis Techniques for Censored and Truncated Data*.SecondEdition.Statistics for Biology and Health, Springer LLC.

Mkenda, T.B., Nokoe, K.S., and Karoki, S. (2019). Statistical Update on the Implementation of HIV/AIDS Guidelines and Policies in Kilimanjaro Region–Tanzania.*Health Science Journal*, 13(2): 640. in the cART era: a retrospective cohort study in China. *Infectious Diseases of Poverty*, 7(25): 1-9. Doi.org/10.1186/s40249-018-0405-8

World Health Organization(2016) .HIV/AIDS.Retrieved July 6, 2016.; Available from: http://www.unaids.org/en/resources/ fact-sheet.