

# A Smooth Test of Goodness-of-Fit for the Weibull Distribution: An Application to an HIV Retention Data

Collins Odhiambo<sup>1</sup>, John Odhiambo<sup>1</sup> and Bernard Omolo<sup>2,\*</sup>

<sup>1</sup>Strathmore Institute of Mathematical Sciences, Strathmore University, Ole Sangale Road, Nairobi, Kenya

<sup>2</sup>Division of Mathematics & Computer Science, University of South Carolina-Upstate, 800 University Way, Spartanburg, South Carolina, USA

**Abstract:** In this study, we fit the two-parameter Weibull distribution to an HIV retention data and assess the fit using a smooth test of goodness-of-fit. The smooth test described here is a score test and is derived as an extension of the Neyman's smooth test. Simulations are conducted to compare the power of the smooth test with the power of each of three empirical goodness-of-fit tests for the Weibull distribution. Results show that the smooth tests of order three and four are more powerful than the three empirical goodness-of-fit tests. For validation, we used retention data from an HIV care setting in Kenya.

**Keywords:** Goodness-of-fit, Loss to follow-up, Neyman's smooth test, Retention in HIV care, Weibull distribution.

## 1. BACKGROUND

The Kenya AIDS Response Progress Report 2014 [1] estimates the number of people living with HIV (PLHIV) to be 1.6 million, while those actively on anti-retroviral therapy (ART) to be 800,000. Retaining patients on ART, however, remains a major challenge. Disruption in care through missed scheduled visits undermines both social as well as clinical outcomes, including risk of virological failure [2]. Discontinuation of ART can lead to drug resistance, HIV-related illnesses and death. It has been shown that individuals who miss visits in the first year of treatment have a higher mortality rate [2]. Studies also show that retention of patients who are on ART treatment remains stable after 12 months of ART initiation, with loss to follow-up (LTFU) being the main cause of attrition [3]. In resource-limited settings, it is common to find patients dropping out of ART treatment. The dropouts are usually attributed to LTFU [4]. Due to significant dropouts, patients may not realize the benefits of ART. Previous studies have singled out associations between frequent LTFU and more severe opportunistic illnesses [5]. More innovation is therefore required for further ART scale-up and to improve retention in care.

Patients who are actively receiving ART are particularly vulnerable to developing drug-resistant infections when virological failure occurs, which could potentially result in broad resistance to ART and transmission of drug-resistant viruses [2]. Determining

correct patterns of LTFU and exploring factors associated with them is therefore crucial in identifying the patients who are at-risk of LTFU. Further, analyses of time to LTFU is useful in informing development of evidence-based interventions that improve patient outcomes [2].

In this paper, we fit a two-parameter Weibull distribution to time to LTFU data (HIV retention data). The probability distribution in this context provides an ideal tool to perform specific risk and probability calculations, and apply the results to make well-informed decisions. This is in cognizance of the fact that an inappropriate distribution may lead to incorrect calculations and eventually wrong decisions. Therefore, finding a distribution that fits the data best is important, particularly to avoid time and financial loss that may arise from an invalid model selection. More about importance and reasons for fitting probability distributions correctly to the data have been elaborated by Hahn and Shapiro [6].

Time-to-failure data arises in many fields (e.g medical, public health, reliability, etc.) and is often assumed to follow the Weibull distribution. The popularity of the Weibull distribution is mainly because of its flexibility and ability to describe survival data [7]. Here, we adopt a smooth test of goodness-of-fit to fit the two-parameter Weibull distribution. The smooth test considered here is a score test obtained by extending the Neyman's goodness-of-fit approach [8], where the score test is obtained by nesting the null hypothesis in a larger class of probability distribution functions [9]. Our study also seeks to validate the result in [10], which showed that the smooth tests for the Weibull distribution are more powerful than several other tests, including the Crammer-Von Moses, Kolmogorov-

\*Address correspondence to this author at the Division of Mathematics & Computer Science, University of South Carolina-Upstate, 800 University Way, Spartanburg, South Carolina, USA; Tel: +1 864-503-5362; Fax: +1 864-503-5930; E-mail: bomolo@uscupstate.edu

Sminorv and Andersen-Darling tests. Furthermore, fitting the Weibull distribution to data simulated from the log-normal and gamma distributions can be reliably assessed by smooth tests [11].

We fit the two-parameter Weibull distribution to an HIV retention data from two public hospitals in Kenya and assess the fit using a smooth test. Our event of interest is time to first LTFU. Other exits from the program (i.e. death, transfer-outs) are removed. Patients who are also active on ART at the end of the observation period are removed. From a routine programmatic arrangement, patients started on ART are expected to regularly attend clinic for either continuous monitoring or drug refill. A patient is considered to be LTFU if s/he fails to show-up within 48 hours after a scheduled appointment and s/he cannot be reached in any way. Testing for goodness of fit for LTFU is therefore important not only to determine which distribution best fits the failure-time data but also to enhance the accuracy in estimating the hazard function, which indicates the instantaneous risk of LTFU. In our formulation, we propose a score test for the null hypothesis that the failure-time data follows a two-parameter Weibull distribution.

To the best of our knowledge, there is no published article on fitting a parametric distribution to time to LTFU data. This paper will therefore significantly contribute the statistical methodology for analysing LTFU data that could give information that is vital in improving HIV care retention and streamline national policy in HIV programming.

In the section 2, we present the methods for our analysis, which includes the description of our data, modelling time to LTFU, the formulation of smooth goodness-of-fit for the Weibull distribution and simulations. We also discuss the performance of the smooth test. In section 3, we discuss the application and results of the test to a locally available HIV retention data. In section 4, we discuss the simulation and validation results. Finally, in section 5, we provide some concluding remarks.

## 2. METHODS

### 2.1. Smooth Test of Goodness-of-fit

#### 2.1.1. Theoretical Framework

Suppose we wish to test the null hypothesis that  $X_1, X_2, \dots, X_n$  is a random sample from a specified continuous distribution with probability density function

$f(x; \boldsymbol{\beta})$ , where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m)^T$  is a vector of nuisance parameters. We can construct a smooth test by extending the Neyman's goodness-of-fit approach of nesting the null hypothesis in a larger class of probability distribution functions [8, 9].

We embed the null probability density function in an order  $k$  alternative as follows:

$$\varphi_k(x, \boldsymbol{\theta}, \boldsymbol{\beta}) = C(\boldsymbol{\theta}, \boldsymbol{\beta}) \exp\left\{\sum_{i=1}^k \theta_i h_i(x, \boldsymbol{\beta})\right\} f(x, \boldsymbol{\beta}), \tag{1}$$

where  $h_i(x, \boldsymbol{\beta})$  is orthonormal to  $f(x, \boldsymbol{\beta})$  and  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^T$  is a vector of real parameters and  $C(\boldsymbol{\theta}, \boldsymbol{\beta})$  is a normalizing constant that ensures  $\varphi_k(x, \boldsymbol{\theta}, \boldsymbol{\beta})$  integrates to one, i.e.

$$\int_{-\infty}^{\infty} \varphi_k(x, \boldsymbol{\theta}, \boldsymbol{\beta}) dx = \int_{-\infty}^{\infty} C(\boldsymbol{\theta}, \boldsymbol{\beta}) \exp\left\{\sum_{i=1}^k \theta_i h_i(x, \boldsymbol{\beta})\right\} f(x, \boldsymbol{\beta}) dx = 1. \tag{2}$$

Testing for the goodness-of-fit of  $f(x; \boldsymbol{\beta})$  is equivalent to testing  $H_0: \boldsymbol{\theta} = \mathbf{0}$  against  $H_A: \boldsymbol{\theta} \neq \mathbf{0}$ . Assuming that the partial derivatives of the log-likelihood function together with their expectations exist, the derivation of the score test statistics using the maximum likelihood function for the observed random sample  $X_1, X_2, \dots, X_n$  has extensively been discussed by [9, 10, 12, 13].

The log-likelihood function is defined as

$$\log L = n \log C(\boldsymbol{\theta}; \boldsymbol{\beta}) + \sum_{j=1}^n \sum_{i=1}^k \theta_i h_i(x_j, \boldsymbol{\beta}) + \sum_{j=1}^n \log(f(x_j), \boldsymbol{\beta}). \tag{3}$$

Given the orthonormal condition for  $\boldsymbol{\theta} = \mathbf{0}$ ,

$$E_0[h_r(X; \boldsymbol{\beta})h_s(X; \boldsymbol{\beta})] = n\delta_{rs}. \tag{4}$$

The partial derivatives for the log-likelihood function (3) generates the score function  $U_{\boldsymbol{\theta}}$  and the asymptotic covariance matrix  $\Sigma$  of  $U_{\boldsymbol{\theta}}$  as follows [9, 14]:

$$U_{\boldsymbol{\theta}} = (h_r(X_1; \boldsymbol{\beta}) + h_r(X_2; \boldsymbol{\beta}) + \dots + h_r(X_k; \boldsymbol{\beta})) \tag{5}$$

and

$$\Sigma = I_{\theta\theta} - I_{\theta\beta} I_{\beta\beta}^{-1} I_{\beta\theta}, \tag{6}$$

where

$$\frac{\partial}{\partial \theta_r} \int_{-\infty}^{\infty} \varphi_k(x, \boldsymbol{\theta}, \boldsymbol{\beta}) dx = \frac{1}{C(\boldsymbol{\theta}, \boldsymbol{\beta})} \frac{\partial C(\boldsymbol{\theta}, \boldsymbol{\beta})}{\partial \theta_r}$$

$$\int_{-\infty}^{\infty} \varphi_k(x, \boldsymbol{\theta}, \boldsymbol{\beta}) dx + \int_{-\infty}^{\infty} h_r \varphi_k(x, \boldsymbol{\theta}, \boldsymbol{\beta}) dx$$

$$= \frac{\partial \log C(\boldsymbol{\theta}, \boldsymbol{\beta})}{\partial \theta_r} + E_r[h_r] = 0.$$

Therefore

$$\frac{\partial \log C(\boldsymbol{\theta}, \boldsymbol{\beta})}{\partial \theta_r} = -E_r[h_r]. \tag{7}$$

$$I_{\theta\theta} = (I_{\theta\theta})_{rs} = -E_k \left[ \frac{\partial^2 \log L}{\partial \theta_r \partial \theta_s} \right] = n\delta_{rs}, \tag{8}$$

$$I_{\theta\beta} = (I_{\theta\theta})_{ru} = -E_k \left[ \frac{\partial^2 \log L}{\partial \theta_r \partial \beta_u} \right] = n\text{cov}_0 \left( h_r, \frac{\partial \log f}{\partial \beta_u} \right), \tag{9}$$

$$I_{\beta\beta} = (I_{\theta\theta})_{uv} = -E_k \left[ \frac{\partial^2 \log L}{\partial \beta_u \partial \beta_v} \right] = n\text{cov}_0 \left( \frac{\partial \log f}{\partial \beta_u}, \frac{\partial \log f}{\partial \beta_v} \right), \tag{10}$$

and

$$I_{\beta\theta} = (I_{\theta\theta})_{ur} = -E_k \left[ \frac{\partial^2 \log L}{\partial \beta_u \partial \theta_r} \right] = n\text{cov}_0 \left( \frac{\partial \log f}{\partial \beta_u}, h_r \right). \tag{11}$$

The score statistic therefore takes the form  $S(\hat{\boldsymbol{\beta}}) = U_{\theta}^T \hat{\Sigma}^{-1} U_{\theta}$ . Here the score function  $U_{\theta} = U_{\theta}(\boldsymbol{\beta})$  has  $r^{\text{th}}$  element  $(h_r(X_1; \boldsymbol{\beta}) + h_r(X_2; \boldsymbol{\beta}) + \dots + h_r(X_n; \boldsymbol{\beta})) / \sqrt{n}$  and  $\Sigma$  is the asymptotic covariance matrix of  $U_{\theta}$ .

But

$$\Sigma = I_{\theta\theta} - I_{\theta\beta} I_{\beta\beta}^{-1} I_{\beta\theta} = nM, \tag{12}$$

where

$$M = I_k - \text{cov}_0 \left( h, \frac{\partial \log f}{\partial \boldsymbol{\beta}} \right) \{ \text{var}_0 \left( \frac{\partial \log f}{\partial \boldsymbol{\beta}} \right) \}^{-1} \text{cov}_0 \left( \frac{\partial \log f}{\partial \boldsymbol{\beta}}, h \right), \tag{13}$$

which essentially reduces to  $M = I_k$  and the score test takes the form  $S(\hat{\boldsymbol{\beta}}_0)$ .  $\hat{\boldsymbol{\beta}}_0$  is the maximum likelihood estimator of  $\boldsymbol{\beta}$  under the null hypothesis and

$$S(\hat{\boldsymbol{\beta}}_0) = U_{\theta}^T(\hat{\boldsymbol{\beta}}_0) \hat{M}^{-1} U_{\theta}(\hat{\boldsymbol{\beta}}_0) = \frac{1}{n} \sum_{j=1}^p \hat{V}_j^2, \tag{14}$$

where  $\hat{V}_j = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_r(x_i; \hat{\boldsymbol{\beta}}_0)$ . The score statistic for testing  $H_0: \boldsymbol{\theta} = \mathbf{0}$  against  $H_A: \boldsymbol{\theta} \neq \mathbf{0}$  is denoted by  $S_k(\hat{\boldsymbol{\beta}})$ . The choice of  $k$  depends on  $\hat{\boldsymbol{\beta}}$  through the model-dependent modified Bayes information criterion (modBIC) given by

$$\text{modBIC}_k = \hat{S}_k - k \log n,$$

in which, relative to BIC, twice the maximized log-likelihood has been replaced by the score statistic. We define  $k$  as the smallest order that maximizes  $\text{modBIC}_k$  i.e.

$$k = \min \{ k \in \{1, 2, \dots, d\} \text{ and } \text{modBIC}_k \geq \text{modBIC}_r, r = 1, 2, \dots, d \}.$$

This is also referred to as the selection rule [9].

**Theorem 2.1** Under null hypothesis, as the sample size  $n$  becomes large ( $n \rightarrow \infty$ ),  $S(\hat{\boldsymbol{\beta}}) \sim \chi_k^2$ .

Proof

Consider a sequence of independent and identical orthonormal samples

$h_r(X_1; \hat{\boldsymbol{\beta}}), h_r(X_2; \hat{\boldsymbol{\beta}}), \dots, h_r(X_n; \hat{\boldsymbol{\beta}})$ . By the orthonormality condition,

1.  $h_0 = 0$ ,
2.  $E_0[h_i(X; \hat{\boldsymbol{\beta}})] = 0$ ,
3.  $E_0[h_i(X; \hat{\boldsymbol{\beta}})h_j(X; \hat{\boldsymbol{\beta}})] = \delta_{ij} = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$

Define  $V_{i1}$  as

$$V_{i1} = \frac{h_i(X_1; \hat{\boldsymbol{\beta}}) - nE_0[h_i(X_1; \hat{\boldsymbol{\beta}})]}{\sqrt{n\delta_{11}}}. \tag{15}$$

The variable  $V_{i1}$  represents a standard score. Using the orthonormal conditions (i.e.  $E_0[h_i(X; \hat{\boldsymbol{\beta}})] = 0$  and  $\delta_{11} = 1$ ), the variable reduces to

$$V_{i1} = \frac{1}{\sqrt{n}} h_i(X_1; \hat{\boldsymbol{\beta}}). \tag{16}$$

Applying the Central Limit Theorem, for each identical and independently distributed  $V_{i1}, V_{i2}, \dots$  the sum of the standard score  $\phi_i = V_{i1} + V_{i2} + \dots$  tends to the standard normal distribution with mean 0 and variance 1 as the size becomes sufficiently large. That is

$$\phi_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{\sqrt{n}} h_i(X; \hat{\boldsymbol{\beta}}) \rightarrow N(0, 1). \tag{17}$$

Since the sum of squares of standard normal variates is  $\chi^2$  distributed,

$$S(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^k \phi_i^2 \sim \chi_k^2. \tag{18}$$

The test statistic of the smooth test ( $\sum_{i=1}^k \phi_i^2$ ) only depends on order  $k$  and not the sample size  $n$ . The smooth test of goodness-of-fit therefore stands out as the preferred test compared to other empirical GOF procedures that are usually affected by the sample size. Another feature of the  $S(\hat{\beta})$  statistic is that its components can be used to indicate alternative probability distribution that would fit a given dataset [15].

**2.1.2. Smooth Test for the Weibull Distribution**

The two-parameter Weibull distribution is defined as

$$f(x; \beta, \eta) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\}, \tag{19}$$

where  $\eta$  is the scale parameter and  $\beta$  is the shape parameter.

The orthonormal polynomials for the Weibull distribution for the first four orders are derived from the Extreme Value distribution because the distribution approaches Weibull as the sample size  $n$  becomes large.

The first five orthonormal polynomials are given as below [9, 16]:

$$h_0 = 1,$$

$$h_1 = \frac{6}{\pi} Z,$$

$$h_2 = \frac{Z^2 - \frac{12\zeta(3)Z}{\pi} - \frac{\pi^2}{6}}{\sqrt{\frac{11}{90}\pi^4 - \frac{24}{\pi^2}\zeta(3)}},$$

$$h_3 = \frac{2,121}{20,000} Z^3 - \frac{309}{625} Z^2 - \frac{1,097}{5,000} Z + \frac{5,583}{10,000},$$

and

$$h_4 = \frac{1}{40} Z^4 - \frac{3,021}{12,500} Z^3 + \frac{269}{1,000} Z^2 + \frac{7,769}{10,000} Z - \frac{22,588}{100,000},$$

where  $Z = \frac{x - \mu}{\sigma - \gamma}$ ,  $\gamma$  is the Euler's constant approximated to be 0.57722, and  $\zeta(3)$  involves the  $\zeta$  function approximated to be 1.20206. The score statistic is given as

$$S(\hat{\beta}, \hat{\eta}) = \sum_{r=2}^k \hat{V}_r^2 = \hat{S}_{k-2}, \tag{20}$$

where  $\hat{V}_r = \frac{1}{\sqrt{n}} \sum_{j=1}^n h_j$ . The first four values of the score statistics for the two-parameter Weibull distribution are:

$$S_1 = 0,$$

$$S_2 = 0,$$

$$S_3 = \frac{\sqrt{b_1} - 1.139547}{\sqrt{20/n}},$$

and

$$S_4 = \frac{b_2 - 7.55\sqrt{b_1} + 3.21}{\sqrt{219.72/n}},$$

where

$$\sqrt{b_1} = \frac{1}{n} \sum_{j=1}^n [(X - \bar{X}) / S]^3, \tag{21}$$

and

$$b_2 = \frac{1}{n} \sum_{j=1}^n \{(X - \bar{X}) / S\}^4. \tag{22}$$

We reject the null hypothesis for large values of the test statistics.

**2.2. Empirical Goodness-of-fit Tests**

Since we are dealing with complete data, we employ other standard empirical goodness-of-fit methods. The conventional empirical goodness-of-fit tests considered here are the Anderson-Darling ( $A^2$ ), the Kolmogorov-Smirnov ( $D_n$ ) and the Cramer-von-Mises ( $\omega^2$ ) tests.

These tests are based on the departure between the empirical distribution function,  $\mathbb{F}_n$ , and theoretical distribution function,  $\mathbb{F}_n$ , of the sampled data. The null hypothesis is rejected when the difference is too large, which would suggest that the sampled data does not come from the underlying distribution. For the case of a Weibull distribution, we consider the Extreme Value Distribution [17] and therefore apply the empirical cumulative distribution function of  $\ln(X_i)$  instead of  $X_i$  [12, 16]. A measure of the difference from the empirical cumulative distribution function of  $\ln(X_i)$  is computed against the estimated theoretical cumulative distribution function using the maximum likelihood method. That is,  $\mathbb{F}_n(x) = \mathbb{F}(x; \ln \hat{\eta}_n, \frac{1}{\hat{\beta}})$ . The goodness-of-fit test statistics are defined as follows.

Anderson-Darling (AD) statistic:

$$A^2 = -n + \frac{1}{n} \sum_{i=1}^n [(2i-1) - 2n] \ln(1 - \hat{U}_i^*) - (2i-1) \ln(\hat{U}_i^*)]. \quad (23)$$

One-sample Kolmogorov-Smirnov (KS) statistic:

$$D_n = \sqrt{n} \sup |F_n(x) - \hat{F}_0(x)| \quad (24)$$

$$= \sqrt{n} \max[\max\{\frac{i}{n} - U_i^*\}, \max\{U_i^* - \frac{i-1}{n}\}]. \quad (25)$$

Cramer-von Mises (CM) statistic:

$$\omega^2 = \sum_{i=1}^n (\hat{U}_i^* - \frac{2i-1}{2n})^2 + \frac{1}{12n}, \quad (26)$$

where  $\hat{U}_i^* = \hat{F}_0(\ln X_i)$ .

### 2.3. Simulation Studies

Simulations were conducted to compare the critical values and power of the empirical goodness-of-fit tests (Anderson-Darling ( $A^2$ ), Kolmogorov-Smirnov test ( $D_n$ ) and Cramer-von-Mises ( $\omega^2$ )) and the smooth test. All computations were performed using the R package *EWGoF*. We generated samples from the two-parameter Weibull distribution, with scale and shape parameters set at 30 and 6, respectively. The number of Monte Carlo runs in each case was set at 1000. Samples of size  $n \in \{5, 20, 50, 100, 500, 1000\}$  were generated and estimates of rejection probabilities computed. We compared the performance of the GOF tests at the 1%, 5% and 10% significance levels. Power was obtained as the percentage of rejection of the null hypothesis.

**Table 1: The Power of Goodness-of-Fit Tests for the Simple Hypothesis  $H_0$  (the Weibull Distribution with Parameters: Scale=30 and Shape=6) Versus the Hypothesis  $H_1$ : a Class of the Weibull Distribution (for the Smooth Test) and Not Weibull (for EDF Tests)**

The power of Smooth Test (order 3)							
$\alpha$	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
.01	0.015	0.006	0.007	0.006	0.006	0.007	0.008
.05	0.053	0.041	0.046	0.035	0.039	0.040	0.047
.1	0.103	0.103	0.090	0.089	0.085	0.091	0.088
The power of Smooth Test (order 4)							
$\alpha$	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
.01	0.016	0.007	0.008	0.008	0.010	0.008	0.009
.05	0.055	0.042	0.049	0.043	0.042	0.042	0.051
.1	0.105	0.104	0.096	0.087	0.088	0.101	0.090
The power of Kolmogorov-Smirnov Test							
$\alpha$	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
.01	0.014	0.010	0.012	0.012	0.012	0.012	0.012
.05	0.049	0.055	0.049	0.056	0.060	0.065	0.069
.1	0.107	0.106	0.097	0.098	0.098	0.113	0.108
The power of Anderson-Darling Test							
$\alpha$	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
.01	0.013	0.010	0.010	0.011	0.011	0.013	0.013
.05	0.047	0.052	0.048	0.055	0.055	0.060	0.066
.1	0.113	0.122	0.095	0.104	0.115	0.118	0.120
The power of Cramer-Von Mises Test							
$\alpha$	n=5	n=20	n=50	n=100	n=200	n=500	n=1,000
.01	0.014	0.013	0.011	0.011	0.012	0.013	0.013
.05	0.050	0.055	0.050	0.053	0.047	0.060	0.067
.1	0.112	0.117	0.096	0.099	0.107	0.116	0.123

Table 1 shows the power of the smooth tests of order 3 and order 4 for the Weibull distribution against the three common empirical distribution function tests (KS, CVM and AD).

Results show that the performance of all the tests is strongly linked to the shape of the simulated distribution. The empirical tests are biased in situations where  $\hat{\beta}$  is fairly minimal (close to 1), whereas when  $\hat{\eta}$  is sufficiently large (i.e.  $\hat{\eta} > 30$ ), they fail to detect the right distribution (Weibull). For smaller samples, the smooth tests are generally more powerful than empirical GOF tests. The components of the smooth test tend to be unbiased when  $\beta = 5$ ,  $\eta = 35$  for large samples (i.e.  $n \geq 500$ ).

The calculation of the asymptotic distributions of the EDF statistics is as follows:

$\phi_n = (F_n(z) - z) / \sqrt{n}$ , where  $F_n(z)$  is the EDF of the set of  $z_i$  and tends to normality  $\phi(z)$  as  $n \rightarrow \infty$ .

### 3. APPLICATION TO LFTU DATA

#### 3.1. Data Description

We conducted a retrospective data analysis for all patients who were initiated ART at two public hospitals in Nairobi, Kenya (Makadara Health Center and Lungalunga Health Center) between October 1, 2011 and December 31, 2014. Considering that ART services were initiated in Kenya in 2003 across all public hospitals, we specifically extracted data from 2011 because by then all the public systems, processes and structures for defaulter tracing were expected to have picked up effectively. Our event of interest was time to first LTFU. The clinical setting considered here is routine regular Comprehensive Care Center (CCC) in typical government hospitals. Data is collected routinely whenever patients come for clinical check-up or drug refill. Since time to first LTFU was the event of interest, other exits (i.e. transfer outs and deaths) were not considered in the analysis. Patients who were actively receiving ART services and did not experience the event were also removed. Only patients who were observed from the time of ART initiation between November 1, 2011 and December 31, 2014, were included in the analysis. The time between ART initiation to first LTFU was given in months. Time to first LTFU was defined as missing routine clinical appointment within 48 hours from the scheduled appointment date and not identified as "active on ART", "dead", or "transferred-out". The time to first LTFU was calculated as the time interval

between the dates of ART initiation and first drop out, as recorded by the ART database, *IQCare*. The cohort was stratified by gender (male and female), WHO Staging (WHO Stage 1, WHO Stage 2, WHO Stage 3 and WHO Stage 4) considered at the time of ART initiation and age groups (<10 years, 10 -14years, 15-24 years and 25+ years). Data was retrieved from a Health Information System(HIS) called *IQCare* without patients identifiers. Only variables of interest were pulled out to an Excel spreadsheet. Data was stored in Excel and thereafter analysed in *R*. Approval was obtained from Pathfinder International.

#### 3.2. Motivation for Analysis of HIV Retention Data

Patients receiving ART can experience LTFU, which may result in discontinuation of treatment, drug toxicity, treatment failure due to poor adherence and drug resistance. This can result in an increased risk of death of up to 40% of ART patients in sub-Saharan Africa [18]. Studies have shown that LTFU has negative impact on immunological benefit of ART and increases AIDS-related morbidity, mortality, and hospitalizations [18]. Individuals who miss visits in the first year of treatment have a higher mortality rate [2]. Stephen and co-authors [4] showed that retention of patients who are on ART treatment remains stable after 12 months of ART initiation, with loss to follow-up being the main cause of attrition [3]. Previous studies have also illustrated associations between frequent LTFU and more severe opportunistic illnesses [5]. Analysis of LTFU have also been used in HIV care to monitor and improve programme effectiveness, using patient retention as a measure of quality of care [19].

The main objective in the analysis of LTFU data was to check retention of patients in care. Programmatically, this is considered an important determinant of successful ART long-term outcomes. Patients who experience LTFU essentially get enrolled in other facilities with different regimen combinations, which is likely to compromise their immune system. Retaining patients for long allows provision of long term Highly Active Antiretroviral Therapy (HAART), tracking WHO staging, tracking immunosuppression profiles and evaluation of emergence of medication toxicities. In resource-limited settings, it is common to find patients dropping out of ART treatment. Due to significant drop-outs, patients may not realize the benefits of ART if they are LTFU. More innovation is therefore required for further ART scale-up and improve retention in care.

### 3.3. Modelling LTFU

The focus is on time to first LTFU, and data in this perspective is primary and has not been published or utilised in any publication. This is a typical Kenyan case, however, different types of LTFU are expected to reflect the general evolution of HIV programming. LTFU is expected to be a stable event that does not evolve much with time, at least in adults. However, in young children, the risk of the event is not likely. Heterogeneity is to be expected as it is well-known that there are various degrees of LTFU. The fact that LTFU is considered as a stable event in any HIV programming suggests that at least in adults there is no event-dependence and no time-dependence. The start time is the time of enrolment on ART. Patients are expected to come for drug refill and routine check-up. During the observation period, a patient can remain active (i.e. does not miss regular appointments), die, transferred-out or LTFU.

### 3.4. Cohort Description

A total of 4,981 patients were initiated ART between November 1, 2011 and December 31, 2014 in two public hospitals. Out of those initiated on ART, 854 patients experienced LTFU and were therefore included in the analysis. The table below shows the patients' status.

**Table 2: Patients' Status**

Status	Frequency	Percentage
Active on ART	2,392	48
Transfer Out	1,405	28
Dead	330	7
First Lost-to-follow up	854	17
Total	4,981	100

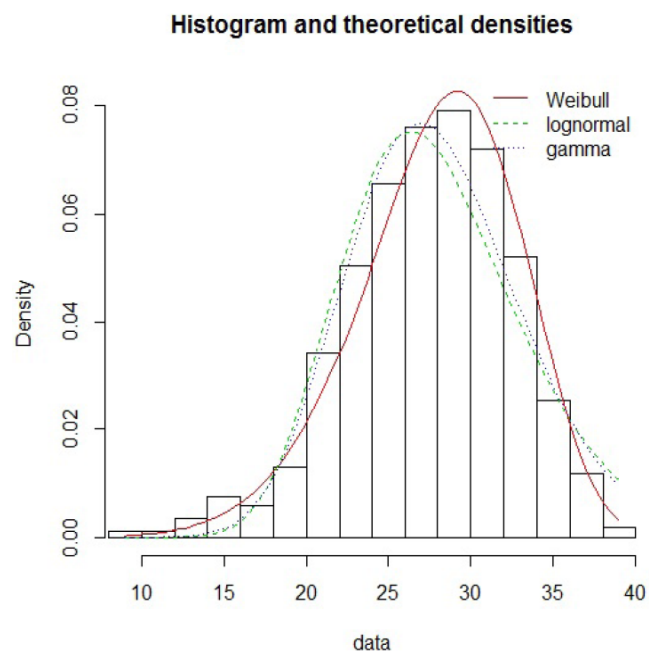
**Table 3: Patients Baseline Characteristics at ART Initiation**

Patients Characteristics	<15 years (Children) (%)	15-24 years (Adolescents) (%)	25+ years (Adults) (%)
Gender (N=854)	87(4%)	458(52%)	309(34%)
Male (N=345)	45(5%)	185(28%)	115(68%)
Female (N=509)	42(8%)	273(54%)	194(38%)
WHO Staging (N=854 )	60(7%)	258(30%)	536(63%)
WHO Stage I(N=376)	18(5%)	112 (30%)	216(65%)
WHO Stage II(N=308)	29(9%)	87(28%)	192(62%)
WHO Stage III(N=136)	8 (6%)	41(30%)	87 (87%)
WHO Stage IV(N=34)	5(15%)	18 (53%)	81 (81%)

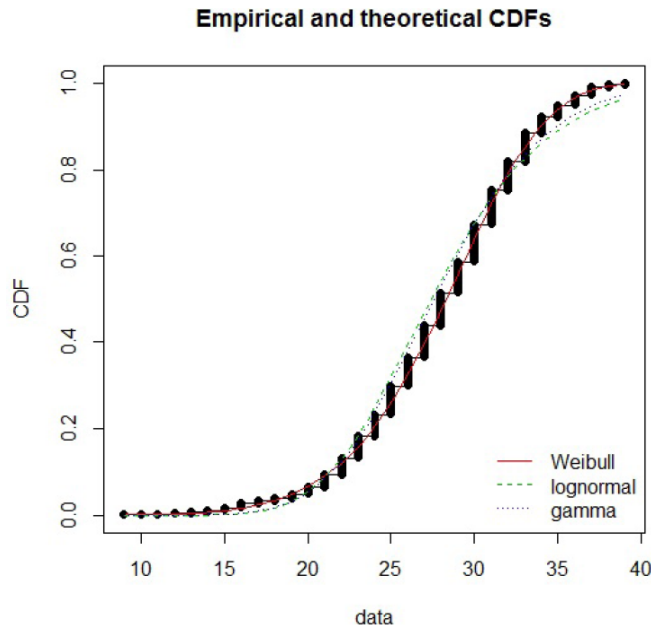
The median age of those lost to follow-up ( $n = 854$ ) was 34.2 years (IQR 30.4 – 38.4), and 59% ( $n=509$ ) of them were female. Forty five percent of patients had advanced/severe immunodeficiency at the start of treatment, and 20% had WHO clinical stage 3 or 4 disease. The mean CD4 count was 449 (SD 9.3) at baseline. Characteristics at baseline during ART initiation is given below.

#### 3.4.1. Graphical Assessment

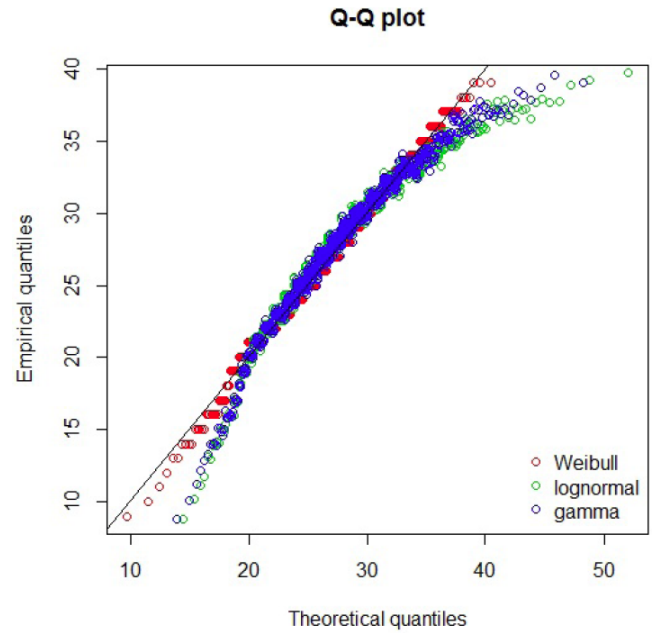
We obtained probability plots to assess the validity of statistical distributions [20] to time to first LTFU data [21]. Graphically, the Weibull distribution seems to be close to the P-P plot line compared to the Gamma and log-normal distribution. See Figures 1, 2, 3 and 4 below.



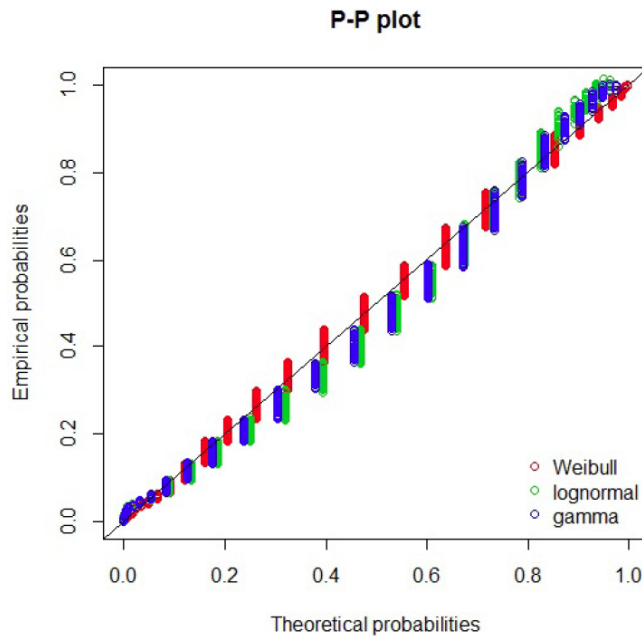
**Figure 1: Test for Theoretical Distributions. The Weibull Distribution is closer to the distribution.**



**Figure 2:** Test for Theoretical Distributions. Here, the data appears to be more coherent with Weibull distribution.



**Figure 4:** Test for Theoretical Distributions. The QQ-plot shows the Weibull distribution is much coherent with the data.



**Figure 3:** Test for Theoretical Distributions. The PP-plot indicates that the Weibull distribution is the ideal distribution.

**3.4.2. Model Fitting**

In order to assess the model fit for the Weibull distribution, we obtained the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). We also fitted the log-normal and the Gamma distributions to the same data. The AIC and BIC for the Weibull distribution (AIC = 5,276.6, BIC = 5,286.1) are lower than that of the Gamma distribution (AIC = 5,777.8, BIC = 5,787.3) and the log-normal (AIC =

6,009.4, BIC = 6,018.9) distributions. Therefore the model that fits the data best is the Weibull distribution.

**3.5. Tests Performance**

The smooth test generated here is constructed using orthonormal functions as opposed to quadratic forms [12]. For the smooth test, the score statistics of order 3 and order 4 are given. The hypothesis regarding the distributional form is rejected by the three empirical distribution tests if their respective test statistic,  $D_n$ ,  $\omega^2$  and  $A_n$  are greater than the critical value obtained from their tabulated values. Also their p-values are considerably lower than the significance level of 0.01. These tests ( $D_n$ ,  $\omega^2$  and  $A_n$ ) are more powerful whenever the sample size is not large. In our situation, however, with a sample size of size of 854, the tests are misleading. The smooth test, on the other hand, rejects the hypothesis when considering upto order four and the p-value is quite large compared with the  $D_n$ ,  $\omega^2$  and  $A_n$  tests.

**3.6. Application Results**

To demonstrate the importance of the smooth test of goodness-of-fit in a real life application, we examined an HIV retention data and fit a two parameter Weibull distribution to LFTU data. We assessed the fit using smooth tests of order 3 and 4 and then compared the results with the three empirical GOF tests. Other exits from the program (i.e. death, transfer-outs and active-on-ART) were removed.



**Table 4: Comparison of the AIC and BIC for the Weibull, Gamma and Lognormal Distributions. All the Parameter Estimates were Obtained by the MLE Method**

Param	Weibull	Gamma	Log-normal
	shape=6.786; scale=30.145	shape=14.958; rate=0.532	meanlog*=3.303; sdlog**=0.299
AIC	5,276.633	5,777.821	6,009.374
BIC	5,286.133	5,787.321	6,018.874

\*mean of the natural logarithm of LTFU, \*\*standard deviation of the natural logarithm of LTFU.

**Table 5: Tests Comparison (N = 864)**

Test type	Test Statistics	p-value
One-sample Kolmogorov-Smirnov test	$D_n = 0.055232$	0.01092
Cramer-von Mises test	$\omega^2 = 1.0238$	7.947e-12
Anderson-Darling test	$A_n = 1.9603$	0.09659
$S_3$	S = 1.2529	0.308
$S_4$	S = 0.66308	0.409

Essentially, we tested the null hypothesis that the Weibull distribution is the underlying distribution of time to first LTFU. The maximum likelihood estimates of the scale and shape parameters under the Weibull model were  $\eta = 30.145$  and  $\beta = 6.786$ , respectively, and the resulting values of the test statistics were  $S_3 = 1.2529, p = 0.308$  and  $S_4 = 1.66308, p = 0.409$ . Hence the Weibull null hypothesis could not be rejected when using smooth tests, suggesting that the Weibull model is the best model for the duration between the start of ART and first LTFU. In comparison, the three empirical GOF tests rejected the null hypothesis. The one-sample Kolmogorov-Smirnov test ( $D_n = 0.055232, p = 0.01092$ ), Cramer-von Mises test ( $\omega^2 = 1.0238, p = 7.947e - 12$ ) and the Anderson-Darling test ( $A_n = 1.9603, p = 0.09659$ ) indicated significance deviations from the null distribution. This suggests that the smooth test is the most reliable test compared with the rest whenever the sample size is sufficiently large.

#### 4. DISCUSSION

We have considered an extension of the smooth tests to non-censored data. For the application to non-censored data, we provided the orthonormal structure so that the smooth tests of order  $k = 3$  and  $k = 4$  are computed. We then evaluated the goodness-of-fit for the two-parameter Weibull distribution fitted on HIV retention data. The empirical GOF tests considered here were powerful whenever the sample size is small. This is consistent with [9, 14, 15, 22]. Our contribution in this article is unique in the sense that we are

considering LTFU data generated from a typical clinical setting.

Lemeshko [23] investigated the gamma distribution with its parameters chosen so that it is closest to the Weibull distribution. The power test was used to assess both simple and composite hypotheses against the simple alternative. Although he found the Kolmogorov-Smirnov, Cramer-von-Mises and Anderson - Darling type nonparametric tests to be most powerful compared with the case when the estimates are found by minimizing the corresponding statistics, the comparison did not take into consideration the smooth tests.

Sururu [24] also did power comparisons in a simulation study of goodness-of-fit tests but the smooth test was not included in his assessment. Few authors have incorporated smooth tests when assessing the power of goodness-of-fit test. In particular, Ledwina *et al.* [22] showed through simulations that the data-driven version of Neyman's smooth tests performs very well over a wide range of alternatives and is competitive with other data-driven procedures. They also showed that the data driven smooth tests are consistent against essentially all alternatives [22]. Rayner and Best [15] also showed that Neyman smooth test for location-scale families are flexible and can be chosen to improve detection of particular alternatives. These tests were shown to perform well against its competitors. This assessment is also consistent with our simulation results.

The smooth test approach fails to reject the null hypothesis when considering up to order four. The  $p$ -value is quite large compared with the EDF tests and is therefore more appropriate when compared to other alternatives for large sample sizes. The smooth GOF test also produced the best estimate of the distribution of the data (Weibull distribution), which will ultimately result in a better estimate of the hazard function for time to LTFU for predicting hazard rates. The Weibull distribution is the best choice from the density plots and graphs and is validated by smooth GOF test. It is important to carefully choose the best GOF tests in order to make the correct inference about the underlying distribution. We have shown that the smooth test is superior and can be used to analyze time to LTFU data in order to determine the underlying distribution. These results agree with those of [25], who showed by simulation that the test for normality based on smooth test has much greater power than the generalized  $\chi^2$  test and the Kolmogorov-Smirnov test. Kang [25] also demonstrated that the test performs generally as well as the Shapiro-Wilk, skewness, and kurtosis tests for a wide range of alternatives.

Several studies have shown that LTFU poses challenges to the successful implementation of ART programs [18]. Studies have shown that patients who discontinued ART developed a rapid increase in viral load and depletion of CD4 cells, putting them at risk of opportunistic infections and early death. Therefore, understanding the underlying pattern and distribution of LTFU is necessary to making sound interventions that maintain adherence to ART treatment. In this study, the two parameter Weibull distribution fits the time to first LTFU well. Several authors [3, 5, 19, 26-28] have shown that the main reason in rising cases of LTFU to be poor patient's defaulter tracing in resource-limited settings. This is likely to compromise positive outcomes of ART in a large scale HIV care center. Pattern of LTFU are therefore crucial in developing practical programmatic interventions.

## 5. CONCLUSION

Loss to follow-up (LTFU) is an important problem both for the care of individuals and the evaluation of anti-retroviral treatment (ART) programmes in low- and middle-income countries. But the evaluation of ART programmes has been difficult because many patients are lost to follow-up [29]. Thus, it is important to model LTFU so as to better understand the factors influencing LTFU. Most of these models are based on the two-parameter Weibull distribution (e.g. [29, 30]) and are

often selected without a prior test of goodness-of-fit of this distribution. Our study, however, evaluated the performance of the smooth test against that of the empirical goodness-of-fit tests for the Weibull distribution and validated the results using loss to follow-up data from HIV retention care.

The smooth goodness-of-fit approach performed better than the empirical GOF tests when fitting a parametric distribution to time-to-event complete data. We described how to fit a two parameter Weibull distribution to an HIV retention data and assess the fit using goodness-of-fit procedures. Our results highlight the need to better understand LTFU of patients initiated on ART. Estimation of nuisance parameters can be performed without changing the test statistics and since the tests rely on maximum likelihood techniques, they asymptotically meet the conditions of the Neyman-Pearson lemma against any simple alternative hypothesis. Future studies should address fitting hazard functions based on the Weibull distribution to censored data in order to determine risks to LTFU in HIV care.

## ACKNOWLEDGEMENTS

The authors wish to acknowledge Pathfinder Kenya for providing the *IQCare* database for analysis. BO was partially supported by a grant from the Simons Foundation (# 282174).

## REFERENCES

- [1] Kenya AIDS Response Progress Report 2014: Progress towards Zero; 2014.
- [2] Rachlis B, Cole DC, van Lettow M, Escobar M, Muula AS, Ahmad F, *et al.* Follow-Up Visit Patterns in an Antiretroviral Therapy (ART) Programme in Zomba, Malawi. *PloS one* 2014; 9(7): e101875.
- [3] Rasschaert F, Koole O, Zachariah R, Lynen L, Manzi M, Van Damme W. Short and long term retention in antiretroviral care in health facilities in rural Malawi and Zimbabwe. *BMC Health Services Research* 2012; 12(1): 444. <https://doi.org/10.1186/1472-6963-12-444>
- [4] Asiimwe SB, Kanyesigye M, Bwana B, Okello S, Muyindike W. Predictors of dropout from care among HIV-infected patients initiating antiretroviral therapy at a public sector HIV treatment clinic in sub-Saharan Africa. *BMC Infectious Diseases* 2016; 16(1): 43. <https://doi.org/10.1186/s12879-016-1392-7>
- [5] Haddow L, Edwards S, Sinka K, Mercey D. Patients lost to follow up: experience of an HIV clinic. *Sexually Transmitted Infections* 2003; 79(4): 349-350. <https://doi.org/10.1136/sti.79.4.349-b>
- [6] Hahn GJ, Shapiro SS. Statistical models in engineering. In: *Statistical models in engineering*. John Wiley & Sons 1968.
- [7] Cousineau D. Fitting the three-parameter Weibull distribution: Review and evaluation of existing and new methods. *IEEE Transactions on Dielectrics and Electrical Insulation* 2009; 16(1): 281-288. <https://doi.org/10.1109/TDEI.2009.4784578>

- [8] Neyman J. Smooth test for goodness of fit. *Scandinavian Actuarial Journal* 1937; 1937(3-4): 149-199. <https://doi.org/10.1080/03461238.1937.10404821>
- [9] Rayner JC, Thas O, Best DJ. *Smooth tests of goodness of fit: using R*. John Wiley & Sons 2009.
- [10] Ledwina T. Data-driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association* 1994; 89(427): 1000-1005. <https://doi.org/10.1080/01621459.1994.10476834>
- [11] Kopecky KJ, Pierce DA. Efficiency of smooth goodness-of-fit tests. *Journal of the American Statistical Association* 1979; 74(366a): 393-397.
- [12] Bargal AI, Thomas DR. Smooth goodness of fit tests for the Weibull distribution with singly censored data. *Communications in Statistics-Theory and Methods* 1983; 12(12): 1431-1447. <https://doi.org/10.1080/03610928308828542>
- [13] Janic-Wróblewska A. Data-driven smooth tests for the extreme value distribution. *Statistics* 2004; 38(5): 413-426. <https://doi.org/10.1080/02331880410001692967>
- [14] Rayner J, Best D. Smooth tests of goodness of fit: an overview. *International Statistical Review/Revue Internationale de Statistique* 1990; pp. 9-17.
- [15] Rayner J, Best D. Neyman-type smooth tests for location-scale families. *Biometrika* 1986; pp. 437-446. <https://doi.org/10.1093/biomet/73.2.437>
- [16] Rayner J, Thas O, De Boeck B. A generalized Emerson recurrence relation. *Australian & New Zealand Journal of Statistics* 2008; 50(3): 235-240. <https://doi.org/10.1111/j.1467-842X.2008.00514.x>
- [17] Krit M. Goodness-of-fit tests in reliability: Weibull distribution and imperfect maintenance models. *Université de Grenoble* 2014.
- [18] Berheto TM, Haile DB, Mohammed S, *et al.* Predictors of loss to follow-up in patients living with HIV/AIDS after initiation of antiretroviral therapy. *North American Journal of Medical Sciences* 2014; 6(9): 453. <https://doi.org/10.4103/1947-2714.141636>
- [19] Sengayi M, Dwane N, Marinda E, Sipambo N, Fairlie L, Moultrie H. Predictors of loss to follow-up among children in the first and second years of antiretroviral treatment in Johannesburg, South Africa. *Glob Health Action* 2013; 6: 19248. <https://doi.org/10.3402/gha.v6i0.19248>
- [20] Kimber A. Tests for the exponential, Weibull and Gumbel distributions based on the stabilized probability plot. *Biometrika* 1985; pp. 661-663. <https://doi.org/10.1093/biomet/72.3.661>
- [21] Team RC. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria 2013; 2014.
- [22] Kallenberg WC, Ledwina T. Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *The Annals of Statistics* 1995; pp. 1594-1608. <https://doi.org/10.1214/aos/1176324315>
- [23] Lemeshko BY, Lemeshko S, Postovalov S. Comparative analysis of the power of goodness-of-fit tests for near competing hypotheses. I. The verification of simple hypotheses. *Journal of Applied and Industrial Mathematics* 2009; 3(4): 462. <https://doi.org/10.1134/S199047890904005X>
- [24] Sürücü B. A power comparison and simulation study of goodness-of-fit tests. *Computers & Mathematics with Applications* 2008; 56(6): 1617-1625. <https://doi.org/10.1016/j.camwa.2008.03.010>
- [25] Kang SI. Performance of generalized Neyman smooth goodness of fit tests 1978.
- [26] Megerso A, Garoma S, Tolosa Eticha TW, Daba S, Tarekegn M, Habtamu Z. Predictors of loss to follow-up in antiretroviral treatment for adult patients in the Oromia region, ethiopia. *HIV/AIDS (Auckland, NZ)* 2016; 8: 83. <https://doi.org/10.2147/hiv.s98137>
- [27] Wang B, Losina E, Stark R, Munro A, Walensky RP, Wilke M, *et al.* Loss to follow-up in a community clinic in South Africa: roles of gender, pregnancy and CD4 count. *SAMJ: South African Medical Journal* 2011; 101(4): 253-257. <https://doi.org/10.7196/SAMJ.4078>
- [28] Ramadhani HO, Thielman NM, Landman KZ, Ndosu EM, Gao F, Kirchherr JL, *et al.* Predictors of incomplete adherence, virologic failure, and antiviral drug resistance among HIV-infected adults receiving antiretroviral therapy in Tanzania. *Clinical Infectious Diseases* 2007; 45(11): 1492-1498. <https://doi.org/10.1086/522991>
- [29] Brinkhof M, Spycher B, Yiannoutsos C, Weigel ea R. Adjusting Mortality for Loss to Follow-up: Analysis of Five ART Programmes in Sub-Saharan Africa. *PLoS ONE* 2010; 5(11): e14149.
- [30] Wu J. Power and Sample Size for Randomized Phase III Survival Trials under the Weibull Model. *J Biopharm Stat* 2015; 25(1): 16-28. <https://doi.org/10.1080/10543406.2014.919940>

Received on 31-01-2017

Accepted on 16-02-2017

Published on 11-04-2017

<https://doi.org/10.6000/1929-6029.2017.06.02.2>© 2017 Odhiambo *et al.*; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.