



# American Journal of Epidemiology & Public Health

## Research Article

# Accelerated Failure Time Model with Application to Data on Tuberculosis/Hiv Co- Infected Patients in Nigeria -

Ogungbola OO<sup>1\*</sup>, Akomolafe AA<sup>1</sup> and Musa AZ<sup>2</sup>

<sup>1</sup>Department of Statistics, Federal University of Technology Akure, P.M.B. 704, Akure, Ondo State Nigeria

<sup>2</sup>National Institute of Medical Research, Yaba, Lagos State

**\*Address for Correspondence:** Ogungbola Opeyemi Oyekola, Federal University of Technology, Akure, P. M. B. 704, Akure, Ondo State, Nigeria. Tel.: +234-806-046-4240, ORCID: [orcid.org/0000-0003-0703-8047](https://orcid.org/0000-0003-0703-8047); Researcher ID: [researcherid.com/rid/Q-1754-2018](https://researcherid.com/rid/Q-1754-2018); E-mail: [ooogungbola@futa.edu.ng](mailto:ooogungbola@futa.edu.ng)

**Submitted: 04 July 2018; Approved: 30 August 2018; Published: 03 September 2018**

**Cite this article:** Ogungbola OO, Akomolafe AA, Musa AZ. Accelerated Failure Time Model with Application to Data on Tuberculosis/Hiv Co-Infected Patients in Nigeria. American J Epidemiol Public Health. 2018;2(1): 021-026.

**Copyright:** © 2018 Ogungbola OO, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

In this research, we considered the Accelerated Failure Time (AFT) model for analyzing survival data on Tuberculosis/HIV co-infected patients in Nigeria. We apply the methods to a cohort of these patients managed in tertiary Directly Observed Treatment Short Course (DOTS) Centre, Nigerian Institute of Medical Research (NIMR) for the period of 6-8 months. The effect of the accelerated failure time model was examined in determining the time to sputum smear and culture conversion following initiation of DOTS treatment and study the factors that influence it on TB patients who are co-infected with HIV. The research established that the model provides a better description than other commonly used models of the dataset because it allows prediction of Hazard function, survival functions as well as time ratio. AFT model was able to prepare some insights into the form of the baseline hazard. The result revealed that the weibull AFT model provided a better fit to the studied data. Hence, it is better for researchers of TB/HIV co-infection to consider AFT model even if the proportionality assumption is satisfied.

**Keywords:** Accelerated failure test model; Survival ratio; TB/HIV infection; Absolute lymphocyte count; Body mass Index

## INTRODUCTION

Survival analysis is a statistical method for data analysis where the length of time,  $t_c$  corresponds to the time period from a well-defined start time until the occurrence of some particular event or end-point  $t_e$ , i.e.  $t = t_e - t_c$  [1]. It is a common outcome measure in medical studies for relating treatment effects to the survival time of the patients. In these cases, the typical start time is when the patient first received the treatment, and the end point is when the patient died or was lost to follow-up.

Accelerated Failure Time model (AFT model) is a parametric model that provides an alternative to the commonly used proportional hazards models. Whereas a proportional hazards model assumes that the effect of a covariate is to multiply the hazard by some constant, an AFT model assumes that the effect of a covariate is to accelerate or decelerate the life course of a disease by some constant [2]. In clinical studies, the failure being investigated is often death. Survival analysis describes the methodologies used in Biostatistics to quantify and describe survival time and to examine the magnitude of differences in survival time [3]. Survival analysis is also appropriate for many other kinds of events such as criminal recidivism, divorce, child-bearing, unemployment, and graduation from school [4]. Many studies in statistics deal with deaths or failure of components: the numbers of deaths, the timing of death, and the risks of death to which different classes of individuals are exposed. Then, in Uganda [5] used data to determine whether Tuberculosis (TB) preventive therapies affect the rate of Acquired Immune Deficiency Syndrome (AIDS) progression and survival in Human Immunodeficiency Virus co-infected adult by using Cox proportional hazards (PH) and Accelerated Failure Time (AFT) models and concluded that TB preventive therapies appeared to have no effect on AIDS progression, death and combined event of AIDS progression and death.

The AFT model is another alternative method for the analysis of survival data. This will be studied by means of real dataset which is from a cohort of TB/HIV co-infected patients managed in tertiary Directly Observed Treatment Short (DOTS) Course center for a period of six months among the Nigerian adults.

The time to sputum conversion is the smear-positive pulmonary TB (PTB+) cases registered in a specified period that converted to smear negative status after the standard two months of the intensive phase of treatment.

The main aim of this research is to fit the Accelerated Failure Test models in analyzing TB/HIV co-infected patients. The specific objectives are:

- To determine the effect of the three regimens of TB/HIV preventive therapy on TB/HIV co-infected adults.
- To determine the time to sputum conversion of TB/HIV patients on therapy.
- To determine the effect of the AFT model on time to sputum conversion in TB/HIV patients on therapy.

PH model and parametric AFT models was critically compared by [5]. The major aims of his work was to support the argument for consideration of AFT model as an alternative to the PH model in the analysis of survival data by means of real life data from TB and HIV in Uganda. There are two advantages of Cox proportional regression models, which are ability to incorporate time varying covariate effects and time-varying covariates [6]. The application of survival analysis has extended the importance of statistical methods for time to event data that incorporate time dependent covariates has used by [7]. The estimation of the long-term survivors' proportion was extensively discussed by [8], also termed as "immune" or "cured" proportion, as well as the survival distribution, via parametric or nonparametric approach. In his work and other related works, ordinary maximum likelihood approach was successfully applied via mixture models to obtain estimators of the parameter  $P$  (the "susceptible proportion") and the parameters associated with the proper survival function for the susceptible population, and their large-sample properties are obtained via classic approaches. Some development that dealt with time varying effect of covariates was presented by [9]. He also emphasized the use of semi-parametric models where some effects are time-varying and some are time-constant, thus giving the extended flexibility only for effects where a simple description is not possible. Time-varying effects may be modelled completely non-parametrically by a general intensity model,  $\lambda_i(t) = \lambda(t, X_i(t))$ . Smoothing techniques have been suggested for estimation of  $\lambda(\cdot)$ ; see, e.g., [10] and the references therein. Such a model may be useful when the number of covariates is small compared to the amount of data, but the generality of the model makes it difficult to get a clear, if any, conclusion about covariate effects. Sy Han C proposed fast and accurate inferences for Accelerated failure time models with applications under various sampling schemes was proposed by [11]. In this research, we applied the model into the sputum conversion of the TB/HIV which are co-infected patients managed in tertiary DOTS centre for a period of 6 months among the Nigeria adults. We also make use of the knowledge of percentage of censoring, variation in sample sizes. All these contribute to the existing knowledge.

## METHODOLOGY

The study was carried out and obtained at the DOTS Clinic of the Nigerian Institute of Medical Research (NIMR). A parastatal under the Federal Ministry of Health that has treated over 5000 TB patients in the last 6 years (2011-2016). The Institute has a DOTS centre where it attends to patients infected with TB co-infected HIV.

All patients that were enrolled between 2011 and 2016 were included in the study; this will enable completion of the 6months treatment cycle for those enrolled in 2016.

We adopt AFT model in order to allow acceleration or deceleration of covariates effects which differentiates from PH models and during our analysis we found out that the data reveals just more AFT models and characteristics than other survival models.

### Log rank test

This was used to compare the death rate between two distinct groups, conditional on the number at risk in the groups. The log rank test hypothesis that:

$H_0$ : All survival curves are the same

$H_1$ : Not all survival curves are the same.

Log rank test approximates a chi-square test which compares the observed number of failures to the expected number of failure under the hypothesis.

$$X^2 = \sum_{i=1}^k (o_i - E_i) / E_i \tag{3}$$

where,  $k - 1$  is the degrees of freedom. A large chi-squared value implies a rejection of the null hypothesis for the alternative hypothesis.

## ACCELERATED FAILURE TIME MODEL

### Exponential and Weibull AFT model

The exponential distribution was studied 1<sup>st</sup> in connection with kinetic theory of gasses [4]. The survival function of  $T_i$  can be expressed by the survival function of . If the, has an extreme value distribution then  $T_i$  follows the exponential distribution. The survival function of Gumbel distribution is given by

$$S_{\epsilon_i}(\epsilon) = \exp(-\exp(\epsilon))$$

The Survival function of Weibull AFT model is given by

$$S_i(t) = \exp[-\exp(\log t - \mu - \beta_1 x_1 - \dots - \beta_p x_p)] / \sigma \tag{4}$$

And the cumulative hazard function of Weibull AFT is

$$H_i(t) = -\log s_i(t) = \exp[(\log t - \mu - \beta_1 x_1 - \dots - \beta_p x_p) / \sigma]$$

### Log-normal AFT model

If the, has standard normal distribution then follows the log-normal distribution. The survival function of log-normal AFT model is given by

$$s_i(t) = 1 - \phi[(\log t - \mu - \beta_1 x_1 - \dots - \beta_p x_p) / \sigma] \tag{5}$$

The cumulative hazard function of Log-normal AFT model is

$$H_i(t) = -\log s_i(t) = -\log(1 - \phi[(\log t - \mu - \beta_1 x_1 - \dots - \beta_p x_p) / \sigma])$$

### Log-logistic AFT model

If the, has logistic distribution then follows the log-logistic distribution .The survival function of log-logistic distribution is given by

$$s_{\epsilon_i}(\epsilon) = \frac{1}{1 + e^{\epsilon}}$$

The survival function of log-normal AFT model is given by

$$s_i(t) = \left\{ \frac{1}{1 + e^{\left( \frac{(\log t - \mu - \beta_1 x_1 - \dots - \beta_p x_p)}{\sigma} \right)}} \right\} \tag{6}$$

The cumulative hazard function of log-logistic AFT is given by

$$H_i(t) = -\log s_i(t) = -\log \left( 1 - \exp \left( - \frac{(\log t - \mu - \beta_1 x_1 - \dots - \beta_p x_p)}{\sigma} \right) \right)$$

### Gamma AFT model

In survival literature, two different gamma models are discussed. The Standard gamma model with 2 parameters and the generalized gamma model with 3 parameters. In this study the standard gamma or gamma model is used.

The probability density function of gamma model

$$f(t) = \frac{\alpha \lambda^{\alpha \gamma}}{\Gamma(\gamma)} t^{\alpha \gamma - 1} \exp[-(\lambda t)^{\alpha}] \quad t > 0, \gamma > 0, \Gamma > 0, \alpha > 0$$

Where  $\gamma$  is the shape parameter of the distribution. The exponential, Weibull and log-normal models are all special cases of the generalized gamma model. The generalized gamma distribution becomes the exponential distribution if  $\alpha = \gamma = 1$ , the Weibull distribution if  $\alpha = 1$ , and the log-normal distribution if  $\gamma \rightarrow \infty$ .

## VARIOUS GOODNESS OF FIT TEST

### AIC

To compare various semi-parametric and parametric models Akaike Information Criterion (AIC) is used. It is a measure of goodness of fit of an estimated statistical model. In this study, AIC is computed as follows

$$AIC = -2(\log - \text{likelihood}) + 2(P + K) \tag{7}$$

Where P is the number of parameters and K is the number of coefficients (excluding constant) in the model. For P = 1, for the exponential, P = 2, for Weibull, Log-logistic, Lognormal etc. The model which as smallest AIC value is considered as best fitted model.

For each distribution of, there is a corresponding distribution for T. The members of the AFT model class include the exponential AFT model, Weibull AFT model, log-logistic AFT model, log-normal AFT model, and gamma AFT model. The AFT models are named for the distribution of T rather than the distribution of or  $\log T$  as it is (Table 1).

| Distribution of $\epsilon$  | Distribution of T |
|-----------------------------|-------------------|
| Extreme value(1 parameters) | Exponential       |
| Extreme value(2 parameters) | Weibull           |
| Logistic                    | Log-logistic      |
| Normal                      | Log-normal        |
| Log-Gamma                   | Gamma             |

## MODEL DIAGNOSTICS

The AFT model is well fitted. Log minus log also revealed the proportionality and the nature be it parallel or not.

## RESULTS

First, descriptive statistics are used to give us information about the distributions of the variables. We get the baseline characteristics in 452 participants using the descriptive statistics (Table 2).

Some continuous variables are grouped into categories according to clinical meaning. The K-M curves show the shape of the survival function for each treatment arm. We can see from (Figure 1) that the cumulative survival proportion appears to be much higher in the Anti-TB/HIV therapy (INHRIFEMB, INHRIFPZAEMB and INHPZAEMB) compared to the groups in which INHPZAEMB was used. In INHPZAEMB group, few participants resume this therapy. It would appear that INHRIFPZAEMB and INHPZAEMB of TB/HIV Therapy significantly prolong the time until participants resume event compared to the other interventions. The median survival time is at 40years of age for INHPZARIF combination TB/HIV therapy while 45years is expected for INPZAEMB therapy group. Many are censored in INHRIFEMB before reaching the age of 60years with regarding sputum conversion of TB patients on therapy in (Figure 1).

**Note:** EMB, INH, PZA, RIF and RPT represent Enthambutol, Isoniazid, Pyrazinamide, Rifampin and Rifapentine respectively.

## LOG RANK TEST

$H_0$ : The effect of the three regimens does not have significant factor to TB preventive therapy for TB/HIV co-infected adults.

$H_1$ : Not  $H_0$ : In the table 3, since  $P_{-value} (.0192) < (\alpha = 0.05)$ , the effect of the three regimens does have significant to TB preventive therapy for TB/HIV co-infected adults. Then survival distributions are different in the population which make the result more statistically significance.

By the log-rank test, in the preventive therapy, there is significant difference among three regimens of TB preventive therapy for TB/HIV co-infected adults, since the  $p$ -value is 0.0192 against 5% level of significance. The Kaplan-Meier (K-M) curves for time to educate length and time to combined event of the preventive therapy is presented (Figure 1), the age is just the median of the co-infected patients'.

In (Figure 2), the log minus log plot was not parallel which revealed to us approximately the suitability of AFT. For this reason, the investigation of AFT Model comes into play.

In (Figure 3), the Log-logistic, Weibull and Log-normal Density functions are approximately normal to the curve while others are not similar.

Accelerated failure time models were compared using statistical

| Variable | T       | Age      | BMI     | LYMPHABS | Creat    | Haemo    |
|----------|---------|----------|---------|----------|----------|----------|
| Total    | 3396    | 16627.32 | 158.11  | 97225.20 | 48092.80 | 23759.02 |
| Mean     | 7.5142  | 36.7861  | 0.3498  | 215.1000 | 106.4000 | 52.5642  |
| Std dev  | 1.2338  | 10.7473  | 0.1054  | 194.6253 | 64.3768  | 26.8265  |
| C.V      | 16.4196 | 29.2157  | 30.1315 | 90.4813  | 60.5045  | 51.0357  |

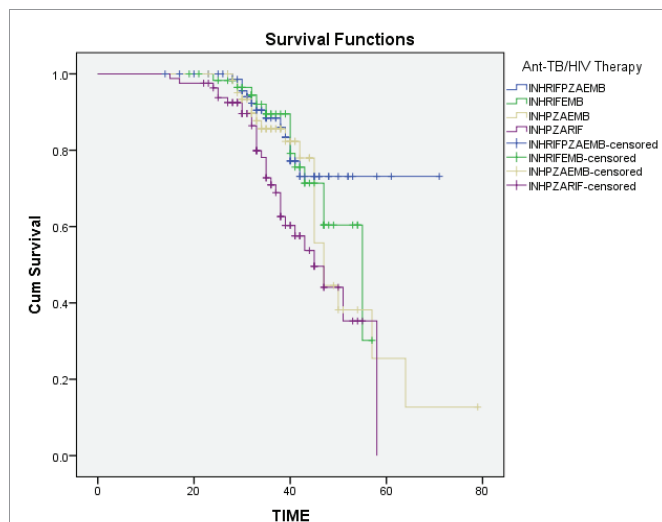


Figure 1: The survival function curve.

Table 3: Test for equality of Survival Distribution for Different level of TB/HIV Therapy.

|                                | Chi-Square | df | Sig.  |
|--------------------------------|------------|----|-------|
| Log Rank (Mantel-Cox)          | 9.930      | 3  | 0.019 |
| Breslow (Generalized Wilcoxon) | 8.570      | 3  | 0.036 |
| Tarone-Ware                    | 9.055      | 3  | 0.029 |

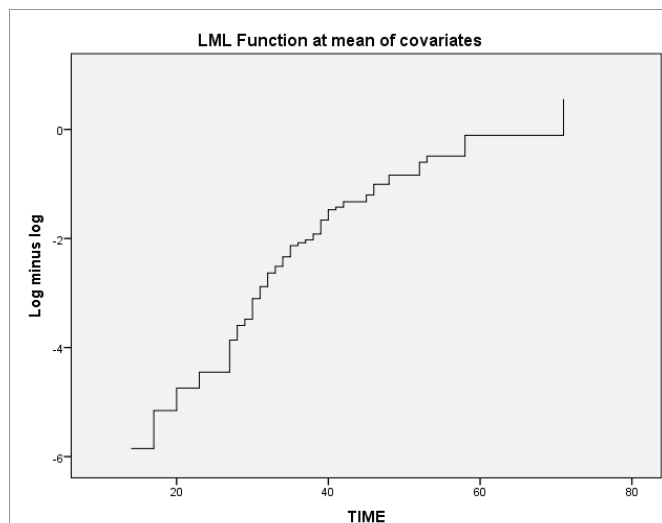


Figure 2: Log minus log plot.

criteria (likelihood ratio test and AIC). The nested AFT models can be compared using the Likelihood Ratio (LR) test. The models in (Table 4 and 5) reveal that (covariates) age and sex are statistically significant while HAEMO GLUC, BMI and LYMPHABS are not significant with their  $p$ -value greater than 0.05, whereas the (covariates) age, sex and haemo are statistically significant with their  $p$ -values. In the AFT model, log-logistic AFT model and the Weibull AFT model are nested within the log-normal AFT model (Table 6). According to the Log-likelihood Ratio (LR) test, the weibull model fits better. However, the LR test is not valid for comparing models that are not nested. In this case, we use AIC to compare the models (Table 7) (The smaller AIC

is the better). The Weibull AFT model appears to be an appropriate AFT model according to AIC compared with other models, although it is only slightly better than Log-logistic or Log-normal model. We also note that the Log-normal model is poorer fit according to LR test and AIC. At last, we conclude that the Weibull model is the best fitting the AFT model based on AIC criteria.

### DISCUSSION AND CONCLUSION

In this paper, attempt has been made to find the best fitted model for studying the survival time TB/HIV patients. To meet the objectives various AFT models following distributions Weibull, Log-normal, Log-logistic are fitted to survival data of TB/HIV patients collected from NIMC, Yaba, Lagos. Different statistical measures such as AIC, Cox- Snell AFT residuals plots and Log minus are used to find the best fitted model. From the results, the Weibull AFT model is found to be the best fitted model.

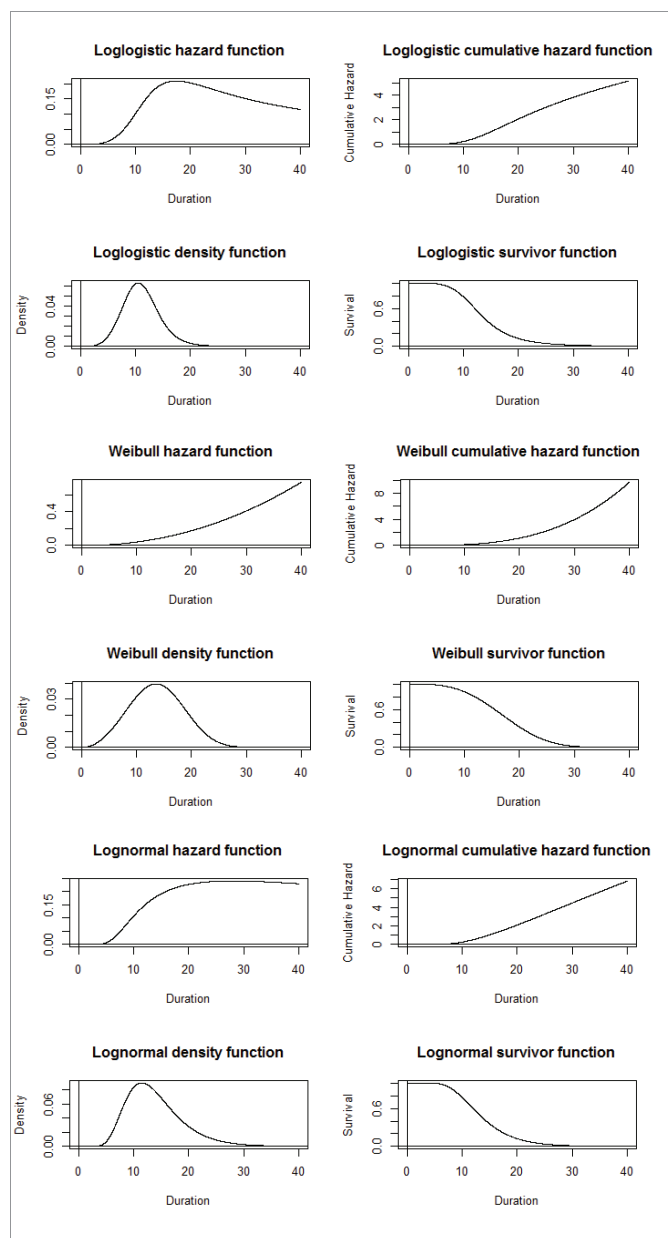


Figure 3: The plot for AFT MODEL: Log-logistic, Weibull, Log-normal with their respective.

**Table 4:** Log-logistic AFT Model.

| Covariates     | $\beta$ | TR    | Sig.  | 95.0% CI for TR) |       |
|----------------|---------|-------|-------|------------------|-------|
|                |         |       |       | Lower            | Upper |
| AGE            | 0.022   | 1.571 | 0.035 | 1.210            | 1.932 |
| SEX            | -0.308  | 0.735 | 0.014 | -0.286           | 1.756 |
| LYMPHAB        | -0.013  | 0.989 | 0.689 | 0.922            | 1.056 |
| HAEMO          | 0.133   | 1.146 | 0.457 | 0.797            | 1.495 |
| CREAT          | 0.000   | 0.999 | 0.984 | 0.987            | 1.011 |
| BMI            | 0.561   | 1.753 | 0.410 | 0.528            | 2.978 |
| WEIGHT         | -0.061  | 0.928 | 0.510 | 0.738            | 1.118 |
| GLUC           | -0.022  | 0.978 | 0.168 | 0.947            | 1.009 |
| Log-likelihood | 225.156 |       |       |                  |       |

**Table 5:** Weibull AFT model.

| Covariates     | $\beta$ | TR    | Sig.  | 95.0% CI for TR) |       |
|----------------|---------|-------|-------|------------------|-------|
|                |         |       |       | Lower            | Upper |
| AGE            | 0.018   | 1.018 | 0.023 | 0.618            | 1.418 |
| SEX            | -0.258  | 0.773 | 0.042 | 0.142            | 1.404 |
| LYMPHAB        | -0.011  | 0.919 | 0.500 | 0.852            | 0.986 |
| HAEMO          | 0.136   | 1.146 | 0.438 | 0.801            | 1.491 |
| CREAT          | 0.000   | 0.999 | 0.984 | 0.981            | 1.009 |
| BMI            | 0.336   | 1.396 | 0.371 | 0.659            | 2.133 |
| WEIGHT         | -0.075  | 0.908 | 0.440 | 0.718            | 1.098 |
| GLUC           | -0.022  | 0.978 | 0.145 | 0.949            | 1.007 |
| Log-likelihood | 218.079 |       |       |                  |       |

**Table 6:** Log-normal AFT model.

| Covariates     | $\beta$ | TR    | Sig.  | 95.0% CI for TR) |       |
|----------------|---------|-------|-------|------------------|-------|
|                |         |       |       | Lower            | Upper |
| AGE            | 0.026   | 1.026 | 0.041 | 0.391            | 1.661 |
| SEX            | -0.158  | 0.854 | 0.036 | 0.437            | 1.271 |
| LYMPHAB        | -0.014  | 0.989 | 0.659 | 0.928            | 1.050 |
| HAEMO          | 0.146   | 1.158 | 0.009 | 0.842            | 1.474 |
| CREAT          | 0.000   | 0.999 | 0.079 | 0.987            | 1.011 |
| BMI            | 0.627   | 1.858 | 0.349 | 0.903            | 2.813 |
| WEIGHT         | -0.061  | 0.928 | 0.465 | 0.763            | 1.093 |
| GLUC           | -0.023  | 0.977 | 0.852 | 0.945            | 1.008 |
| Log-likelihood | 235.009 |       |       |                  |       |

**Table 7:** The log-likelihoods and Likelihood Ratio (LR) tests, for comparing alternative AFT models.

| Distribution | No of parameter | Log-likelihood | Testing against the Log-normal distribution |    |
|--------------|-----------------|----------------|---|----|
|              |                 |                | LR  | df |
| Log-logistic | 2               | -100.532       | 326.46                                      | 1  |
| Weibull      | 3               | -263.762       | 440.452                                     | 2  |
| Log-normal   | 2               | -43.536        |   |    |

In this study, the result shows that Weibull AFT model is better than the other models in case of explaining the survival esophagus cancer data. The factor, TB/HIV directed treatments, has a significant role in case of survival of TB/HIV patients. The patients who undergo the TB/HIV directed treatment other than surgery has lower risk of dying than the patients who has underwent the treatment of surgery and its combinations according to the laid down principles.

This study is based on a large number of participants from Lagos residents in Nigeria, where the prevalence of TB infection and HIV

**Table 8:** Akaike Information Criterion (AIC) in the AFT models.

| Distribution | Log-likelihood | k | c | AIC     |
|--------------|----------------|---|---|---------|
| Log-logistic | -100.532       | 6 | 2 | 225.156 |
| Weibull      | -263.762       | 6 | 1 | 218.079 |
| Log-normal   | -43.536        | 6 | 2 | 235.019 |

are very high. In this study, the Cox model and the Accelerated Failure Time model have been compared using TB/HIV co-infected data. Association of the TB/HIV preventive therapies with the sputum conversion is examined through the linkage of the signs and symptoms to replication of the virus.

The AFT model provides an estimate of the survival function time ratios. In this research, we have analyzed the TB/HIV dataset the methods. This study provides an example of a situation where the AFT model is appropriate and the description of the data reveal that log-minus-log plot is not parallel.

### ACKNOWLEDGEMENT

We like to appreciate the management of National Institute Medical Research (NIMR) for their ethical approval to make use of their health survival data. God bless them all.

### REFERENCES

1. Ata N, Sozer M. Cox regression models with non-proportional hazards applied to lung cancer survival data. Hacettepe Journal of Mathematics and Statistics. 2007; 36: 157-167. <https://goo.gl/4ZpTy6>

2. Wikipedia the free encyclopedia. The accelerating failure time model. 2018; <https://goo.gl/Qwfk3L>

3. World Health Organization (WHO). Global tuberculosis control: surveillance, planning, financing. Geneva. 2006; 79. <https://goo.gl/i3EqAZ>

4. Whalen CC, Johnson JL, Okwera A, Hom DL, Huebner R, Ellner JJ, et al. A trial of three regimens to prevent tuberculosis in Ugandan adults infected with the human immunodeficiency virus. N Engl J Med. 1997; 337: 801-808. <https://goo.gl/T1J5Bw>

5. Jiezhi Qi. Comparison of proportional hazards and accelerated failure time models. A Master of Science Thesis Submitted to the College of Graduate Studies and Research in the Department of Mathematics and Statistics, University of Saskatchewan Saskatoon, Saskatchewan Canada. 2009; <https://goo.gl/9qQo6o>

6. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society. 1972; 34: 187-220. <https://goo.gl/QAzMQ6>

7. Kazeem AA, Abiodun AA, Ipinyomi RA. Semi-parametric non-proportional hazard model with time varying covariate. Journal of Modern Applied Statistical Methods. 2015; 14: 68-87. <https://goo.gl/GtVXQC>

8. Maller RA, Zhou X. Survival analysis with long-term survivors. New York: Wiley; 1996. <https://goo.gl/rovUPc>

9. Scheike TH. Time-varying effects in survival analysis. Handbook of Statistics. 2003; 23: 61-85. <https://goo.gl/5Cx1aY>

10. Nielsen JP, Linton OB. Kernel estimation in a nonparametric marker dependent hazard model. Ann Statist. 1995; 23: 1735-1748. <https://goo.gl/WyDHaa>

11. Sy Han C. Statistical methods and computing for Semi-parametric and accelerated failure time model with induced Smoothing. 2013. <https://goo.gl/vhrBo6>