Mikailalsys Journal of Mathematics and Statistics

e-ISSN : 3030-816X
p-ISSN : 3030-8399

Index : Harvard, Boston, Sydney University, Dimensions, Scilit, Semantic, Google, etc

https://doi.org/10.58578/MJMS.v4i3.10125

Novel Regression Model for Investigating the Survival Times of HIV Patients in Gombe State, Nigeria

U. F. Abbas[1] & O. D. Adubisi[2]

[1]Gombe State Polytechnic, Bajoga, Nigeria; [2]Federal University Wukari, Wukari, Nigeria
ufabbas@gspb.edu.ng; adubisiobinna@fuwukari.edu.ng

Submitted:	Revised:	Accepted:	Published:
Jun 15, 2026	Aug 30, 2026	Sep 11, 2026	Sep 16, 2026

Abstract

Conventional survival models often lack the flexibility required to capture the complex hazard structures observed in biomedical data. This study proposes a survival regression model based on the log-DUS Topp–Leone Burr–Hatke exponential (LDUSTLBHE) distribution for modeling censored survival times among patients with HIV in Gombe State, Nigeria. The model extends the Burr–Hatke exponential framework by incorporating additional shape flexibility to accommodate diverse survival patterns. Its probability density function, survival function, and regression structure were derived, and the parameters were estimated using maximum likelihood estimation under censoring. The model was applied to data from 200 patients with HIV, with age, gender, CD4 count, World Health Organization disease stage, and number of opportunistic infections included as explanatory variables. Its performance was compared with the log-Topp–Leone Burr–Hatke exponential, log-alpha-power Burr–Hatke exponential, and log-Burr–Hatke exponential regression models using the log-likelihood, Akaike information criterion, Bayesian information criterion, consistent Akaike information criterion, and Hannan–Quinn information criterion. The LDUSTLBHE model provided the best fit among the competing models, while diagnostic assessments based on Cox–Snell residuals and Kaplan–Meier survival comparisons further supported its adequacy. Advanced disease stage, a greater number of opportunistic infections, and older age were associated with shorter survival times. The

MJMS Journal is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

proposed model therefore offers a flexible framework for analyzing censored survival data and provides clinically relevant insights into HIV survival dynamics.

Keywords: Censored Survival Data; HIV; LDUSTLBHE Distribution; Nigeria; Survival Regression

INTRODUCTION

Human Immunodeficiency Virus (HIV) is a retrovirus that primarily infects CD4+ T lymphocytes, which play a critical role in coordinating the body's immune response. By progressively destroying these immune cells, HIV weakens the immune system and increases susceptibility to opportunistic infections and other life-threatening conditions. Since its discovery, the HIV epidemic has remained one of the most serious global public health challenges. According to recent global estimates, approximately 38 million people are currently living with HIV worldwide. Although substantial progress has been made in the development and widespread implementation of antiretroviral therapy (ART), which has transformed HIV infection from a fatal disease into a manageable chronic condition, significant disparities still exist in early diagnosis, treatment initiation, and long-term viral suppression, particularly in resource-limited settings (UNAIDS, 2023; NACA, 2022).

Sub-Saharan Africa continues to bear a disproportionate share of the global HIV burden. Structural and socio-economic challenges such as poverty, inadequate healthcare infrastructure, limited access to treatment, and delayed diagnosis contribute significantly to the continued spread and impact of the disease in the region (Sheehy et al., 2014; Onovo et al., 2023). Nigeria, in particular, remains one of the countries most affected by HIV in Africa and despite improvements in prevention and treatment programs, the country still experiences a large number of late-stage diagnoses, which often result in advanced immune suppression at the time patients seek medical care (NACA, 2022). Gombe State, located in northeastern Nigeria, reflects many of these challenges associated with HIV management. The effectiveness of HIV treatment and survival outcomes in the region is influenced by several contextual factors including social stigma, inadequate healthcare infrastructure, economic constraints, and limited access to early testing services (Adamu et al., 2018; Onovo et al., 2023). As a result, many patients present for treatment at advanced stages of

the disease when immune system damage has already progressed significantly. These conditions produce complex survival patterns that vary across individuals and over time.

Understanding the survival dynamics of HIV patients is therefore essential for improving clinical management and guiding public health interventions. Survival analysis provides an important statistical framework for analysing time-to-event data, particularly in the presence of censoring, which frequently occurs in medical studies. Classical survival models such as the exponential, Weibull, and Cox proportional hazards models have been widely applied in biomedical research (Kaplan and Meier, 1958; Cox, 1972; Kleinbaum and Klein, 2012). However, these traditional models often impose restrictive assumptions on the hazard rate function and may not adequately capture the complex and heterogeneous survival patterns associated with HIV disease progression. To address these limitations, recent research in statistical modeling has focused on the development of more flexible lifetime distributions and survival regression models capable of accommodating diverse hazard rate shapes and data characteristics (Mezher and Ibrahim, 2024; Ali and Abbas, 2025; Khikman et al., 2025; Li and Zang, 2025; Sadiq et al., 2025; Olanrewaju and Adeniran., 2026). Flexible parametric models constructed through generator families and distributional transformations have demonstrated improved performance in modeling survival data exhibiting skewness, heavy tails, and non-monotonic hazard functions (Abdulkadir et al., 2024; Ishag et al., 2024; Lima et al., 2025; Isa et al., 2025; Yusuf et al., 2025; Liu et al., 2026). These models provide greater adaptability and have been successfully applied in several areas including reliability analysis, engineering studies, and biomedical research.

Despite these methodological advances, relatively few studies have applied flexible distribution-based regression models to HIV survival data in Nigeria. Existing studies in the region have largely focused on epidemiological trends, prevalence analysis, and basic survival estimation. For instance, Udofia et al. (2021), Onovo et al. (2023), Anyiam et al. (2024) and Falgore et al. (2024) examined HIV survival patterns with emphasis on CD4 cell count progression and general survival outcomes. Although these studies provide valuable insights into HIV epidemiology, they primarily rely on conventional survival analysis techniques that may not fully capture the complex dynamics of HIV progression and immune suppression. Consequently, the absence of flexible statistical models capable of accommodating the intricate survival patterns observed among HIV patients remains a significant limitation in the existing literature. Advanced regression models based on

flexible probability distributions have the potential to address this gap by allowing for more realistic modeling of hazard rate behaviour and heterogeneous survival experiences among patients.

Motivated by these limitations, the present study proposes a novel statistical regression model that integrates advanced probability distributions within a regression framework for analysing censored survival data. The proposed model aims to provide improved flexibility in capturing the survival dynamics of HIV patients. Using HIV patient data from Gombe State, Nigeria, this study evaluates the performance of the proposed model and compares it with existing competing models in order to provide deeper insights into the factors influencing survival outcomes among individuals living with HIV in the region.

METHODS

The Log DUSTopp-Leone Burr-Hatke Exponential model ($LDUSTL_{BHE}$)

The DUSTopp-Leone Burr Hatke exponential distribution introduced by Abbas and Adubisi (2026) has the cumulative distribution function (cdf) and probability density function (pdf) specified as

$$F(x;\gamma,\beta)=\frac{e^{\left[1-\left(\frac{e^{-\beta x}}{1+\beta x}\right)^2\right]^{\gamma}}-1}{e-1},\ x>0,\ \gamma,\beta>0 \tag{1}$$

and

$$f(x;\gamma,\beta)=\frac{2\gamma\beta(2+\beta x)e^{-2\beta x}}{(e-1)(1+\beta x)^3}\left[1-\left(\frac{e^{-\beta x}}{1+\beta x}\right)^2\right]^{\gamma-1} e^{\left[1-\left(\frac{e^{-\beta x}}{1+\beta x}\right)^2\right]^{\gamma}} \tag{2}$$

where γ is the shape parameter and , β is the scale parameter. Let $Y=\log X$, $\beta=e^{-\frac{\mu}{\sigma}}$, then the random variable Y follows the $LDUSTL_{BHE}$ parameters (γ,μ,σ). Given that $Y=\log X$, then $x=e^{y}$ and $\frac{dy}{dx}=\frac{1}{x}\Rightarrow\frac{dx}{dy}=x$.

The density function of Y is derived using the transformation expressed as

$$f(y;\mu,\sigma,\gamma)=f(x;\mu,\sigma,\gamma)\left|\frac{dx}{dy}\right| \tag{3}$$

$$= \frac{2\gamma e^{-\frac{\mu}{\sigma}e^{y}}\left(2+e^{-\frac{\mu}{\sigma}e^{y}}\right)e^{-2e^{-\frac{\mu}{\sigma}e^{y}}}}{\sigma(e-1)\left(1+e^{-\frac{\mu}{\sigma}e^{y}}\right)^{3}}\left[1-\left(\frac{e^{-e^{-\frac{\mu}{\sigma}e^{y}}}}{1+e^{-\frac{\mu}{\sigma}e^{y}}}\right)^{2}\right]^{\gamma-1} e^{\left[1-\left(\frac{e^{-e^{-\frac{\mu}{\sigma}e^{y}}}}{1+e^{-\frac{\mu}{\sigma}e^{y}}}\right)^{2}\right]^{\gamma}} \quad (4)$$

Hence, the transformed density function is specified as

$$\mathrm{F}(y;\mu,\sigma,\gamma) = \frac{2\gamma e^{\frac{y-\mu}{\sigma}}\left(2+e^{\frac{y-\mu}{\sigma}}\right)e^{-2e^{\frac{y-\mu}{\sigma}}}}{\sigma(e-1)\left(1+e^{\frac{y-\mu}{\sigma}}\right)^{3}}\left[1-\left(\frac{e^{-e^{\frac{y-\mu}{\sigma}}}}{1+e^{\frac{y-\mu}{\sigma}}}\right)^{2}\right]^{\gamma-1} e^{\left[1-\left(\frac{e^{-e^{\frac{y-\mu}{\sigma}}}}{1+e^{\frac{y-\mu}{\sigma}}}\right)^{2}\right]^{\gamma}} \quad (5)$$

where $\gamma, \sigma > 0$ and $y, \mu \in \mathbb{R}$. The survival function to the preceding function is specified as

$$\mathrm{S}(y;\mu,\sigma,\gamma) = 1 - \left\{\frac{e^{\left[1-\left(\frac{e^{-e^{\frac{y-\mu}{\sigma}}}}{1+e^{\frac{y-\mu}{\sigma}}}\right)^{2}\right]^{\gamma}}-1}{e-1}\right\} \quad (6)$$

The standardized random variable Z is defined as $z = \frac{y-\mu}{\sigma}$. Hence, Equations (5) and (6) can be expressed as

$$\mathrm{F}(z;\mu,\sigma,\gamma) = \frac{2\gamma e^{z}(2+e^{z})e^{-2e^{z}}}{\sigma(e-1)(1+e^{z})^{3}}\left[1-\left(\frac{e^{-e^{z}}}{1+e^{z}}\right)^{2}\right]^{\gamma-1} e^{\left[1-\left(\frac{e^{-e^{z}}}{1+e^{z}}\right)^{2}\right]^{\gamma}} \quad (7)$$

$$\mathrm{S}(z;\mu,\sigma,\gamma) = 1 - \left\{\frac{e^{\left[1-\left(\frac{e^{-e^{z}}}{1+e^{z}}\right)^{2}\right]^{\gamma}}-1}{e-1}\right\} \quad (8)$$

The regression model which connects the i[th] response variable y_i with the independent variables $K_{i1}, K_{i2}, \ldots, K_{im,}$ is specified as

$$y_i = \sum_{j=1}^{m} K_{ij}\xi_j + \sigma z_i, \quad i = 1,2,\ldots,n. \quad (9)$$

where $\mu_i = \sum_{j=1}^{m} K_{ij}\xi_j$, ξ_j, $j = (1,2\ldots,m)$ is an $m \times 1$ parameter vector associated with the independent variables and Z_i is the random error term which follows Equation (7) with $\mu = 0$ and $\sigma = 1$. Additionally, Equation (9) can be utilized in modeling situations were the response variable significantly affected by the independent variables.

Figure 1 displays the density and survival functions of the LDUSTL$_{BHE}$ distribution for selected parameter values. The plots demonstrate the high degree of flexibility offered by the proposed distribution. The density function is capable of exhibiting different shapes depending on the parameter values, including right-skewed and unimodal forms commonly observed in survival data. Similarly, the survival function exhibits varying rates of decline, indicating that the distribution can effectively model different hazard behaviors. This flexibility is particularly important in survival modeling, where hazard functions may exhibit non-monotonic or complex patterns due to disease progression or treatment effects.

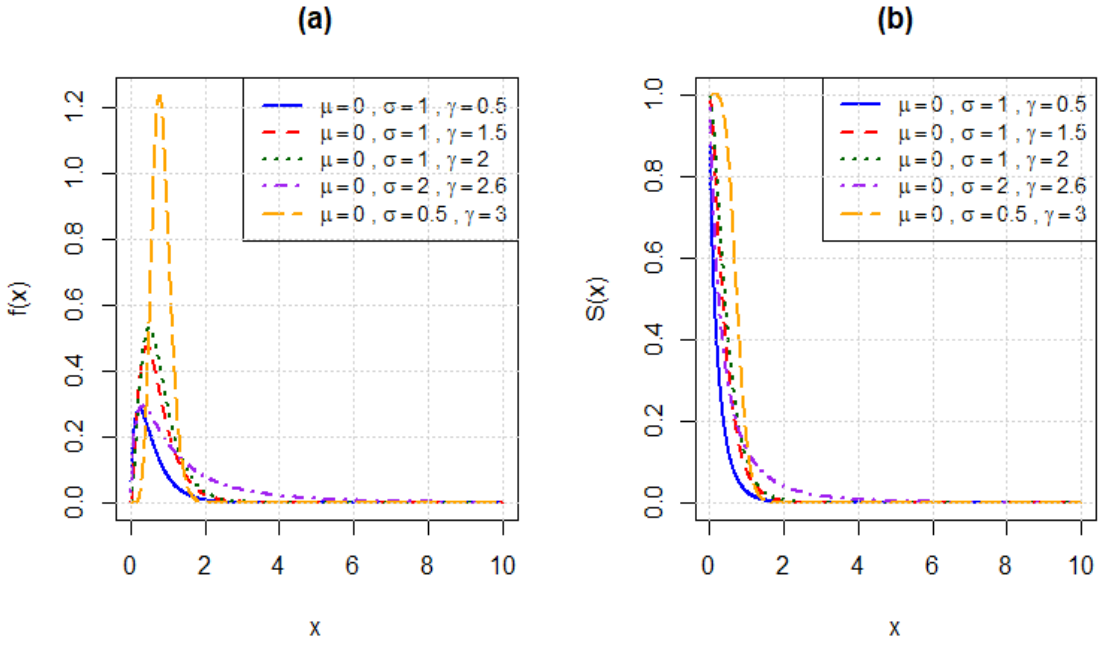


Figure 1. The density and survival functions plots of the LDUSTL$_{BHE}$.

Inference for the LDUSTL$_{BHE}$

Given that $(y_1, \Delta_1, K_1), (y_2, \Delta_2, K_2), \ldots, (y_n, \Delta_n, K_n)$ is random sample of size (n) with

$$y_i = min[\log(T_i), \log(C_i)] \tag{10}$$

where Δ_i is the censoring indicator (0 = censored, 1 = failure) and K_i denote the i^{th} individual. Assuming the observed lifespan and censoring time are independent with no informative censoring. The loglikelihood, say ℓ with $\psi = (\eta, \sigma, \xi^T)^T$ is specified as

$$\ell(\psi; y) = \sum_{i\epsilon F} \log[f(y;\psi)] + \sum_{i\epsilon C} \log[S(y;\psi)] \tag{11}$$

where $f(y;\psi)$ is the density function in Equation (5), $S(y;\psi)$ is the survival function in Equation (6), C and F denote the set of censored and uncensored observations (See, Hosmer et al., 2008; Lawless, 2011; Falgore et al., 2024).

Residual Analysis for the LDUSTL$_{BHE}$ Model

The examination of the residual is an important step following the formulation of the LDUSTL$_{BHE}$. Hence, the residual analysis is employed to confirm whether the underlying assumtions of the model has significantly changed. The Cox-Snell residual analysis which offers a means of assessing the degree to which a model matches the observed survival data is normally utilized to evaluate the goodness of fit of survival models. The Cox-Snell residual is specifies as

$$e_i = -\log[S(y_i;x)], \quad i = 1,2,\dots,n. \tag{12}$$

The corresponding Cox-Shell residuals to the LDUSTL$_{BHE}$ regression model is specified as

$$e_i = -\log\left[1-\left\{\frac{e^{\left[1-\left(\frac{e^{-e^{\frac{y-\mu}{\sigma}}}}{1+e^{\frac{y-\mu}{\sigma}}}\right)^2\right]^{\gamma}-1}}{e-1}\right\}\right], \quad i = 1,2,\dots,n. \tag{13}$$

According to Lawless (2011), if the fitted model is sufficient, the residuals takes an exponential distributional form.

RESULTS

Data Description

The HIV datasets which comprise seven variables with 200 observations, were obtained from the federal medical center and Gombe state specialist hospital in North-East, Nigeria. The variables are the survival times (in months), age (in years), Gender (Male = *1*, Female = *2*), CD4-count, WHO-stage (Acute infection = *1*, Chronic (Asymptomatic) infection = *2*, Symptomatic HIV = *3*, AIDS = *4*), number of opportunist infections (Ols)

and Censoring (Censored = *0*, Un-censored = *1*). The dataset utilized in this study will be made available upon request.

For this research, the survival times is considered as the response variable y_i, while the independent variables are age (K_{i1}), Gender (K_{i2}), CD4 count (K_{i3}), WHO-stage (K_{i4}) and number of opportunist infections (K_{i5}). The LDUSTL$_{BHE}$ model is fitted alongside the Log-Topp Leone Burr Hatke exponential (LTLBHE) model by Anyiam et al. (2024), Log-Alpha power Burr Hatke exponential (LAPBHE) model by Adubisi et al. (2025) and Burr Hatke exponential (LBHE) model by Yadav et al. (2021). The fitted regression model is specified as

$$y_i = \xi_0 + \sum_{j=1}^{5} K_{ij}\xi_j + \sigma z_i\,, \quad i = 1,2,\ldots,200. \tag{14}$$

Table 1 presents the maximum likelihood estimates of the parameters of the LDUSTL$_{BHE}$ regression model together with those of the competing models (LTLBHE, LAPBHE, and LBHE). The table also reports model comparison criteria including the negative log-likelihood and several information criteria such as AIC, cAIC, BIC, and HQIC. The results indicate that several regression covariates significantly influence the survival times of HIV patients. In particular, the WHO disease stage and the number of opportunistic infections are statistically significant predictors in the LDUSTL$_{BHE}$ regression model at the 5% significance level. The estimated coefficients suggest that patients at more advanced WHO stages tend to experience shorter survival times. This finding is consistent with clinical understanding of HIV progression, where individuals in advanced stages of infection generally exhibit greater immune suppression and are more vulnerable to life-threatening complications.

Similarly, the number of opportunistic infections shows a significant negative association with survival duration. Opportunistic infections are known to occur when the immune system is severely compromised, and their presence often indicates deteriorating health status among HIV patients. The statistical significance of this variable therefore reflects its critical role in determining patient survival outcomes. Age is also found to have a notable effect on survival time. The estimated regression coefficient indicates that increasing patient age is associated with reduced survival duration. This observation aligns with existing epidemiological evidence suggesting that older individuals may experience more rapid disease progression due to weaker immune responses and the presence of comorbidities. In contrast, the variables gender and CD4 count do not appear to be

statistically significant predictors in the fitted models. While CD4 count is commonly regarded as an important indicator of immune function, its statistical insignificance in this dataset may be influenced by treatment effects such as antiretroviral therapy, which can stabilize CD4 levels over time.

From the standpoint of model performance, the $LDUSTL_{BHE}$ regression model exhibits the lowest values of the negative log-likelihood and information criteria among all competing models as reported in Table 1. This clearly indicates that the proposed model provides the best fit to the HIV survival data. The superior performance of the $LDUSTL_{BHE}$ model can be attributed to its enhanced flexibility, which enables it to capture complex survival patterns that cannot be adequately represented by simpler Burr-Hatke-based models.

Table 1. The MLEs for the fitted regression models to the HIV dataset.

Pa.	$LDUSTL_{BHE}$		LTLBHE		LAPBHE		LBHE	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
γ	0.3112	3.04e-19	0.3797	8.59e-23	46.1282	3.770e-01	-	-
σ	0.1051	5.59e-95	0.1049	1.17e-101	0.3045	1.767e-12	0.2027	1.225e-52
ξ_0	6.3555	0.0000	6.3677	0.0000	5.9949	0.0000	6.1943	0.0000
ξ_1	0.2670	1.32e-53	0.2806	1.97e-51	0.3598	2.329e-35	0.3583	8.593e-45
ξ_2	0.0078	8.16e-01	0.0041	9.09e-01	-0.0219	5.913e-01	-0.0277	5.138e-01
ξ_3	-0.0008	9.57e-01	0.0019	9.15e-01	-0.0031	8.683e-01	0.0002	9.939e-01
ξ_4	-0.0399	5.43e-03	-0.0418	8.88e-03	-0.0281	1.506e-01	-0.0320	1.162e-01
ξ_5	0.1022	7.23e-04	0.1173	3.44e-04	0.1360	2.822e-04	0.1482	1.458e-04
$-\mathcal{L}$	-26.564		-29.509		-29.082		-34.629	
AIC	69.128		75.018		74.163		83.257	
cAIC	103.474		109.364		108.510		113.310	
BIC	95.474		101.364		100.51		106.310	
HQIC	79.791		85.681		84.826		92.587	
Rank	1		3		2		4	

The likelihood ratio test (LRT) results reported in Table 2 provide additional statistical evidence supporting the superiority of the proposed $LDUSTL_{BHE}$ regression model. The likelihood ratio test is commonly used to compare nested statistical models by evaluating whether the inclusion of additional parameters leads to a statistically significant improvement in model fit. The test statistic is computed based on the difference in the log-

likelihood values of the competing models and follows a chi-square distribution under the null hypothesis.

In this study, the LRT compares the proposed $LDUSTL_{BHE}$ model with its nested or simpler alternatives, namely the LTLBHE, LAPBHE, and LBHE regression models. The null hypothesis assumes that the simpler model adequately explains the survival data, while the alternative hypothesis states that the more flexible $LDUSTL_{BHE}$ model provides a significantly better representation of the data. The results in Table 2 indicate that the likelihood ratio statistics are statistically significant when the $LDUSTL_{BHE}$ model is compared with each of the competing models. This implies that the additional parameter structure introduced in the $LDUSTL_{BHE}$ distribution significantly improves the model's ability to capture the underlying survival dynamics of the HIV dataset. Consequently, the null hypothesis that the simpler models are sufficient is rejected in favour of the more flexible $LDUSTL_{BHE}$ specification.

From a modeling perspective, these findings suggest that the extra flexibility incorporated into the $LDUSTL_{BHE}$ distribution plays an important role in describing the complex hazard behaviour observed in the HIV survival data. Survival times associated with chronic diseases such as HIV often exhibit heterogeneous patterns influenced by disease progression, treatment response, and patient-specific characteristics. Simpler parametric models may fail to accommodate such complexity due to restrictive assumptions about the shape of the hazard function. The significant LRT results therefore reinforce the conclusions obtained from the information criteria presented in Table 1. Both sets of results consistently indicate that the $LDUSTL_{BHE}$ regression model provides a more adequate statistical framework for analysing the HIV survival data. This improved fit highlights the practical value of developing flexible parametric survival models that can capture non-linear and heterogeneous survival patterns often observed in biomedical studies.

Overall, the LRT confirms that the additional distributional flexibility introduced in the $LDUSTL_{BHE}$ model leads to a statistically meaningful improvement in model performance. This supports the adoption of the proposed model for analysing censored survival data and provides further evidence of its suitability for studying HIV survival outcomes.

Table 2. Likelihood ratio test for the fitted regression models

Model comparison			LR-stat	df	p-value
BHE	vs	DUSTLBHE	16.12956	1	0.0000592
APBHE	vs	DUSTLBHE	5.03547	0	0.0000000
TLBHE	vs	DUSTLBHE	5.88998	0	0.0000000

Figures 2–5 present the Cox–Snell residual plots for the $LDUSTL_{BHE}$, LTLBHE, LAPBHE, and LBHE regression models. Cox–Snell residual analysis is a standard diagnostic technique used to assess the adequacy of fitted survival models. For a correctly specified survival model, the cumulative hazard function of the Cox–Snell residuals should approximate a straight line through the origin with slope equal to one. Among the four models, the residual plot corresponding to the $LDUSTL_{BHE}$ model exhibits the closest alignment with the theoretical reference line. This indicates that the proposed model provides a better representation of the underlying survival distribution of the HIV data. The residual plots for the other models show greater deviations from the expected line, suggesting a less adequate fit.

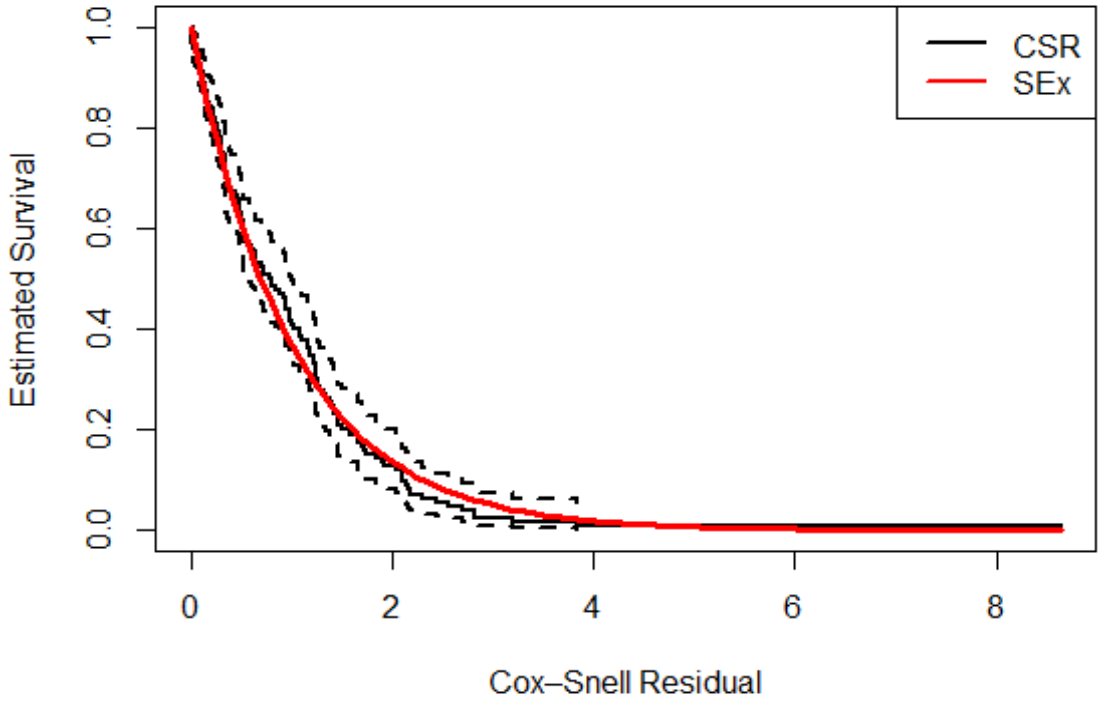


Figure 2. The $LDUSTL_{BHE}$ model Cox-Shell residual plot for the fitted HIV data

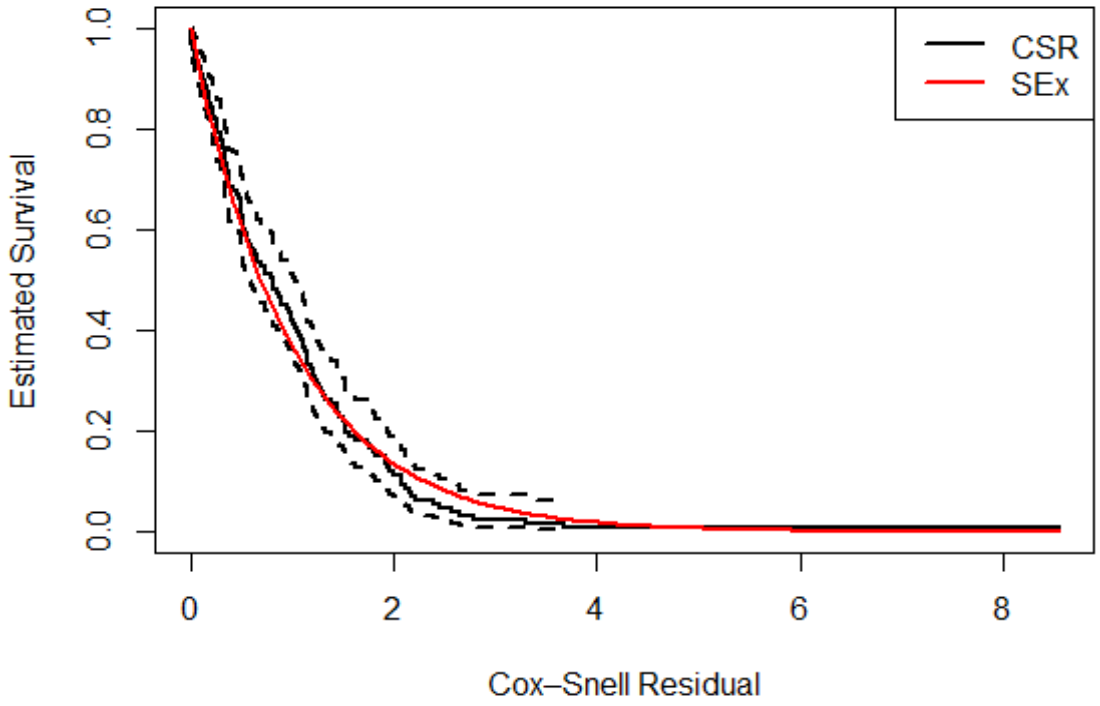


Figure 3. The LTLBHE model Cox-Shell residual plot for the fitted HIV data

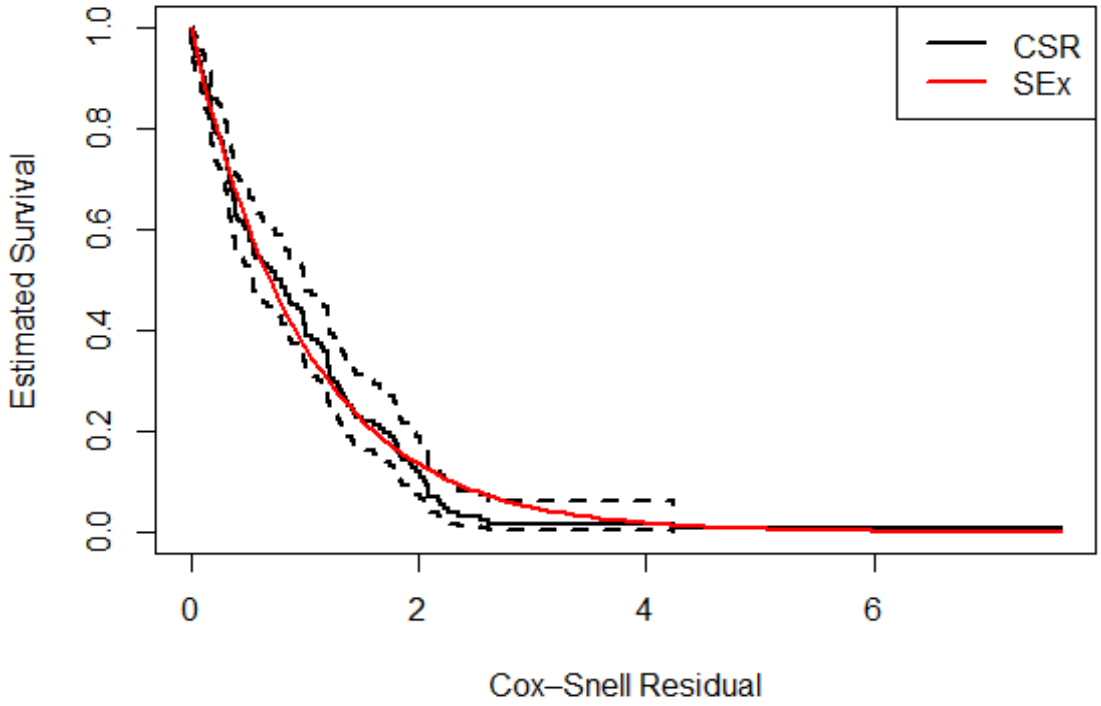


Figure 4. The LAPBHE model Cox-Shell residual plot for the fitted HIV data

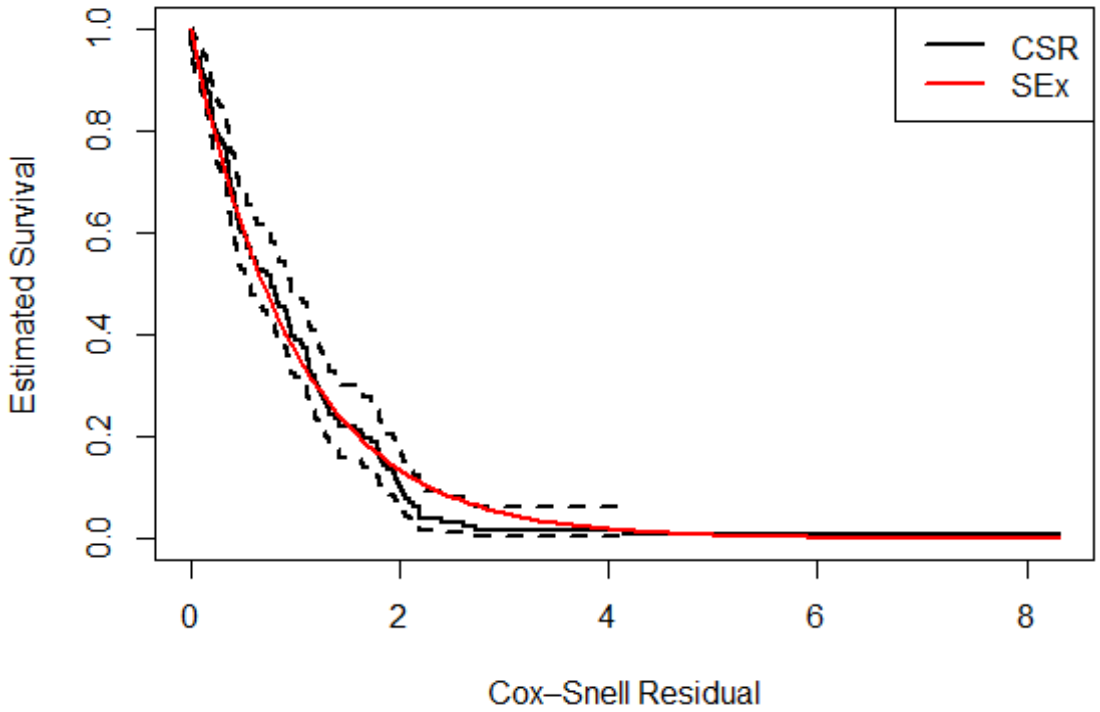


Figure 5. The LBHE model Cox-Shell residual plot for the fitted HIV data

Applications of Kaplan-Meier Survival Probability

The Kaplan–Meier estimator, commonly referred to as the product-limit estimator, is a non-parametric technique used in survival analysis to estimate the survival function from observed lifetime data, particularly in the presence of censored observations. The method introduced by Kaplan–Meier (1958), has become one of the most widely used tools in medical and reliability studies. It enables researchers to estimate the probability that a patient survives beyond a specified time and provides an intuitive graphical representation of survival experience over time.

Table 3 presents the Kaplan–Meier survival estimates for the HIV dataset across multiple time intervals. The table reports the number of individuals at risk, the number of observed events, the estimated survival probability, the corresponding standard error, and the 95% confidence interval. The results show a gradual decline in survival probability as time progresses. During the early months of follow-up, the survival probability remains relatively high, indicating that most patients survive during the initial stages after diagnosis or treatment initiation. However, as time increases, the survival probability steadily decreases, reflecting the progressive nature of HIV infection. A notable feature of the Kaplan–Meier estimates is the steeper reduction in survival probability observed during certain periods. These declines correspond to intervals with higher numbers of observed

events, suggesting increased mortality risk during these stages of disease progression. The widening confidence intervals at later time points are also expected in survival analysis. As the number of individuals remaining at risk decreases over time, the uncertainty associated with the survival estimates increases, leading to broader confidence bounds.

Overall, the Kaplan–Meier estimates provide valuable non-parametric insight into the survival dynamics of HIV patients and serve as an important benchmark for evaluating the adequacy of the parametric survival models considered in this study.

Table 3. Kaplan-Meier Survival Probability for the HIV Data

Time (in months)	Number of Risks	Number of events	Survival probability	Survival error	95% Confidence Interval	
					Lower	Upper
24	199	1	0.994975	0.005038	0.985199	1.00000
36	198	1	0.989950	0.007143	0.976188	1.00000
60	197	0	0.989950	0.007143	0.976188	1.00000
72	195	2	0.979796	0.010206	0.960392	0.99959
84	188	0	0.979796	0.010206	0.960392	0.99959
96	176	1	0.974229	0.011689	0.952164	0.99681
108	174	1	0.968630	0.013033	0.944202	0.99369
120	173	1	0.963031	0.014264	0.936481	0.99033
132	172	0	0.963031	0.014264	0.936481	0.99033
144	171	1	0.957400	0.015422	0.928893	0.98678
156	169	1	0.951735	0.016525	0.921403	0.98306
168	168	1	0.946069	0.017570	0.914044	0.97922
180	167	2	0.934739	0.019527	0.899641	0.97121
192	165	0	0.934739	0.019527	0.899641	0.97121
204	164	1	0.929040	0.020462	0.892517	0.96706
216	163	2	0.917640	0.022247	0.878488	0.95854
228	160	1	0.911905	0.023113	0.871516	0.95417
240	159	1	0.906170	0.023959	0.864601	0.94974
252	157	2	0.894626	0.025617	0.850818	0.94069
276	155	2	0.883083	0.027213	0.837216	0.93146
288	153	3	0.865767	0.029517	0.817102	0.91733
300	150	4	0.842680	0.032464	0.790732	0.89804
312	146	2	0.831137	0.033898	0.777712	0.88823
336	143	3	0.813700	0.036040	0.758205	0.87326
348	138	3	0.796011	0.038209	0.738577	0.85791
360	133	4	0.772071	0.041147	0.712251	0.83692

372	129	3	0.754116	0.043332	0.692714	0.82096
384	126	3	0.736161	0.045511	0.673339	0.80484
396	122	2	0.724092	0.046988	0.660386	0.79394
408	119	3	0.705838	0.049246	0.640895	0.77736
420	114	3	0.687263	0.051597	0.621161	0.76040
432	110	2	0.674768	0.053203	0.607950	0.74893
444	108	2	0.662272	0.054821	0.594803	0.73739
456	106	5	0.631033	0.058926	0.562204	0.70829
468	101	4	0.606041	0.062294	0.536386	0.68474
480	95	5	0.574144	0.066824	0.503664	0.65449
492	90	5	0.542248	0.071547	0.471299	0.62388
504	84	4	0.516426	0.075593	0.445311	0.59890
516	77	7	0.469478	0.083743	0.398413	0.55322
528	69	3	0.449066	0.087588	0.378229	0.53317
540	65	6	0.407614	0.096105	0.337632	0.49210
552	58	5	0.372475	0.104225	0.303655	0.45689
564	52	4	0.343823	0.111648	0.276248	0.42793
576	48	4	0.315171	0.119830	0.249199	0.39861
588	43	3	0.293182	0.126899	0.228623	0.37597
600	39	6	0.248077	0.144102	0.187037	0.32904
612	33	3	0.225525	0.154259	0.166682	0.30514
624	29	3	0.202195	0.166657	0.145852	0.28030
636	25	0	0.202195	0.166657	0.145852	0.28030
648	23	6	0.149448	0.207653	0.099480	0.22452
660	17	2	0.131866	0.225750	0.084718	0.20525
672	14	0	0.131866	0.225750	0.084718	0.20525
684	13	1	0.121723	0.239527	0.076118	0.19465
708	12	1	0.111579	0.254851	0.067710	0.18387
720	10	1	0.100421	0.275790	0.058489	0.17242
732	9	2	0.078105	0.328338	0.041039	0.14865
744	7	0	0.078105	0.328338	0.041039	0.14865
768	6	1	0.065088	0.375685	0.031168	0.13592
780	5	2	0.039053	0.523902	0.013986	0.10904
792	3	1	0.026035	0.664183	0.007083	0.09570
840	2	1	0.013018	0.970123	0.001944	0.08716
948	1	0	0.013018	0.970123	0.001944	0.08716

Figure 6 illustrates the Kaplan–Meier survival curve for the HIV dataset. The curve displays the probability of surviving beyond a given time point and provides a graphical representation of the survival experience of the patients. The curve shows a monotonic decline over time, reflecting the natural progression of HIV infection. Several step decreases are observed in the curve, corresponding to observed failure events in the dataset.

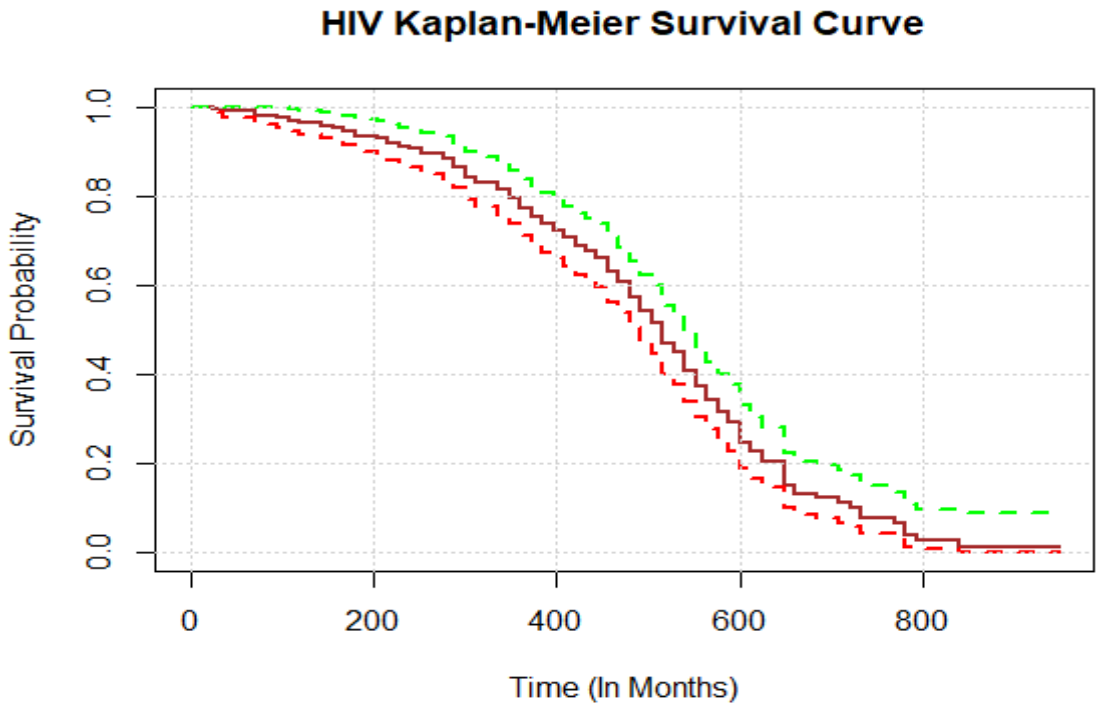


Figure 6. HIV dataset Kaplan-Meier Survival Curve

Figures 7–10 compare the Kaplan–Meier survival curve with the fitted survival functions obtained from the $LDUSTL_{BHE}$, LTLBHE, LAPBHE and LBHE models. The fitted survival curve from the $LDUSTL_{BHE}$ model closely follows the empirical Kaplan–Meier curve across most of the time range. This indicates that the proposed model successfully captures the underlying survival behaviour of the HIV dataset. In contrast, the fitted curves corresponding to the competing models exhibit larger deviations from the Kaplan–Meier curve, particularly at later time periods. These discrepancies suggest that the competing models lack the flexibility required to adequately represent the survival dynamics present in the data.

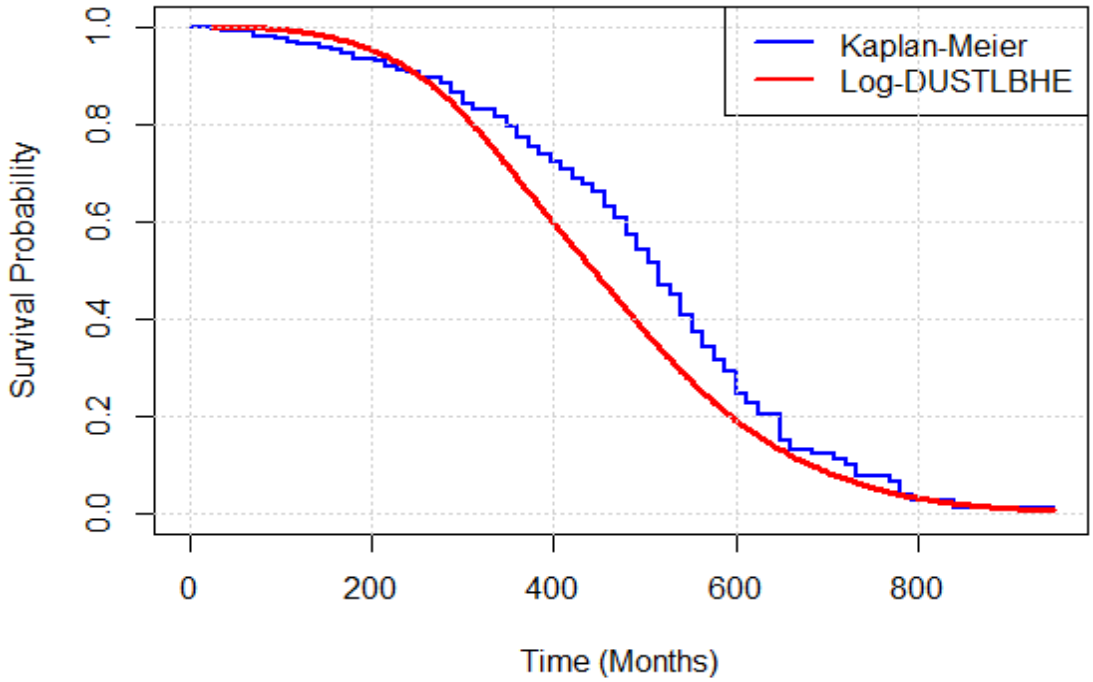


Figure 7. HIV Kaplan-Meier and LDUSTL$_{BHE}$ Survival Curve

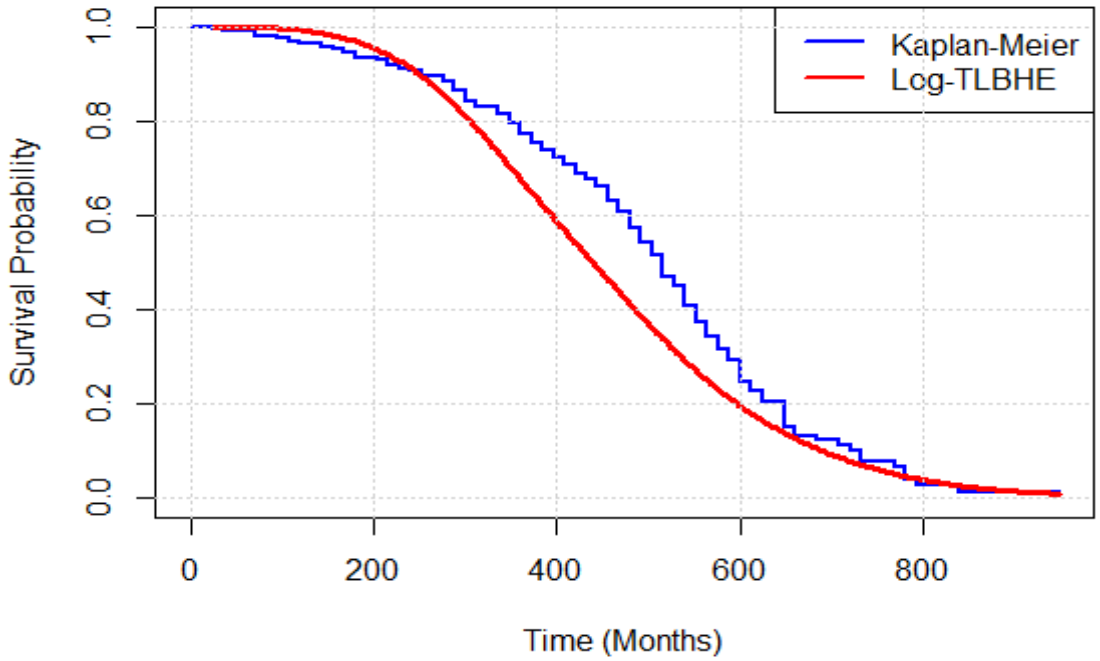


Figure 8. HIV Kaplan-Meier and LTLBHE Survival Curve

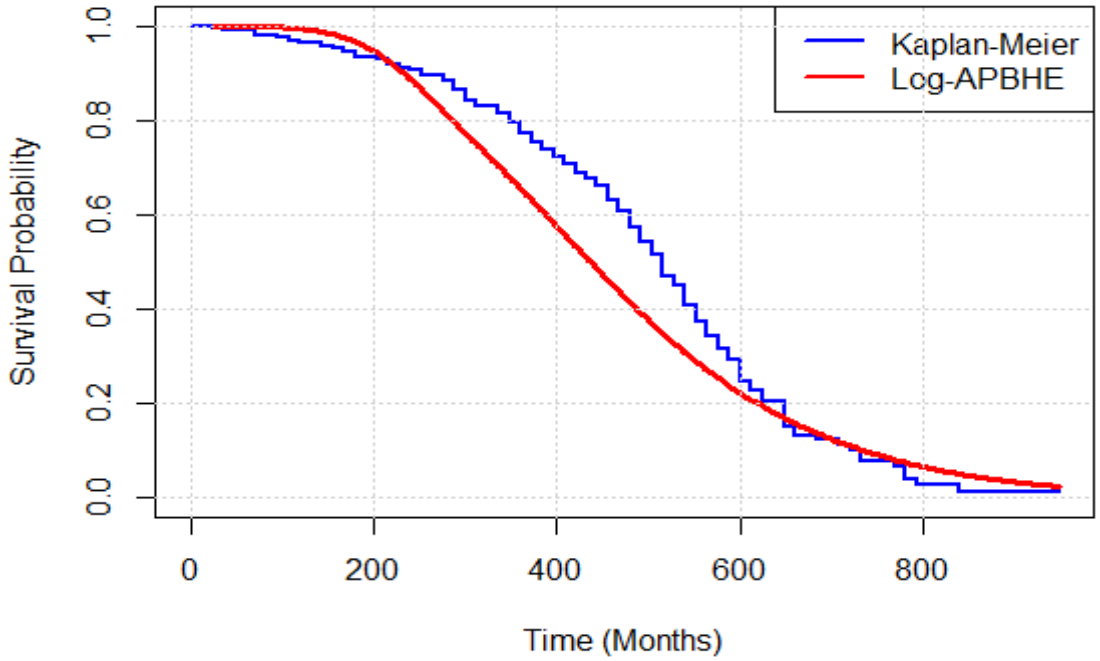


Figure 9. HIV Kaplan-Meier and LAPBHE Survival Curve

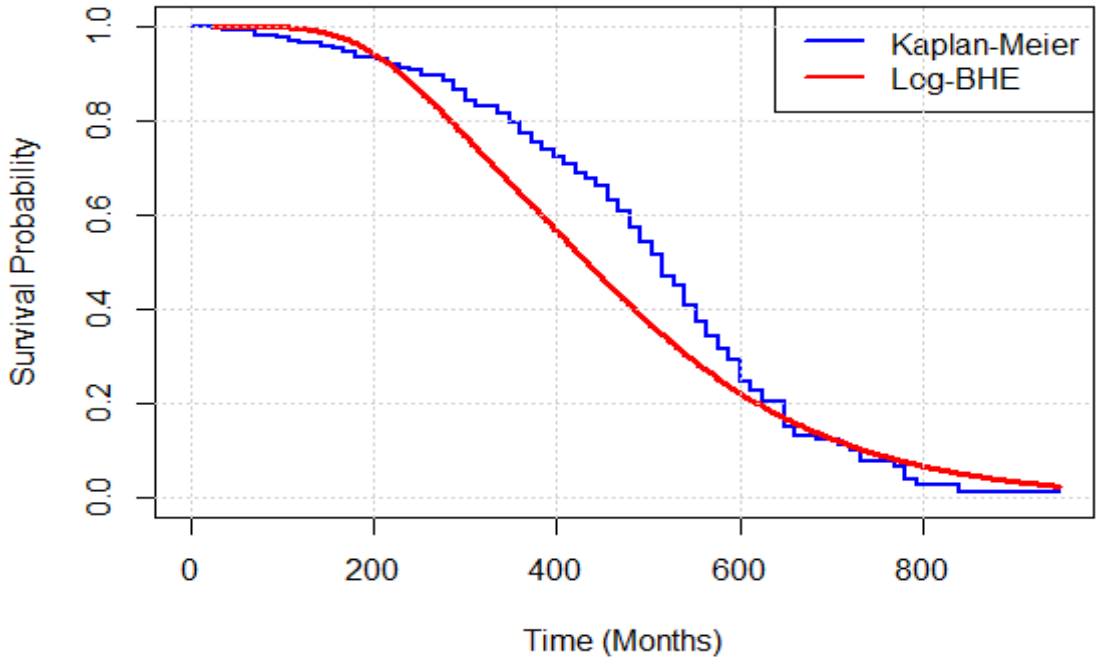


Figure 10. HIV Kaplan-Meier and LBHE Survival Curve

DISCUSSION

This study investigated the survival patterns of HIV patients using both non-parametric and parametric survival modelling techniques, with particular emphasis on the proposed LDUSTL$_{BHE}$ regression model. The analysis combined the Kaplan–Meier

survival estimator with flexible parametric regression models in order to provide a comprehensive understanding of the survival behaviour of the HIV dataset. The Kaplan–Meier survival curves provide an initial non-parametric assessment of survival experience among the patients. The estimated survival function indicates a gradual decline in survival probability as time progresses, reflecting the progressive nature of HIV infection. The Kaplan–Meier approach is particularly useful because it accommodates censored observations and provides an empirical description of the survival distribution without imposing strict distributional assumptions. The curves therefore serve as a useful benchmark for evaluating the adequacy of the parametric survival models considered in this study.

The results presented in Table 1 compare the goodness-of-fit of the proposed $LDUSTL_{BHE}$ regression model with several competing models, including the LTLBHE, LAPBHE, and LBHE models. Model performance was evaluated using commonly used information criteria such as the AIC, cAIC, BIC and HQIC. These criteria balance model complexity and goodness of fit by penalizing excessive parameterization. The results show that the $LDUSTL_{BHE}$ regression model produced the smallest values of these information criteria compared with the alternative models. Smaller values of these statistics indicate a better trade-off between model fit and parsimony, suggesting that the proposed model provides the most adequate representation of the survival data among the models considered.

The likelihood ratio test results reported in Table 2 provide further statistical support for the superiority of the proposed model. The likelihood ratio test compares nested models by examining whether the inclusion of additional parameters leads to a statistically significant improvement in model fit. In this study, the $LDUSTL_{BHE}$ model was compared with its simpler counterparts, and the test statistics indicated statistically significant improvements in fit when the additional parameters were included. These findings suggest that the increased flexibility of the $LDUSTL_{BHE}$ model enables it to capture important characteristics of the survival distribution that may not be adequately represented by the simpler models.

The improved performance of the $LDUSTL_{BHE}$ model can be attributed to its ability to accommodate complex hazard rate behaviours commonly observed in medical survival data. In many biomedical applications, particularly in chronic diseases such as

HIV/AIDS, the hazard function may exhibit non-monotonic patterns due to variations in disease progression, treatment effectiveness, and patient heterogeneity. Traditional parametric survival models often impose restrictive assumptions on the shape of the hazard function, which may lead to inadequate model fit. By contrast, the proposed $LDUSTL_{BHE}$ model provides greater flexibility in modelling these complex survival dynamics.

The graphical results presented in the figures further support the adequacy of the proposed model. The fitted survival curves closely follow the empirical Kaplan–Meier curves, indicating that the model provides a realistic representation of the observed survival patterns. In addition, the density and hazard function plots demonstrate the flexibility of the LDUSTLBHE distribution in capturing different shapes of survival behaviour. Such graphical diagnostics are important because they allow researchers to visually assess how well the model reproduces the characteristics of the observed data. From a practical standpoint, the findings of this study have important implications for survival analysis in medical research. Accurate modelling of survival data is essential for understanding disease progression, evaluating treatment outcomes, and informing clinical decision-making. The results suggest that flexible parametric survival models such as the $LDUSTL_{BHE}$ model may provide improved analytical tools for studying survival outcomes in HIV patients and other chronic disease populations.

Overall, the combination of information criteria, likelihood ratio tests, and graphical diagnostics consistently indicates that the $LDUSTL_{BHE}$ regression model offers a superior fit to the HIV survival dataset. These findings highlight the value of developing flexible statistical models capable of capturing complex survival patterns often encountered in real-world biomedical data.

CONCLUSION

This study proposed a novel survival regression model based on the $LDUSTL_{BHE}$ distribution for analysing censored survival data. The model was applied to HIV survival data from Gombe State, Nigeria, and its performance was compared with several competing Burr-Hatke exponential based survival models. The empirical analysis demonstrated that the proposed model provides the best fit to the data according to several model selection criteria and diagnostic plots. The results also revealed that advanced WHO

disease stage, increasing number of opportunistic infections, and higher patient age are associated with reduced survival time among HIV patients. These findings highlight the importance of incorporating flexible statistical models in survival analysis to better understand disease progression and patient outcomes. Future research may explore extensions of the model to other datasets, investigate Bayesian estimation approaches, and develop multivariate or frailty versions of the model for more complex survival data structures.

REFERENCES

Abbas, U. F., & Adubisi, O. D. (2026). Novel DUS Topp–Leone Burr–Hatke exponential model with properties and application to COVID-19 datasets. *Modern Journal of Statistics, 2*(2), 30–56. https://doi.org/10.64389/mjs.2026.02268

Abdulkadir, A., Adubisi, O. D., & Madaki, R. M. (2024). Novel extended Weibull regression model for investigating the survival times of breast cancer patients. *Mikailalsys Journal of Mathematics and Statistics, 2*(3), 135–155.

Adamu, P. I., Oguntunde, P. E., Okagbue, H. I., & Agboola, O. O. (2018). On the epidemiology and statistical analysis of HIV/AIDS patients in the insurgency-affected states of Nigeria. *Open Access Macedonian Journal of Medical Sciences, 6*(7), 1315–1321. https://doi.org/10.3889/oamjms.2018.229

Adubisi, O. D., Joshua, T., & David, A. A. (2025). Alpha-power-Burr–Hatke-exponential model: Properties, simulations, and applications based on uncensored and progressive type-II-censored samples. *Reliability: Theory & Applications, 20*(3), 455–472. https://doi.org/10.24412/1932-2321-2025-386-455-472

Ali, I., & Abbas, M. (2025). Application of Weibull survival regression for predicting hospital length of stay among hepatitis B patients in Maiduguri, Nigeria. *FUDMA Journal of Sciences, 9*(12), 262–268.

Anyiam, K. E., Alghamdi, F. M., Nwaigwe, C. C., Aljohani, H. M., & Obulezi, O. J. (2024). A new extension of Burr–Hatke exponential distribution with engineering and biomedical applications. *Heliyon, 10*(19), e38293. https://doi.org/10.1016/j.heliyon.2024.e38293

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187–220. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Falgore, J. Y., Yahaya, A., Doguwa, S. I., Mohammed, A. S., & Imam, A. T. (2024). A new survival regression model with application to HIV data. *European Journal of Statistics, 4*, Article 10. https://doi.org/10.28924/ada/stat.4.10

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data* (2nd ed.). Wiley. https://doi.org/10.1002/9780470258019

Isa, A. M., Doguwa, S. I., Alhaji, B. B., & Dikko, H. G. (2025). Extension of the Weibull survival regression based on sine type-II Topp–Leone-G family of distribution: Properties and application. *Iraqi Statisticians Journal, 2*(1), 74–90. https://doi.org/10.62933/z1y1sr28

Ishag, M. A. S., Wanjoya, A., Adem, A., Alsultan, R., Alghamdi, A. S., & Afify, A. Z. (2024). The exponentiated-Weibull proportional hazard regression model with application to censored survival data. *Alexandria Engineering Journal, 104*, 587–602. https://doi.org/10.1016/j.aej.2024.08.007

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457–481. https://doi.org/10.1080/01621459.1958.10501452

Khikman, M. A., Multiyaningrum, R., Kholifah, R. I. N., Sa'adah, L. N., Safira, E. L., Sarah, A. D., Amri, I. F., & Al Haris, M. (2025). Survival analysis using Kaplan–Meier and Cox regression in hypertension patients at Kefamenanu Regional Hospital. *Eigen Mathematics Journal, 8*(2), 123–132. https://doi.org/10.29303/emj.v8i2.270

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4419-6646-9

Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (2nd ed.). Wiley. https://doi.org/10.1002/9781118033005

Li, J., & Zang, K. (2025). Survival regression with the new Weibull–Pareto distribution under right censoring. *Bulletin of Computer and Data Sciences, 6*(2), 39–50. https://doi.org/10.71448/bcds2562-3

Lima, I. S., Silva, S. J., Brighenti, C. R. G., Nakamura, L. R., Oliveira, T. A., Oliveira, M. E. C., & Ramires, T. G. (2025). Survival time analysis in women with breast cancer using distributional regression models. *Cadernos de Saúde Pública, 41*(8), e00073324. https://doi.org/10.1590/0102-311XEN073324

Liu, H., Lopa, S. H., & Wahed, A. S. (2026). Log-skew-logistic distribution and regression model for censored survival data. *Journal of Statistical Theory and Practice, 20*, Article 15. https://doi.org/10.1007/s42519-025-00537-0

Mezher, A. S., & Ibrahim, W. S. (2024). Using the Bayesian method to estimate the exponential survival regression model for stomach cancer patients in Iraq. *Journal of Statistics and Management Systems, 27*(8), 1691–1700. https://doi.org/10.47974/JSMS-1268

National Agency for the Control of AIDS. (2022). *Global AIDS response: Nigeria country progress report 2022*.

Olanrewaju, S. O., & Adeniran, E. K. (2026). Survival analysis of prostate cancer patients using Cox regression model and log-logistic model. *International Journal of Latest Technology in Engineering, Management & Applied Science, 15*(1), 577–589. https://doi.org/10.51583/IJLTEMAS.2026.150100051

Onovo, A. A., Adeyemi, A., Onime, D., Kalnoky, M., Kagniniwa, B., Dessie, M., Lee, L., Parrish, D., Adebobola, B., Ashefor, G., Ogorry, O., Goldstein, R., & Meri, H. (2023). Estimation of HIV prevalence and burden in Nigeria: A Bayesian predictive modelling study. *eClinicalMedicine, 62*, 102098. https://doi.org/10.1016/j.eclinm.2023.102098

Sadiq, I. A., Kajuru, J. Y., Usman, A., Doguwa, S. I., & Yahaya, A. (2025). The NGOF-Et-Weibull survival regression model with application to liver cancer time-to-event data. In *Proceedings of the 1st International Conference on Statistical Sciences and Data Analytics (ICSSDA 2025)* (pp. 108–127).

Sheehy, M., Tun, W., Vu, L., Adebajo, S. B., Obianwu, O., & Karlyn, A. (2014). High levels of bisexual behavior and factors associated with bisexual behavior among men having sex with men (MSM) in Nigeria. *AIDS Care, 26*(1), 116–122. https://doi.org/10.1080/09540121.2013.802281

Udofia, E. M., Umeh, E. U., & Onyekwere, C. K. (2021). Modeling of survival of HIV patients by stages of immune suppression and opportunisic infections. *American Journal of Theoretical and Applied Statistics, 10*(6), 233–242. https://doi.org/10.11648/j.ajtas.20211006.12

UNAIDS. (2023). *The path that ends AIDS: UNAIDS Global AIDS Update 2023*. https://www.unaids.org/en/resources/documents/2023/global-aids-update-2023

Yadav, A. S., Altun, E., & Yousof, H. M. (2021). Burr–Hatke exponential distribution: A decreasing failure rate model, statistical inference and applications. *Annals of Data Science, 8*, 241–260. https://doi.org/10.1007/s40745-019-00213-8

Yusuf, H., Sadiq, I. A., Hephzibah, A. A., Usman, A., Abdullahi, J., David, R. O., Adamu, S. A., & Abdulkarim, H. (2025). An application of the Kaplan–Meier survivability probability and a log-extended Weibull survival regression model to the acute myocardial infarction dataset. In *Proceedings of the 1st International Conference on Statistical Sciences and Data Analytics (ICSSDA 2025)* (pp. 267–276).