International Journal of Preventive Medicine and Health (IJPMH)
ISSN: 2582-7588 (Online), Volume-5 Issue-5, July 2025

Survival Model of Cervical Cancer Patients using the 3-Parameter Weibull Distribution Model

Hassan Swedy Lunku, Ismail Juma Kaudunde, Kidney Chillingo

Abstract: ***This study aimed to evaluate a parametric survival model for cervical cancer patients treated with ORCI, and a case study was conducted to describe the model. The survey of survival times of cervical cancer patients may help reduce cervical cancer outcomes. Data on socio-demographic characteristics, reproductive status, stages, treatment, and follow-up of the treatment, abstracted from medical files, were considered in model development. The primary objective of this study was to analyse cervical cancer survival times from the diagnosis period using a three-parameter Weibull distribution model. The analysis was performed using the open-source statistical software R and Minitab. The three-parameter Weibull distribution is highly flexible for fitting random data; moreover, it exhibits strong adaptability to various types of probability distributions. When the three parameters are well chosen, it can be equal to or approximate some other statistical distribution. However, the three parameters were estimated to utilize the Weibull model successfully. The distribution of survival times of cervical cancer patients, as analysed, follows the three-parameter Weibull distribution, with required test statistics including the Anderson-Darling significant value and standard probability plots. The use of other parametric distribution models, such as the Gamma, three-parameter Gamma, and Weibull distributions, which encompass various types of hazard functions, is recommended for future studies.***

Key terms: 3-Parameter Weibull Distribution Model, Parametric Model, Survival Times, Cervical Cancer

Abbreviations:
ACS: American Cancer Society
LN: Lymph Node
MLE: Maximum Likelihood Estimation
AIC: Akaike Information Criterion
DFR: Decrease in Failure Rate
RSS: Residual Sum of Squares

I. INTRODUCTION

2017 WHO reports cancer as the leading cause of death in economically developed countries and the second globally, where cancer disease killed almost 9.6 million in 2018, approximately 70% of deaths from cancer occurred in low- and middle-income countries and globally, about 1 in 6 deaths is due to cancer, and is expected to increase to 11.5 million in 2030 (WHO, 2007 [21], 2013) [22]. Every year in developing countries, at least 7 million people die of cancer, more than HIV/AIDS, malaria, and tuberculosis combined. In developing countries, the burden of cancer increases due to the adoption of Western lifestyles and an increase in the number of older adults (WHO, 2008 [25], 2010 [23]). Poor cancer survival in developing countries is attributed to patients' late diagnosis, usually at an advanced or late stage, and limited access to timely and standard treatment. According to the GLOBOCAN (2012) [28], respectively, an estimated 14.1 million and 8.2 million of new cases of cancer and deaths due to cancer occurred in 2012 and the estimation of 5 years prevalent cases showed that there were 32.5 million alive who had diagnosed with cancer during the previous 5 years (Ferlay at al, 2014; Swaminathan and Sankaranarayanan, 2011) [18].

Cervical cancer is a cancer that starts in the cervix, part of the woman's reproductive system, which is the lower part of the uterus (womb). The cervical cancer stages are categorised into four; stage I is when there is a small amount of tumor present that has not spread to a lymph node (LN), stage II when the cancer spreads beyond the cervix and uterus but not the pelvic wall or lower part of the vagina, stage III when the cancer grows to the lower part of the vagina and pelvic wall and stage IV which is the most advanced stage when cancer spread to bladder, rectum or other areas of the body (Plummer et al, 2016) [29]. Invasive cervical cancer, which consists of stage IB and stage IIA, is one of the most successfully treatable cancers when detected early through regular screening, and its treatments include chemo-radiotherapy given at the same time, while surgery was included for the late stage of the disease (Swaminathan et al, 2002) [17].

The most common cause of cervical cancer is human papillomavirus (HPV) infection, and risk factors such as many sexual partners and engaging in early sexual contact, having a first full-term pregnancy before age 17, long-term use of birth control drugs, smoking, and alcohol consumption (Kalbfleish and Prentice, 2002) [9]. Few women with HPV infection progress to a cancer diagnosis, even though HPV infection is the primary cause of cervical cancer. 90% of women diagnosed with cervical cancer survive after the first year of diagnosis, according to the American Cancer Society (ACS).

In statistics, a parametric model refers to a family of distributions that can be described using a finite number of parameters, and these parameters are typically combined to form a single-dimensional parameter vector (Yang et al., 2019) [30]. These models incorporate various techniques for modelling and analysing different variables when the focus is

Manuscript received on 07 March 2025 | First Revised Manuscript received on 18 April 2025 | Second Revised Manuscript received on 19 June 2025 | Manuscript Accepted on 15 July 2025 | Manuscript published on 30 July 2025.

*Correspondence Author(s)

Hassan Swedy Lunku*, Assistant Lecturer, Department of Local Government Accounting and Finance, Local Government Training Institute (LGTI), Dodoma, Tanzania. Email ID: mtakwimu88@gmail.com, ORCID ID: 0000-0001-5668-8551

Ismail Juma Kaudunde, Lecturer, Department of Local Government Accounting and Finance, Local Government Training Institute (LGTI), Dodoma, Tanzania. Email ID: ismail.juma@yahoo.com

Kidney Chillingo, Department of Local Government Accounting and Finance, Local Government Training Institute (LGTI), Dodoma, Tanzania. Email ID: chillingokjc@gmail.com

on the relationship between the covariates and the survival times of the available survival data, considering the distribution (Harrel, 2001) [6]. In this study, interval censoring was employed to estimate the survival and failure rate using a 3-parameter Weibull distribution model, and the estimation method used was Maximum Likelihood Estimation (MLE). Parametric models are of interest to many statistical researchers due to their flexibility and variety in function and performance; widely used survival analysis models such as non-parametric model (Kaplan-Meier), semi-parametric models (Cox (1972, 1975) [4] proportional hazard model) and parametric models (Exponential, Weibull, Lognormal, Gamma) are often reviewed in this regard to estimate the survival of cervical cancer patients and prognostic factors (Lee and Wang, 2003 [11]; Schafer, 2002 [14]; Walter and Samuel, 2004) [20].

II. METHODS AND MODELS

A. Study Design, Sampling Procedure, and Sample Size

Retrospective review of female patients diagnosed with proven cervical cancer treated at ORCI and the available information abstracted using Kobo toolbox, from medical files from January 2014 to December 2015, and followed up to December 2017. The sample size of 161 cervical cancer patients was estimated by using the following criteria: test survival rate of 70%, anticipated survival rate of 80%, 5% significance level, 95% confidence level, and 90% power of the test using the following formulae for two-tailed $n=\left\{\frac{z_{1-\alpha/2}\sqrt{P_0(1-P_0)}+z_{1-\beta}\sqrt{P_a(1-P_a)}}{(P_0-P_a)}\right\}^2$ (Lwanga and Lemeshow, 1991) [13].

B. Study Models

i. Survival Functions and Estimation Method: The Weibull Function

A more complex but often more realistic model for survival is given by the Weibull function by the formulae: S (t) = exp (-λt$^{\gamma}$), where survival time, t ≥ 0, scale and shape parameters are estimated, denoted as λ > 0, γ > 0 (Weibull, 1951 [27]; Woolson, 1981 [24]; Walter and Samuel, 2004). The exponential survival function is a special case of the Weibull distribution with a parameter equal to 1, γ = 1. The hazard function is given by h(t) = λγt^ (γ-1). It increases as t increases if γ > 1 and decreases as t increases if 0 < γ < 1, and thus the different values of γ will reveal the nature of the distribution (Collet, 2003) [2]. The non-zero shift (Threshold) Weibull distribution has three parameters, and its probability density function is given by

$$f(x \setminus \gamma,\beta,\alpha) = \gamma\beta^{-\gamma}(x-\alpha)^{\gamma-1}\exp\left(-\left(\frac{x-\alpha}{\beta}\right)^{\gamma}\right)$$

Where β>0 is the scale parameter, α is the shift parameter, which is also a lower bound, and γ>0 is the shape parameter responsible for the skewness of the distribution. Basic statistics of the distribution are given as the Mean of the distribution, $E(x) = \alpha + \beta\Gamma\left(1+\frac{1}{\gamma}\right)$

Variance of the distribution, $V(x) = \beta^2\left(\Gamma\left(1+\frac{2}{\gamma}\right) - [\Gamma(1+1/\gamma]^2\right)$

Fisher skew, $Sk(x) = \frac{[\Gamma(1+1/\gamma]^3 - 3\Gamma\left(1+\frac{1}{\gamma}\right)\Gamma\left(1+\frac{2}{\gamma}\right) + \Gamma(1+\frac{3}{\gamma})}{\left(\Gamma\left(1+\frac{2}{\gamma}\right) - [\Gamma(1+1/\gamma]^2\right)^{3/2}}$

Where Γ represents the Gamma function, its shape varies from hyper exponential when γ<1 to nearly symmetrical when $\gamma \approx 3.6$ and when $\gamma \to \infty$ to a negatively skewed distribution. The best method to estimate the parameters of 3-parameter Weibull distribution is Maximum Likelihood Estimation (MLE) method, however, its application is problematic to Weibull distribution due to the following reasons: the distribution must satisfy the regularity conditions and Weibull distribution does not due to the domain of the random variable depends on the lower position of the lower bound, MLE solutions are biased and it is not known by what amount, and MLE solutions are not available in a direct form for two of three parameters (Yang et al, 2019). The likelihood function of the 3-parameter Weibull distribution for a sample of size n is given by

$$L(\gamma,\beta,\alpha/X) = \prod_{i=1}^{n} f(x_i/\gamma,\beta,\alpha)$$

$$= \prod_{i=1}^{n} \gamma\beta^{-\gamma}(x_i-\alpha)^{\gamma-1}\exp\left(-\left(\frac{x_i-\alpha}{\beta}\right)^{\gamma}\right)$$

Log-likelihood of the function simplified to

$$\log\left(L(\gamma,\beta,\alpha/X)\right) = -n\gamma\log(\beta) + n\log(\gamma) + (\gamma-1)\sum_{i=1}^{n}\log(x_i-\alpha) - \beta^{-\gamma}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}$$

Maximizing the equation by computing the derivative concerning the scale parameter β equaling zero, and the parameter estimated given by

$$\beta = \sqrt[\gamma]{\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}$$

And the derivative of the log-likelihood function concerning γ, replace $\beta^{-\gamma}$ with $1/\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}$ and dividing by n yields

$$\frac{1}{\gamma} + \frac{1}{n}\sum_{i=1}^{n}\log(x_i-\alpha) - \frac{\sum_{i=1}^{n}\log(x_i-\alpha)^{1+\gamma}}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}} = 0$$

Similarly, for α, we get

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i-\alpha}x\frac{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}} - \frac{\gamma}{\gamma-1} = 0$$

One way to estimate γ and α is to square the equations and search for the minimum, so that the complete MLE solution is given by $\{\hat{\gamma},\hat{\alpha}\}_{MLE}$ which satisfies the following two constraints

$$\underset{\gamma,\alpha}{Min}\left(\frac{1}{\gamma}+\frac{1}{n}\sum_{i=1}^{n}\log(x_i-\alpha)-\frac{\sum_{i=1}^{n}\log(x_i-\alpha)^{1+\gamma}}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}\right)^2$$

$$\underset{\gamma,\alpha}{Min}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i-\alpha}x\frac{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}{\sum_{i=1}^{n}(x_i-\alpha)^{\gamma-1}}-\frac{\gamma}{\gamma-1}\right)^2$$

$$\text{And by } \hat{\beta}_{MLE}=\sqrt[\gamma]{\frac{1}{n}\sum_{i=1}^{n}(x_i-\alpha)^{\gamma}}$$

ii. Weibull Regression Model

Models assume that the patient's survival time has a continuous probability Weibull distribution with probability distribution function given by the following formulae.

$$f(t/z)=\frac{\alpha}{\exp(\beta z)}\left(\frac{t}{\exp(\beta z)}\right)^{\alpha-1}\exp\left(-\frac{t}{\exp(\beta z)}\right)^{\alpha},\ t>0,\ \alpha>0$$

The following formulae can express the hazard function of the Weibull regression model.

$$h(t/z)=\frac{\alpha}{\exp(\beta z)}\left(\frac{t}{\exp(\beta z)}\right)^{\alpha-1}$$

With the survival function given by the following formula

$$S(t/z)=\exp\left(-\frac{t}{\exp(\beta z)}\right)^{\alpha}$$

The estimation of the parameter with the maximum likelihood function of the following formulae

$$L(\beta,t,z)=\prod_{i=1}^{n}\left\{\frac{\alpha t^{\alpha-1}}{[\exp(\beta z)]^{\alpha}}\exp\left(\frac{t}{\exp(\beta z)}\right)^{\alpha}\right\}\left\{\exp\left(\frac{t}{\exp(\beta z)}\right)^{\alpha}\right\}$$

The log of the Weibull hazard is a linear function of log time with constant (p log λ + log p) and slope (p – 1). The risk is thus rising if p > 1, constant if p = 1, follows the exponential, and declining if p < 1 to produce a bathtub curve (Anderen et al, 1993 [1]; Cox, 1972, 1975 [3]; Lee and Wang, 2003). The Weibull is also related to the extreme-value distribution, T ~W (λ; p) if and only if Y = log-T = α + σW; where W has the extreme value distribution, α = -log λ, and the surviving probability is given as, p = 1/σ where σ is the estimated parameter of the distribution (Theaune and Schoenfeld, 1982 [15]; Sharma, 1996 [16]; Gambach, 2001) [19].

iii. Akaike Information Criterion (AIC)

AIC is a measure of selecting a model from a set of different models, where smaller AICS indicate a better fit of the model, and estimates the quality of each model relative to each of the other models (Zhou, 2000 [26]; Leung et al, 1997 [12]; Harrel, 2001). AIC is given by -2log(likelihood) + 2k, where k is the number of parameters in the model, and for this study, k=3. AIC can also be calculated using residual sum of squares (RSS) from the regression AIC = nlog(nRSS) + 2k, where n is the number of observations (Everitt and Horton, 2005 [5]; Hosmer and Lemeshow, 1999) [7]; Izerman and Tran, 1990) [8].

iv. Log-Likelihood Value

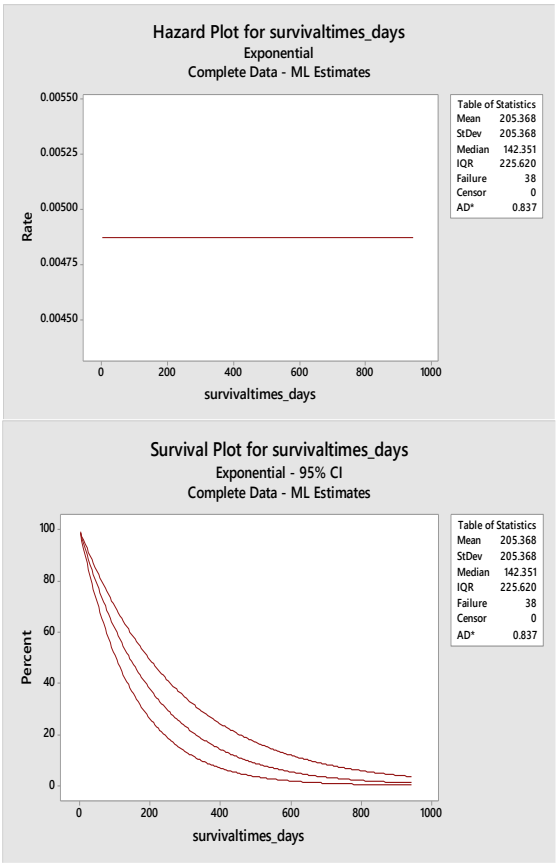
The estimated parameters of the three parametric models will be calculated by using maximum likelihood functions, and the selection of the best fit depends on the likelihood values of the observed data under the three parametric models (Kardaum, 1993) [10]. The function given by log L (θ; y) $=\sum \log f_i(y_i;\theta)$ $n_i=1$. The model that yields the highest likelihood value will provide the best fit.

III. RESULTS

A. Parametric Distributions

i. Exponential Function

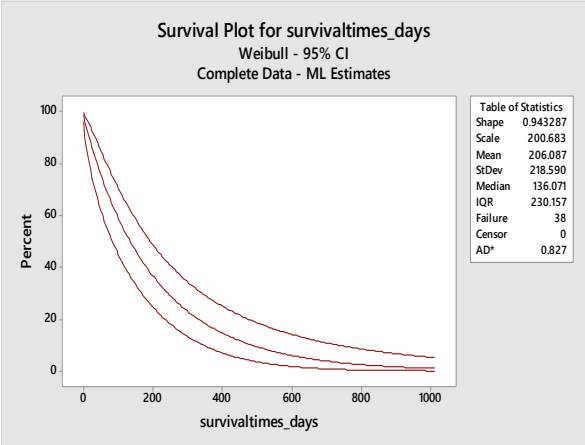
The simplest function was used to describe survival with one parameter; therefore, the approximated hazard rate remains a constant of 0.0048, with mean and median survival times of 205.368 and 142.351 days, respectively, as shown in Figure 1. The estimated mean parameter is 205.368 days (SE = 33.3152, and 95% CI of 149.434 and 282.239, respectively). Log-Likelihood = -240.343 and Anderson-Darling (adjusted) Goodness-of-Fit of 0.837.

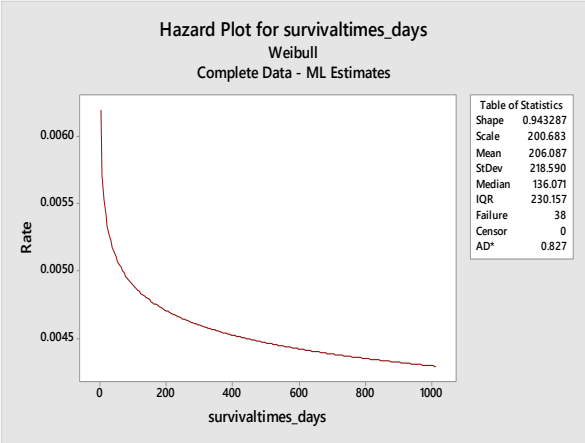


[Fig.1: Exponential Hazard and Survival Function]

ii. Weibull Function

More realistic but complex model for survival was given by Weibull where there was a gradual decrease in failure rate (DFR) as days of survival time increased with mean survival time of 206.08 days, median survival time 136.071 days and respective estimated shape and scale parameters of 0.9433±0.1267 and 200.683±36.0983 (95% lower limit of 0.7249 and 141.059, upper limit 1.2274 and 285.511). Log-Likelihood = -240.246 and Anderson-Darling Goodness-of-fit 0.827. As shown in Figure 2, the hazard rate decreases and remains constant at a specific rate.

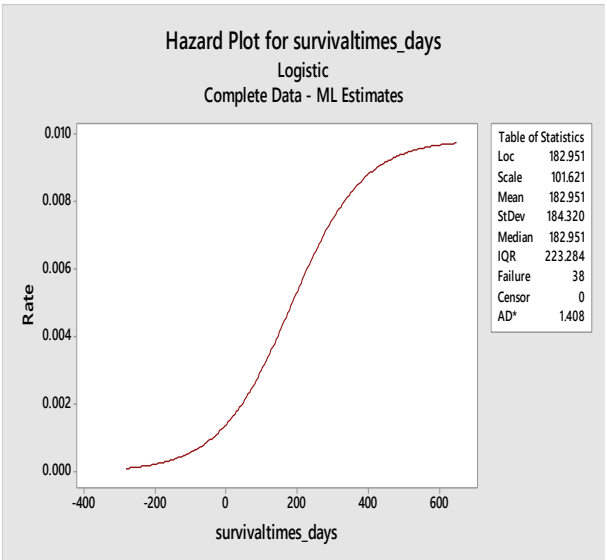


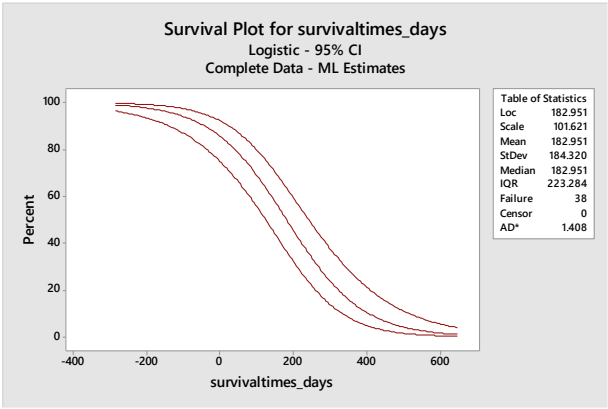


[Fig.2: Weibull Hazard Function and Survival Plot]

iii. Logistic Function

The logistic distribution is among the class of parametric survival models where the hazard rate increases; as a life testing model, it has its standing as an increasing failure rate model. The two-parameter logistic function was fitted with estimated location and scale parameters of 182.951 and 101.621, respectively (the respective standard errors are 28.7899 and 13.7572, with 95% lower limits of 126.524 and 77.9378, and upper limits of 239.379 and 132.5). The mean and median survival times were 182.951 days with a standard deviation of the survival time of 184.320 days. As shown in Figure 3, the hazard rate increases and remains constant when the rate reaches above 0.008.





[Fig.3: Logistic Hazard and Survival Plot]

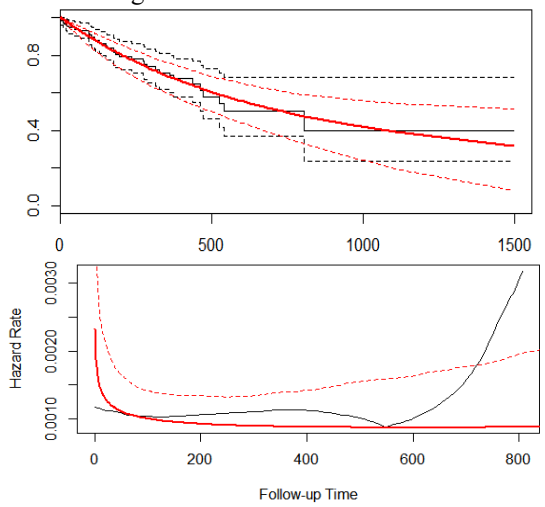
iv. Gamma Distribution and Generalized Gamma Function

The gamma distribution with parameters λ and k, denoted T(λ; k), has density: $f(t) = \frac{\lambda(\lambda t)^{k-1}e^{-\lambda t}}{T(k)}$ and survivor function: S(t) = 1- $I_k(\lambda t)$ and $I_k(x)$ is the incomplete gamma function, defined as: $Ik(x) = \int_0^x \frac{\lambda^{k-1}e^{-x}dx}{\Gamma(k)}$. Table 1 presents the results related to this function, estimating the shape parameter k = 0.8261 and the rate parameter as 0.0007. There is no explicit formula for the hazard, but it can be computed easily as the ratio of the density to the survivor function: h(t) = f(t) /S (t).

Table-I: Gamma Distribution Estimate

	Estimate	95% Lower	95% Upper	Standard Error
Shape	0.826141	0.609608	1.119587	0.128117
Rate	0.000701	0.000337	0.001457	0.000262

The gamma hazard increases monotonically if k > 1, from a value of 0 at the origin to a maximum of λ, λ is constant if k = 1 converts to an exponential distribution, decreases monotonically if k < 1, from ∞ at the origin to an asymptotic value of λ as shown in the Figure 4 follows a bath tub curve. Log-likelihood and AIC were respectively -298.8456 and 601.6911 at 2 degrees of freedom.



[Fig.4: Gamma Distribution, Survival, and Hazard Rate]

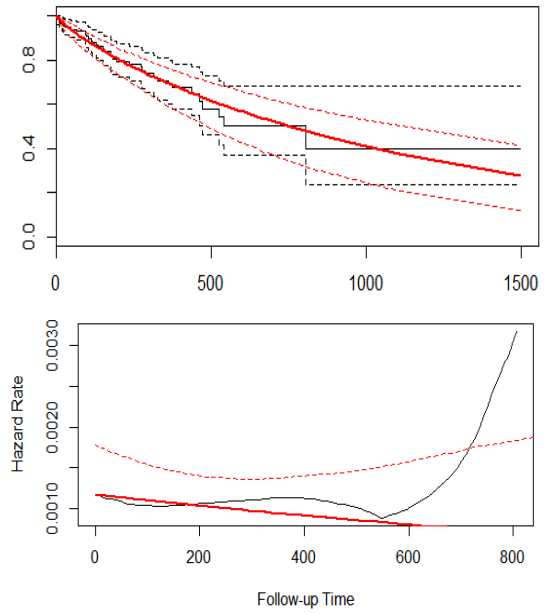
As introduced by Stacy and depicted in Figure 5, the generalised gamma distribution presents a flexible family with varying shapes and hazard functions, which are

often suitable for modelling survival data. It is a three-parameter distribution with respective estimated scale, shape, and location values presented as μ, σ, and Q, with values of 7.0848, 0.9826, and 1.2763, respectively, as shown in Table 2. The log-likelihood and AIC were -298.8052 and 603.6105, respectively, at 3 degrees of freedom.

$$f(t;\theta) = \frac{k\,(k^{-2})^{(k^{-2})}(\lambda t)^{(k^{-2})}\frac{k}{\sigma}\exp\left[-k^{-2}(\lambda t)^{\frac{k}{\sigma}}\right]}{[\Gamma(k^{-2})\sigma t]}$$

Table-II: Generalized Gamma Distribution Estimates

	Estimates	95% Lower	95% Upper	Standard Error
Mu	7.0848	6.5935	7.5761	0.2506
Sigma	0.9826	0.4266	2.2631	0.4183
Q	1.2763	-0.0568	2.6093	0.6802



[Fig.5: Generalized Gamma Distribution, Survival, and Hazard]

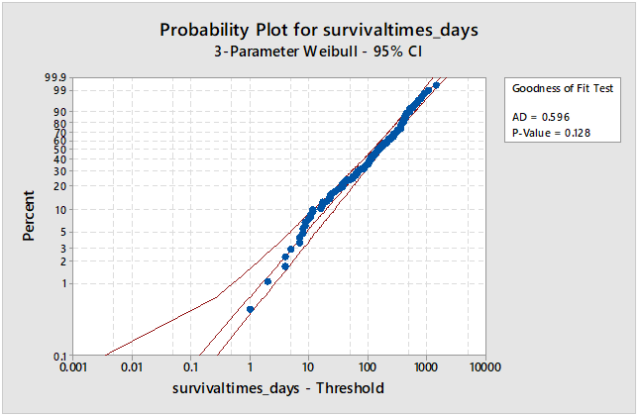
B. Goodness-of-fit Test

The hypotheses for the Anderson-Darling test were: H_0: The data follow a specified distribution VS, H_1: The data do not follow a specified distribution. From the Goodness-of-fit, presented in Table 3, the 3-parameter Gamma distribution had a small adjusted Anderson-Darling value of 0.504, while the 3-parameter Weibull had a small AD value of 0.596 with a p-value of 0.128; hence, the data fit well.

Table-III: Goodness-of-fit Test

Distribution	Anderson-Darling	p-value
Normal	6.363	<0.005
3-parameter lognormal	1.865	*
2-parameter Exponential	0.895	0.147
3-parameter Weibull	0.596	0.128
Smallest Extreme value	15.634	<0.010
Largest extreme value	3.108	<0.010
3-parameter Gamma	0.504	*
Logistic	4.581	<0.005
3-parameter Loglogistic	2.322	*

Probability plots are a great way to visually identify the distribution that survival data follow. If the data points follow a straight line, the distribution is considered to be a good fit. Figure 6 shows a 3-parameter Weibull distribution in the probability plot for survival times. The distribution is a good fit for the data among the distributions, as the points fall closely along the fitted distribution line and the confidence bound lines.



[Fig.6: Three-Parameter Weibull Probability Plot]

Since the probability plot, Table 4 presents ML estimates of the distribution parameters, AD and p-value indicate a 3-parameter, Weibull is a good fit of the survival time of cervical cancer with shape parameter 0.93695, scale 225.23005, and threshold of -1.02393 as shown by the maximum likelihood estimates of the distribution parameter table below. For 3-parameter distributions, only a low value indicates that adding the third parameter is a significant improvement over the 2-parameter version.

Table-IV: ML Estimates of Distribution Parameters

Distribution	Location	Shape	Scale	Threshold
Normal*	230.77640		234.65877	
3-parameter Lognormal	4.98998		1.10669	-12.63121
2-parameter Exponential			232.21869	-1.44235
3-parameter Weibull		0.93695	225.23005	-1.02393
Smallest extreme value	364.61500		352.72600	
Largest extreme value	134.35442		249.40483	
3-parameter Gamma		0.87858	263.53913	-0.76506
Logistic	198.63323		120.10248	
3-parameter loglogistic	3.94217		0.76048	-2.05067

C. Three-Parameter Weibull

The Weibull distribution is characterised by its shape, scale, and threshold parameters, and is also referred to as the 3-parameter Weibull distribution. A 3-parameter Weibull distribution can work with zeros and negative data, but all data for a 2-parameter Weibull distribution must be greater than zero. From Table 5, the probability function of the 3-parameter Weibull is given as; $f(T;k,\lambda,\theta) = \frac{k}{\lambda}\left(\frac{T-\theta}{\lambda}\right)^{k-1}e^{-\left(\frac{T-\theta}{\lambda}\right)^{k}}$, where for the fitted distribution the shape parameter, k=0.920705, scale parameter, λ = 1085.09 and the threshold parameter, θ= -3.52284.

Table-V: Three-Parameter Weibull Distribution Parameter Estimates

Parameter	Estimate	Standard error	95% Normal Lower CI	95% Normal Upper CI
Shape	0.920705	0.119488	0.713925	1.18738
Scale	1085.09	250.384	690.319	1705.61
Threshold	-3.52284	0	-3.52284	-3.52284

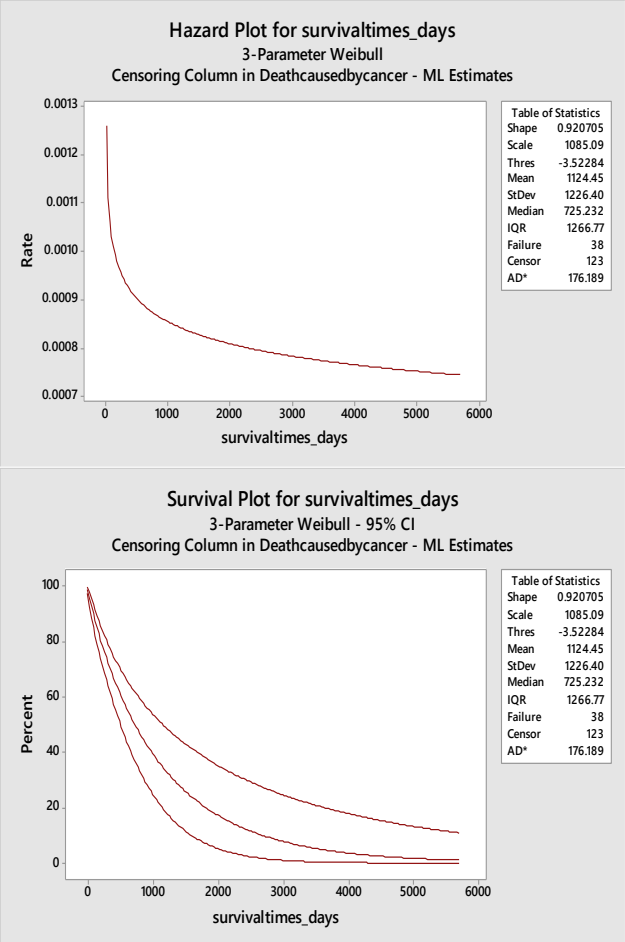
Log-Likelihood = -300.005

As shown in Table 6, the estimated mean time to failure (MTTF) of the 3-parameter distribution is higher, 1124.45 days, compared to the other distributions fitted for the cervical cancer survival data, also with a higher median survival time of 725.232 days compared to the different distributions

Table-VI: Characteristics of Distribution

	Estimate	Standard Error	95% Lower	95% Upper
Mean (MTTF)	1124.45	314.613	507.820	1741.08
Standard deviation	1226.40	479.876	569.597	2640.57
Median	725.232	146.693	437.718	1012.75
First Quartile	276.876	49.9647	178.847	374.805
Third quartile	1543.65	406.661	746.605	2340.69
Interquartile range	1266.77	381.364	702.168	2285.35

As shown in the Figure 7, the hazard plot and survival plot for survival time for 3-parameter Weibull, there was a steep decrease failure rate as survival time increase before 1000 days with hazard rate greater than 0.0012 and then gradual decrease failure rate above 1000 days, which means that the patients had lower survival time.



[Fig.7: Three-Parameter Weibull Hazard and Survival Plot]

Table-VII: Survival Time Characteristics of Different Distributions

	Mean	Median	S.E.	95% confidence interva	
				Lower bound	Upper bound
Exponent	205.368	142.351	33.3152	149.434	282.239
Weibull	206.080	136.071	36.0983	141.059	285.511
Logistic	182.951	184.320	28.7899	126.524	239.379
Three-parameter Weibull	1124.45	725.232	314.613	507.82	1741.08

D. Parametric Regression Models

The logistic regression obtained for available data for cervical cancer patients with minimum log-likelihood and AIC, where AIC=-2loglik (MLE) + 2p when W~N (0,1), W~logistic, and W~Extreme values. The table below presents the coefficients of the AFT logistic regression model, which was found to be significantly greater than those of other models. The M.L.E. of the scale parameter was 176.243, the intercept M.L.E. of 589.560, and the estimated log of the scale parameter [Log (scale)] was 5.381. As shown in Table 8, the patient's cancer stage was more significant than the other covariates in the model; thus, patients with the latter stages, i.e., stages III and IV, had lower survival rates than those with the other covariates in the study.

Table-VIII: AFT Logistic Regression Model Coefficients

Covariates	Coefficients	Std. error	Z	P
Intercept	475.64698	354.030	1.34	0.17910
Stage II	-11.42700	166.356	-0.07	0.94524
Stage III	-417.10416	156.896	-2.66	0.00785
Stage IV	-609.11332	172.926	-3.52	0.00043
Menopause status category	127.54061	81.062	1.57	0.11563
Multiple sexual partners	35.41634	89.929	0.39	0.69371
Smoking history	59.32155	139.022	0.43	0.66959
Cancer grade	-10.31298	26.547	-0.39	0.69766
Log(scale)	5.172	0.129	40.08	<2e-16

Loglik (model) =-297.1, Loglik (intercept) =-316.6, Chisq = 39.05 with 7 degrees of freedom, p=1.9e-06

IV. DISCUSSION

The available data fit well with the three-parameter Weibull distribution, and the patient's surviving probability decreases significantly. The survival times for non-parametric and semi-parametric approaches were similar to those of the available data on non-parametric, semi-parametric, and parametric models. In contrast, the parametric approach had a higher mean survival time. The findings of the study showed that the survival of patients was poor and patients with the latter cancer stage had an increased risk of death compared to those with earlier cancer stage on survival probability of patients with cervical cancer was 0.194, 0.166, 0.0973, and 0.0387 for stage I, II, III and IV. The chance of survival for stage IV was lower compared to the other stages, I, II, and III. The cancer stage had a significant impact on patient survival in the fitted model, surpassing the effect of other covariates. Similarly, the cancer stage was found to have a greater impact on patient survival, as indicated by the fitted model,

compared to the other covariates. The study provides essential information for public health decisions and policymakers, as well as estimates the patient survival probability.

V. CONCLUSION

Estimating survival functions for different diseases has interested statisticians for several years, and since the survival function gives information on the probability of a time-to-event of interest, which was death caused by cervical cancer for this study. Researchers and biostatisticians prefer semi-parametric models under certain conditions. The parametric models provide more precise estimates due to their specific conditions; however, they are invalid when the PH assumption does not hold or when the survival times of the available data do not follow the parametric distribution, as shown in this study. Detection of cervical cancer at early stages through regular screening programs for women and comprehensive treatment should be taken up to improve the overall survival of the patients. Improved awareness is necessary in controlling cervical cancer, and can be done by having health education and regular screening programs carried out to create awareness.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/Competing Interests:** Based on my understanding, this article does not have any conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted with objectivity and without any external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Andersen P.K., organ O., Gill R.D., Keiding N. (1993); *Statistical Models Based On Counting Processes*, Springer, New York. DOI: http://dx.doi.org/10.1007/978-1-4612-4348-9
2. Collet D. (2003); *Modelling survival data in medical research*, Chapman and Hall, London. DOI: https://doi.org/10.1201/b18041
3. Cox D.R. (1975); *Partial likelihood,* Biometrika, Vol. 62, No. 2, 269-276, DOI: https://doi.org/10.1093/biomet/62.2.269
4. Cox, D. R (1972): *Regression models and life tables*. Journal of the Royal Statistical Society: Series B, 34:187-220, DOI: https://doi.org/10.1111/j.2517-6161.1972.tb00899.x
5. Everitt B.S. and Hothorn T. (2005); A handbook of Statistical analyses using R, CRC Press, London. https://digitallibrary.tsu.ge/book/2019/september/books/A-Handbook-of-Statistical-Analyses.pdf
6. Harrel F.E., (2001); *Regression modeling strategies: With applications to linear models, logistic regression and survival analysis*, Springer, New York. DOI: https://doi.org/10.1007/978-1-4757-3462-1
7. Hosmer D.W and Lemeshow S (1999); *Applied survival analysis: Regression modeling to time to event data*, John Wiley & Sons Inc, New York. DOI: https://doi.org/10.1002/9780470258019
8. Izenman A.J. and Tran L.T. (1990); *Estimation of the survival function and hazard rate*, Journal of Statistical Planning and Inference, Vol 24, 233-247. DOI: https://doi.org/10.1016/0378-3758(90)90044-U
9. Kalbfleish & Prentice (2002); *The statistical analysis of failure time data*, 2nd edition, Wiley and Sons, New York. DOI: http://doi.org/10.1002/9781118032985
10. Kardaum, (1993); *Statistical analysis of male larynx cancer patients: A case study*. Statistical Nederlandica, 37:103-126. https://www.iosrjournals.org/iosr-jhss/papers/Vol.%2021%20Issue5/Version-1/E0215012234.pdf
11. Lee E. T. and Wang J. W. (2003); *Statistical Methods for Survival Data Analysis,* 3rd edition, Wiley & Sons Inc, New Jersey. DOI: http://doi.org/10.1002/0471458546
12. Leung, K.M., R.M. Elashoff and A.A. Afifi, (1997); *Censoring issues in survival analysis*, Annual Review of Public Health, 18:83-104, DOI: https://doi.org/10.1146/ANNUREV.PUBLHEALTH.18.1.83
13. Lwanga S.K. and Lemeshow S. (1991); *Sample size determination in health studies: A practical guide,* WHO manual. https://iris.who.int/handle/10665/40062
14. Schafer J. L. (2002); *Analysis of Incomplete Multivariate Data*, Chapman and Hall, London. https://www.taylorfrancis.com/books/mono/10.1201/9780367803025/analysis-incomplete-multivariate-data-schafer
15. Schoenfeld D. (1982), *Residuals for the proportional hazards regression model,* Biometrika, 69(1): 239-241, DOI: http://dx.doi.org/10.1093/biomet/69.1.239
16. Sharma S. (1996); *Applied multivariate techniques*, John Wiley & Sons Inc., Canada. https://www.wiley.com/en-us/Applied+Multivariate+Techniques+-p-9780471310648
17. Swaminathan R, et al (2002); *Effect of loss to follow-up on population-based cancer survival rates in developing countries,* International Journal of Cancer Suppl, 13:172. DOI: https://doi.org/10.2471/blt.07.046979
18. Swaminathan R, Lucas E and Sankaranarayanan R (2011); *Cancer survival in Africa, Asia, the Caribbean and Central America: Database and Attributes*, IARC Scientific Publications No. 162, International Agency for Research on Cancer, Lyon. https://pubmed.ncbi.nlm.nih.gov/21675403/
19. Theaune, T. and Gambach (2001); *Modelling survival data- Extending the Cox model*, Springer, London. https://catalog.nlm.nih.gov/discovery/fulldisplay?docid=alma9910960353406676&context=L&vid=01NLM_INST:01NLM_INST&lang=en&search_scope=MyInstitution&adaptor=Local%20Search%20Engine&tab=LibraryCatalog&query=lds56,contains,Models,%20Statistical,AND&mode=advanced&offset=110
20. Walter A. S. and Samuel S. W. (2004); *Weibull Models*, John Wiley and Sons, New York. https://www.wiley.com/en-us/Weibull+Models+-p-9780471360926
21. WHO (2007). The World Health Organisation's fight against cancer: Strategies that prevent, cure, and care. WHO Cancer Brochure. https://www.afro.who.int/sites/default/files/2017-06/WHOCancerBrochure2007.FINAL_.pdf
22. WHO*, Human Papillomavirus (HPV) and Cervical Cancer, s.l.: World Health Organization (WHO).* 2013. https://www.who.int/news-room/fact-sheets/detail/human-papilloma-virus-and-cancer
23. WHO/ICO*, HPV Information Centre on HPV and Cervical Cancer (HPV Information Centre),* 2010, WHO. https://hpvcentre.net/
24. Woolson, R.F., (1981); *Rank test and a one-sample log-rank test for comparing observed survival data to standard population*. Biostatistics, 37:687-696. DOI: http://doi.org/10.2307/2530150
25. World Health Organisation (WHO), 2008, *The Global Burden of Disease: 2004 Update*. Geneva: World Health Organization. https://www.who.int/publications/i/item/9789241563710
26. Zhou Mai (2000); *Using software R for survival analysis and simulation—a tutorial*, The American Statistician, Kentucky. https://www.ms.uky.edu/~mai/rsurv.pdf
27. Weibull, W (1951). *A statistical Distribution of wide applicability*, Journal of Applied Mechanics, 18, 293-297. DOI: https://doi.org/10.1115/1.4010337
28. GLOBOCAN (2012); *Cancer Incidence and Mortality Worldwide*, GLOBOCAN, Lyon, France: International Agency for Research on Cancer.

https://publications.iarc.fr/Databases/Iarc-Cancerbases/GLOBOCAN-2012-Estimated-Cancer-Incidence-Mortality-And-Prevalence-Worldwide-In-2012-V1.0-2012
29. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S. (2016): *Global burden of cancers attributable to infections in 2012: a synthetic analysis*. Lancet Global Health, London. DOI: https://doi.org/10.1016/s2214-109x(16)30143-7
30. Yang F, Ren H, and Hu Z (2019) Maximum Likelihood Estimation for Three-Parameter Weibull Distribution using Evolutionary Strategy, Mathematical Problems in Engineering. DOI: http://dx.doi.org/10.1155/2019/6281781

AUTHOR'S PROFILE

Hassan Swedy Lunku is an Assistant Lecturer in Statistics at the Local Government Training Institute (LGTI), with research interests in statistical modelling, data analysis, and public health performance. His academic work focuses on applying quantitative methods to enhance the health sector, improve service delivery, and inform policy evaluation. He is particularly passionate about using statistics to support evidence-based decision-making within Tanzania's public health systems.

Ismail Juma Kaudunde is a Lecturer in Accounting and Finance at the Local Government Training Institute (LGTI). His research spans public financial management and fiscal decentralisation. With a strong background in finance, He focuses on improving financial practices in the health sector through applied research and capacity-building initiatives.

Kidney Chillingo holds a Master of Science in Mathematics and brings a sharp analytical mind to his writing. With a strong foundation in logical reasoning and abstract problem-solving, he has a particular interest in exploring the intersection between mathematical theory and real-world application. His work often reflects a passion for clarity, precision, and the elegant simplicity of mathematical thought. When he's not immersed in equations, he enjoys mentoring students and engaging with projects that promote education.