Proceedings of the 1st International Conference on Statistical Sciences and Data Analytics

THE NGOF-ET-WEIBULL SURVIVAL REGRESSION MODEL WITH APPLICATION TO LIVER CANCER TIME-TO-EVENT DATA

Ibrahim Abubakar Sadiq [1]*, Jibril Yahaya Kajuru [1], Abubakar Usman [1], Sani Ibrahim Doguwa [1] and Abubakar Yahaya [1]

[1] Department of Statistics, Ahmadu Bello University, Zaria-Nigeria

Corresponding Author email: isabubakar@abu.edu.ng

ABSTRACT

This study applies the novel Log-NGOF-Et-Weibull Survival Regression (Log-NGOFEtWSR) model to liver cancer time-to-event data to enhance survival analysis with improved estimation and fit. Using data from 40 patients (72.5% uncensored, 27.5% censored) reported by Charles (2023), survival time in weeks was modelled against age, gender, AFP levels, and tumour stage. Parameters were estimated using Maximum Likelihood Estimation (MLE). The Log-NGOFEtWSR model outperformed the conventional Log-Weibull Survival Regression (Log-WSR), achieving lower -2LL (105.18 compared to 128.65), AIC (137.18 compared to 140.65), and BIC (164.20 compared to 174.92). Significant shape parameters included $\beta = 0.4517$ $(p = 0.0060)$, $\gamma = 2.6418$ $(p = 0.0375)$, and $\delta = 0.3348$ $(p = 0.0375)$. Cox-Snell residual analysis confirmed superior adequacy, while significant covariates were age >50 years ($\pi 3 = -3.6144, p < 0.001$), female gender ($\pi 5 = 3.0275, p < 0.001$), and tumour stage 1 ($\pi 8 = -1.9897, p < 0.001$). Kaplan-Meier analysis showed survival declining from 0.95 at 2 weeks to 0.32 at 41 weeks, with death observed up to 80 weeks. The Log-NGOFEtWSR survival curve closely matched the Kaplan-Meier estimate, confirming its robustness. Generally, the Log-NGOFEtWSR model offers a flexible, accurate, and reliable framework for liver cancer survival analysis, outperforming traditional models.

Keywords: Survival Analysis, Liver Cancer, NGOF-Et-Weibull Distribution, Kaplan-Meier, Maximum Likelihood Estimation, Cox-Snell Residuals, Model Fit.

1. INTRODUCTION

Liver cancer, one of the leading causes of cancer-related deaths globally, presents a significant challenge in clinical oncology due to its late diagnosis and poor prognosis (Anwanwan *et al*., 2020; Li *et al*., 2024). According to the Global Cancer Observatory, liver cancer was responsible for over 830,000 deaths in 2020, making it the third leading cause of cancer mortality worldwide (Bray *et al*., 2021). The high fatality rate is largely attributed to late-stage detection, limited treatment options, and the complex biological behaviour of the disease (Hong& Ding, 2025; Mathur *et al*., 2025; Mamun *et al*., 2024; Crosby *et al*., 2022). Survival analysis plays a crucial role in assessing the prognosis of liver cancer patients and evaluating the effect of potential risk factors and treatment strategies (Akor *et al.*, 2025; Usman *et al.*, 2025; Sadiq *et al.*, 2025). Traditional models such as the Kaplan-Meier estimator and Cox proportional hazards model have been widely used in medical statistics for analyzing time-to-event data (Semary *et al*., 2025; Klein & Moeschberger, 2003). However, these models have limitations, Kaplan-

Meier is non-parametric and does not incorporate covariates, while the Cox model assumes proportional hazards, which may not hold in many real-life scenarios, especially in heterogeneous clinical data (Semary *et al*., 2025; Sadiq *et al.,* 2025).

In recent years, parametric survival regression models have gained attention for their flexibility in modelling complex survival patterns. These models allow for the incorporation of covariates and can provide better estimates when the underlying hazard function is known or can be approximated (Lawless, 2003). Among the various parametric families, Weibull models are popular due to their ability to model increasing or decreasing hazard rates depending on the shape parameter (Murphy, 2007).

To further enhance the flexibility of survival models, several generalized versions of the Weibull distribution have been proposed. One such innovation is the NGOF-Et-Weibull (New Generalized Odd Fréchet Exponentiated-Weibull) distribution, which extends the baseline Weibull model by introducing additional shape parameters that can capture a wider range of hazard behaviours, including bathtub-shaped, monotonic increasing, and decreasing hazards (Sadiq *et al.*, 2023). This flexibility is particularly valuable in modelling the survival times of liver cancer patients, where hazard rates may change drastically across different stages of the disease or based on biomarker levels.

Incorporating the NGOF-Et-Weibull distribution into a regression framework allows for simultaneous modelling of the survival time and identification of prognostic covariates such as age, sex, alpha-fetoprotein levels, and tumour size. These variables are clinically known to influence liver cancer prognosis (Marrero *et al*., 2005). By integrating both parametric modelling and covariate analysis, the NGOF-Et-Weibull survival regression model offers a powerful tool for understanding survival dynamics and guiding medical decision-making (Sadiq *et al.*, 2023).

Furthermore, comparing the fitted NGOF-Et-Weibull model to the Kaplan-Meier estimator provides insight into its empirical adequacy. Finally, residual analysis using Cox-Snell residuals offers a means to evaluate the goodness-of-fit of the model, helping validate its application in clinical settings (Collett, 2015).

In summary, this study aims to apply the NGOF-Et-Weibull survival regression model to liver cancer time-to-event data, estimate survival probabilities, identify prognostic factors, and compare model performance with standard non-parametric techniques.

2. METHODOLOGY

The NGOF-Et-Weibull Distribution: Semary *et al*. (2025) built on the research of Sadiq *et al*. (2023) to introduce a new flexible model called the NGOF-Et-Weibull Distribution, which is part of the New Generalized Odd Frechet-exponentiated-G family of distributions earlier developed by Sadiq *et al*. (2023). A random variable X is said to have an NGOF-Et-Weibull distribution with PDF, CDF, hazard, and survival functions defined for all $x;\alpha,\beta,\gamma,\delta,\phi,\omega>0$ by Semary *et al*. (2025) as

$$f(x)=\beta\gamma\alpha^{\beta}\delta\left(\omega\phi^{-\omega}x^{\omega-1}\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-(\gamma\delta+1)}$$
$$\times\left(\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-\gamma\delta}-1\right)^{\beta-1}\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-\gamma\delta}-1\right)\right)^{\beta}\right\} \quad (1)$$

$$F(x)=\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-\gamma\delta}-1\right)\right)^{\beta}\right\} \quad (2)$$

$$h(x)=\beta\gamma\alpha^{\beta}\delta\left(\omega\phi^{-\omega}x^{\omega-1}\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-(\gamma\delta+1)}$$
$$\times\left(\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-\gamma\delta}-1\right)^{\beta-1}\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-\gamma\delta}-1\right)\right)^{\beta}\right\} \quad (3)$$
$$\times\left(1-\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-\gamma\delta}-1\right)\right)^{\beta}\right\}\right)^{-1}$$

$$S(x)=1-\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\left(\frac{x}{\phi}\right)^{\omega}\right\}\right)^{-\gamma\delta}-1\right)\right)^{\beta}\right\} \quad (4)$$

The plots of the PDF and CDF of this distribution with varying δ and keeping other parameters constant are shown in Figures 1 and 2 respectively. The plots of the hazard function and survival function of this distribution with varying δ and keeping other parameters constant are shown in Figures 3 and 4 respectively.

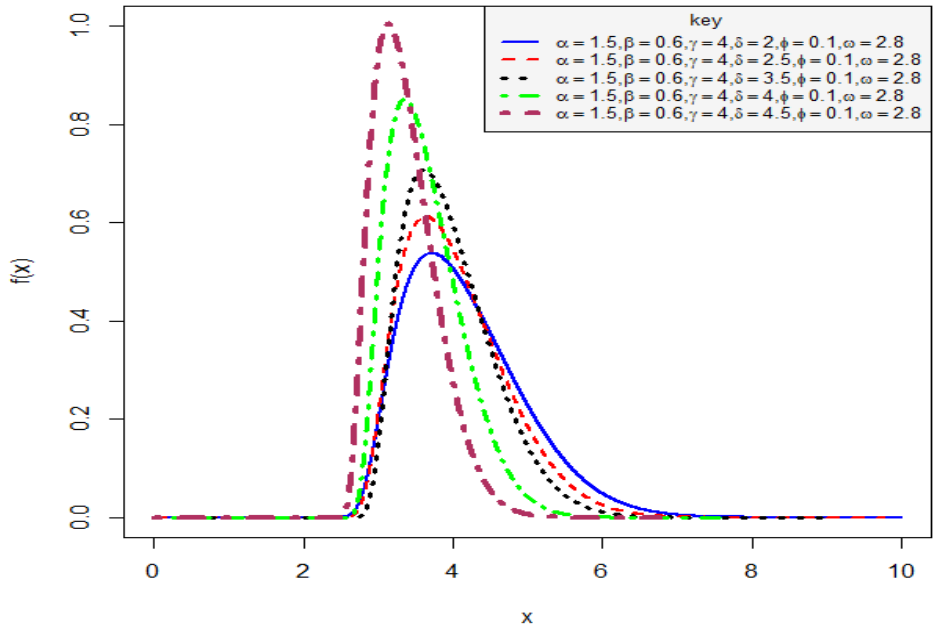

Figure 1: PDF Plot of NGOF-Et-Weibull Distribution

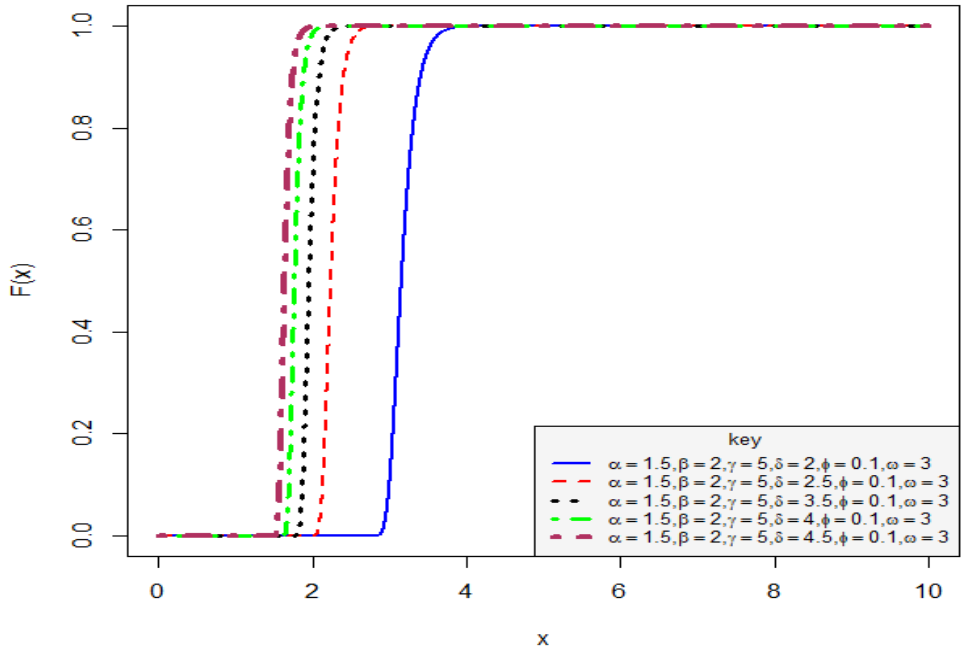

Figure 2: CDF Plot of NGOF-Et- Weibull Distribution

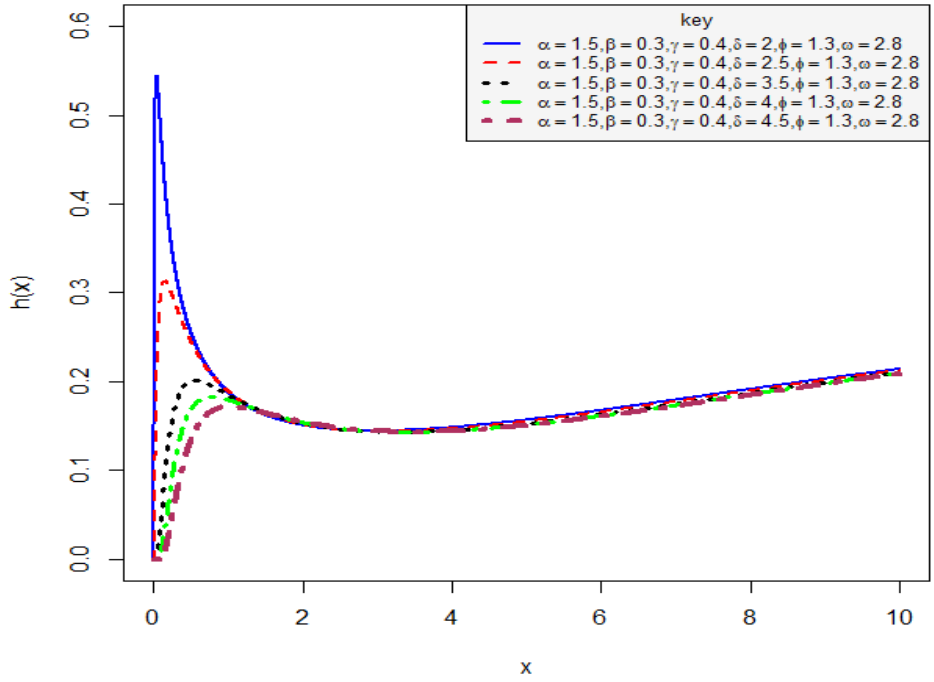

Figure 3: HF Plot of NGOF-Et-Weibull Distribution

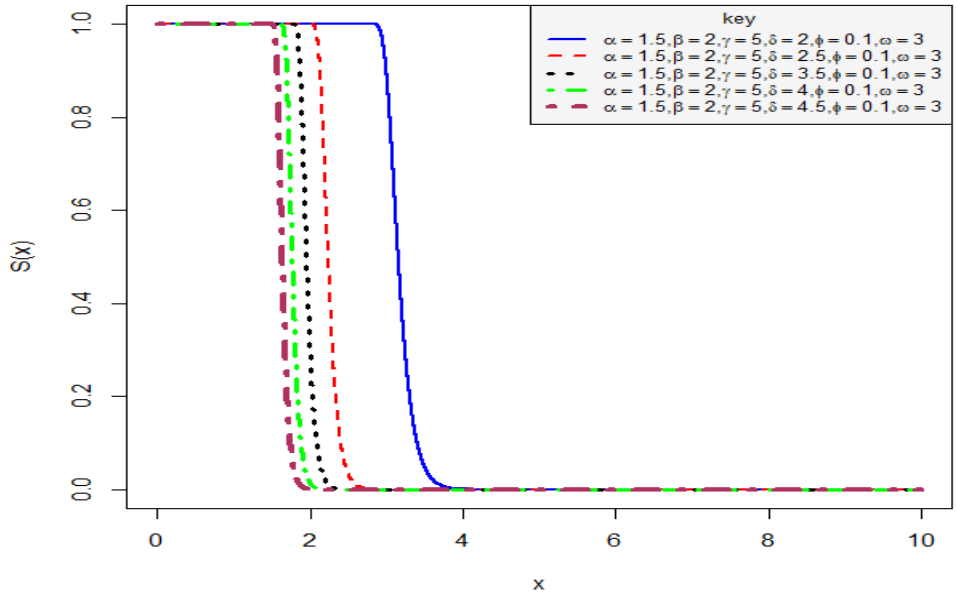

Figure 4: SF Plot of NGOF-Et-Weibull Distribution

The plots of the PDF in Figure 1 of the NGOF-Et-W distribution with varying δ and keeping other parameters constant explore the impact of the delta (δ) parameter. Based on the impact of the shape of the distribution, delta (δ) primarily controls the shape of the right tail of the distribution. Lower delta (δ < 2.5) the tail becomes lighter, indicating a lower probability of observing lower values. Higher delta (δ > 2.5) the tail becomes heavier, indicating a higher probability of observing larger values, gradually approaching an exponential-like decay. While the left tail is also slightly affected by delta, the dominant effect is on the right side.

The plots of the CDF in Figure 2 of the NGOF-Et-W distribution with varying δ and keeping other parameters constant explore the impact of the alpha (δ) parameter and show several CDF curves. Each curve corresponds to a specific value of the exponentiated parameter delta. Since all other parameters are constant, the curves illustrate the effect of varying delta on the CDF. The curve with the lowest delta (δ = 2) reaches a specific F(x) value at a lower x value compared to the other curves. This means the probability of X being less than or equal to that x value is higher for the lower delta. As delta increases (δ = 2.5, 3.5, 4 and 4.5), the curves shift to the right. The same F(x) value is now reached at a higher x value. This indicates a lower probability of X being less than or equal to that x value for a higher delta.

Figure 3 shows the NGOF-Et-W distribution hazard function curves. Each curve corresponds to a specific value of delta keeping all other parameters constant, the curves illustrate the effect of varying delta on hazard. The hazard function reveals the chance of failure after surviving previous periods. It typically starts at a certain value, increases or stays constant for a while, and then eventually decreases towards zero and then remains constant. As delta (δ) increases in the NGOFW distribution, the peak of the hazard function curve shifts to the right. For a given time value x, the corresponding h(x) value will be lower for a higher delta. This indicates the likelihood of failure at a particular time x decreases as the delta increases. This indicates that the system or component has a higher likelihood of surviving for a longer period before failure becomes more probable.

Figure 4 shows several survival function curves for the NGOF-Et-W distribution. Each curve corresponds to a specific value of the exponentiated parameter delta. Since all other parameters are constant, the curves illustrate the effect of varying delta on survival.

The curve with the highest delta (δ = 4.5) has a higher S(x) value for a given time x compared to the other curves. This means the probability of surviving beyond that time x is greater for the higher delta. As delta decreases (δ = 4, 3.5 and 2), the curves shift to the right. The same S(x) value is now reached at a lower x value. This indicates a lower probability of surviving beyond that time x for a lower delta. In essence, by keeping all other parameters constant, a higher delta in the NGOF-Et-W distribution explains a higher probability of a system or component surviving for a longer duration.

The Log-NGOF-Et-Weibull Distribution: Semary *et al*. (2025) made a logarithmic transformation and came up with a survival model called the Log-NGOF-Et-Weibull Distribution with PDF and survival function as

$$g(y)=\frac{\beta\gamma\alpha^{\beta}\delta}{\rho}\exp\left\{\frac{y-\mu}{\rho}\right\}\exp\left\{-\exp\left(\frac{y-\mu}{\rho}\right)\right\}\left(1-\exp\left\{-\exp\left(\frac{y-\mu}{\rho}\right)\right\}\right)^{-(\gamma\delta+1)}$$
$$\times\left(\left(1-\exp\left\{-\exp\left(\frac{y-\mu}{\rho}\right)\right\}\right)^{-\gamma\delta}-1\right)^{\beta-1}\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\exp\left(\frac{y-\mu}{\rho}\right)\right\}\right)^{-\gamma\delta}-1\right)\right)^{\beta}\right\} \quad (5)$$

$$S(y)=1-\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\exp\left(\frac{y-\mu}{\rho}\right)\right\}\right)^{-\gamma\delta}-1\right)\right)^{\beta}\right\} \quad (6)$$

where $y\in\Re$, $\beta,\gamma,\alpha,\delta,\rho>0$ and $-\infty<\mu<\infty$

However, we can re-write $Z=(y-\mu)/\rho$ from equation (5.7) as a log-linear model (Semary *et al*., 2025):

$$y=\mu+\rho Z \quad (7)$$

The log-linear model of equation (7) so that the model of *Y* given *X* can be represented by (Semary *et al*., 2025)

$$y_i=\underset{\sim}{\pi}^{T}\underset{\sim}{x}_i+\rho z_i,\qquad i=1,2,...,n \quad (8)$$

where the location parameter in equation (7) is given by $\mu_i=\underset{\sim}{\pi}^{T}\underset{\sim}{x}_i$, $\underset{\sim}{x}_i^{T}=\left(x_{i1},x_{i2},...,x_{ip}\right)$ is the covariate vector and $\underset{\sim}{\pi}=\left(\pi_1,\pi_2,...,\pi_p\right)^{T}$ is a $p\times1$ vector of unknown parameters and $\rho>0$ is an unknown scale parameters (Semary *et al*., 2025). In the model presented in equation (8), the random error z_i has a density function given by equation (5) if $\mu=0$ and $\rho=1$. Therefore, equation (8) is referred to as the Log-NGOF-Et-Weibull survival regression model, with the survival function given as (Semary *et al*., 2025):

$$S(y\mid x)=1-\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\exp\left(\frac{y_i-\underset{\sim}{\pi}^{T}\underset{\sim}{x}_i}{\rho}\right)\right\}\right)^{-\gamma\delta}-1\right)\right)^{\beta}\right\} \quad (9)$$

where $\underset{\sim}{x}$ is the $n\times p$ matrix of exposure variables, and $\underset{\sim}{\pi}$ is the p coefficient vector.

The plots of the density function and survival function for the Log-NGOF-W model are presented in Figures 5 and 6, respectively.

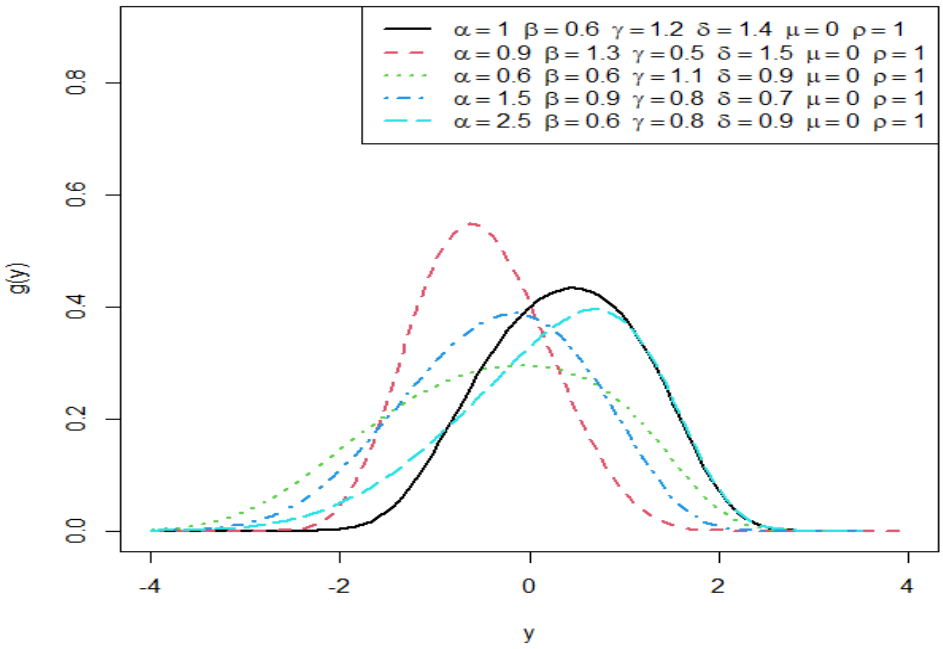

Figure 5: PDF Plot of Log NGOF-Et-Weibull Distribution

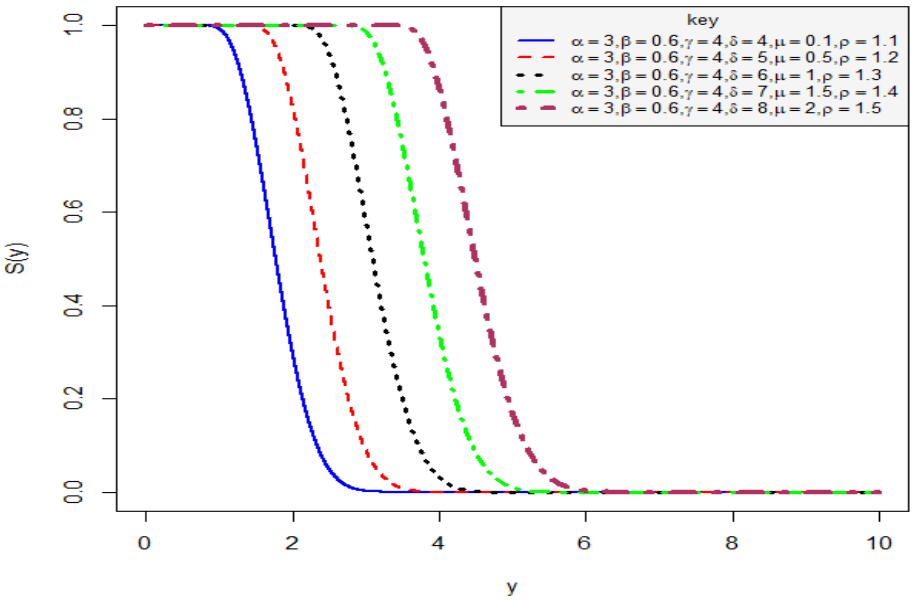

Figure 6: Survival Plot of Log NGOF-Et-Weibull Distribution

Figure 5 shows a plot of the log-NGOF-Et-Weibull distribution as a function of y. The shape of the distribution is determined by four parameters, alpha, beta, gamma and delta. Alpha (α) controls the scale of the distribution. A higher alpha value results in a distribution that is stretched out to the right. Beta (β) controls the shape of the distribution. A higher beta value results in a more peaked distribution. Gamma (γ) controls the location of the distribution. A higher gamma value shifts the distribution to the right. Delta (δ) controls the general shape of the distribution. A higher delta value results in a peaked distribution The y-axis of the plot is labelled "g(y)", which is the probability density function (pdf) of the log-NGOF-Et-Weibull distribution. The pdf expresses how likely it is to observe a particular value of y. The area under the curve between two points represents the probability that y will fall within that range.

The plot in Figure 6 shows the survival function of the log-NGOF-Et-Weibull distribution which can be used to model a wide variety of survival shapes, depending on the values of the parameters. The specific shapes of the distributions in the plot are determined by the values of alpha, beta, gamma, delta, mu and rho that are shown for each curve.

Consider a random sample of size n: $(y_1, d_1, x_1), ..., (y_i, d_i, x_i), ..., (y_n, d_n, x_n)$, where $y_i = \min\left[Log(T_i), Log(C_i)\right]$, d_i is the censoring indicator ($d_i = 0$ if censored and $d_i = 1$ if uncensored) and x_i is the p covariate vector associated with the i[th] subject (Semary *et al*., 2025). We assumed no informative censoring and independence of the observed survival time and censoring time (Semary *et al*., 2025). The log-likelihood function of the model given in equation (8) for the vector of parameter $\Omega = (\alpha, \gamma, \beta, \delta, \rho, \pi^T)^T$ takes the form (Semary *et al*., 2025):

$$l(\Omega; y_i) = \sum_{i \in U} \log \left[\begin{array}{l} \dfrac{\beta\gamma\alpha^{\beta}\delta}{\rho} \exp\left\{\dfrac{y-\mu}{\rho}\right\} \exp\left\{-\exp\left(\dfrac{y-\mu}{\rho}\right)\right\} \\ \times \left(1-\exp\left\{-\exp\left(\dfrac{y-\mu}{\rho}\right)\right\}\right)^{-(\gamma\delta+1)} \left(\left(1-\exp\left\{-\exp\left(\dfrac{y-\mu}{\rho}\right)\right\}\right)^{-\gamma\delta} - 1\right)^{\beta-1} \\ \times \exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\exp\left(\dfrac{y-\mu}{\rho}\right)\right\}\right)^{-\gamma\delta} - 1\right)\right)^{\beta}\right\} \end{array} \right] + \sum_{i \in C} \log \left[1-\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\exp\left(\dfrac{y-\mu}{\rho}\right)\right\}\right)^{-\gamma\delta} - 1\right)\right)^{\beta}\right\}\right] \quad (10)$$

where U denotes the set of uncensored observations, C denotes the set of censored observations by Semary *et al*. (2025).

The examination of residuals is a crucial step following the formulation of the model. It is employed to confirm whether the model assumptions differ significantly. We take into account the Cox and Snell

residuals in this work. The overall fit of the model is examined using these residuals (Semary *et al*., 2025). They can be articulated as (Semary *et al*., 2025).

$$Re_i = -\log\left\{1-\exp\left\{-\left(\alpha\left(\left(1-\exp\left\{-\exp\left(\frac{y_i-\underset{\sim}{\pi}^T\underset{\sim}{x}_i}{\rho}\right)\right\}\right)^{-\gamma\delta}-1\right)^{\beta}\right)\right\}\right\}, \quad i=1,2,...,n \qquad (11)$$

These residuals follow the standard exponential distribution if the fitted regression models are adequate (Semary *et al*., 2025).

Liver Cancer Survival Regression Models

The liver cancer dataset was sourced from Charles (2023). The dataset contains $n=40$ liver cancer patients, where 27.5% of them are censored and 72.5% are uncensored. The response variable y is the natural logarithm of the observed survival time (in weeks); d = censoring (0 = alive at study end or lost to follow-up; 1 = death due to liver cancer); with the following exposure variables: $x1$: the patient's age is less than 35 years; $x2$: the patient's age fall between 35-50 years; $x3$: the patient that is over 50 years; $x4$: the patient is male; $x5$: the patient is a female; $x6$: the patients' Alpha-fetoprotein blood level-1; $x7$: the patients' Alpha-fetoprotein blood level-2; $x8$: the size of the liver cancer stage-1; $x9$: the size of the liver cancer stage-2 and $x10$: the size of the liver cancer stage-3.

The liver cancer survival regression models: log-NGOF-Et-WSR have the form:

$$y_i = \pi_0 + \sum_{k=1}^{10}\pi_k x_{ik} + \rho z_i, \quad i=1,2,...,40 \qquad (12)$$

where $z_{i's}$ are the independent and identically distributed random errors having density function given in equations (5) for the regression models, and $y_i = \log_e(t_i)$ is the natural logarithm of the observed survival times in weeks. The $\underset{\sim}{x}_i^T = (x_{i1}, x_{i2},..., x_{ip})$ is the covariate vector and $\underset{\sim}{\pi} = (\pi_1, \pi_2,..., \pi_p)^T$ is a $p\times 1$ vector of unknown parameters and $\rho > 0$ is an unknown scale parameter.

3. RESULTS

Table 1: MLE of the Parameters for the Fitted Survival Regression Model and the Competing Model to the Liver Cancer Dataset

	Log-NGOFEtWSR		Log-WSR	
Parameters	Parameter Estimates	P-value	Parameter Estimates	P-value
α	2.0049	0.3247	-	-
β	0.4517	0.0060	-	-
γ	2.6418	0.0375	-	-
δ	0.3348	0.0375	-	-
ρ	0.9350	0.0841	0.8011	6.1e-14
π_0	4.9886	2.2e-16	1.7684	2.2e-16
π_1	-3.8789	2.2e-16	3.2462	2.2e-16
π_2	-3.4211	2.2e-16	3.5213	2.2e-16
π_3	-3.6144	2.2e-16	3.0816	2.2e-16
π_4	2.7359	3.63e-13	4.7751	2.2e-16
π_5	3.0275	4.39e-13	5.0320	2.2e-16
π_6	-0.0402	0.9374	-4.6255	2.2e-16
π_7	1.0778	0.0609	-4.2163	2.2e-16
π_8	-1.9897	1.46e-07	-1.8585	2.4e-11
π_9	-1.6974	2.40e-09	-1.9715	2.2e-16
π_{10}	-0.9835	0.0041	-1.3784	2.6e-07
-2LL	105.1809		128.6544	
AIC	137.1809		140.6544	
BIC	164.203		174.9209	
Rank	1		2	

Table 1 presents the results of the fitted log-NGOF-Et-WSR and log-WSR models for the liver cancer dataset. We compared the survival regression models for the cancer data based on the AIC and BIC statistics. The preferred model is the one that gives the lowest AIC and BIC values. The figures in Table 1 indicate that log-NGOF-Et-WSR has a lower value of AIC and BIC than Log-WSR, inferring that, while the former regression model is more appropriate, the new regression model fits the dataset better than the log-WSR model. The results for the best-fitted survival regression model (log-NGOF-Et-WSR) show that there are statistically significant effects of the exposure variables except for Alph-fetoprotein blood level-2, indicating that a patient will survive beyond time t (at the 5% significance level).

Table 2: Goodness-of-Fit Comparison

Criterion	**Log-NGOFEtWSR**	**Log-WSR**	**Interpretation**
−2LL	105.18	128.65	Lower is better → Log-NGOFEtWSR fits better
AIC	137.18	140.65	Lower AIC → Log-NGOFEtWSR is preferred
BIC	164.20	174.92	Lower BIC → Log-NGOFEtWSR still preferred
Rank	1	2	The best model is Log-NGOFEtWSR

Table 2, the Log-NGOFEtWSR model outperforms the standard Log-Weibull model in all criteria (−2LL, AIC, BIC). This suggests it provides a better fit to the liver cancer time-to-event data, likely due to its greater flexibility in capturing complex hazard shapes.

Table 3: Interpretation of Distribution Parameters (Log-NGOFEtWSR)

Parameter	Estimate	P-value	Meaning
α	2.0049	0.3247	Shape parameter – not statistically significant
β	0.4517	0.0060	Shape parameter – significant; affects hazard rate shape.
γ	2.6418	0.0375	Tuning shape – significant
δ	0.3348	0.0375	Tuning shape – significant
ρ	0.9350	0.0841	Shape – marginal significance

Table 3, several shape parameters (β, γ, δ) are statistically significant, indicating that the NGOF-Et-Weibull distribution can flexibly adapt to various hazard rate shapes (increasing, decreasing, bathtub, etc.), which is important in cancer survival modelling.

Table 4: Interpretation of Regression Coefficients (Covariates $\pi_0 - \pi_{10}$)

Covariate	Estimate	P-value	Interpretation
π_0 (Intercept)	4.9886	2.2e-16	Baseline log-survival time (highly significant)
π_1 (x_1: Age < 35)	−3.8789	2.2e-16	Significantly shorter survival than baseline (older patients)
π_2 (x_2: Age 35–50)	−3.4211	2.2e-16	Also significantly lower survival than the older group
π_3 (x_3: Age > 50)	−3.6144	2.2e-16	Reduced survival but close to group effect
π_4 (x_4: Male)	2.7359	3.63e-13	Males have a higher survival time than females.
π_5 (x_5: Female)	3.0275	4.39e-13	Also shows high survival, but the comparison should be based on which is a reference.
π_6 (x_6: AFP level-1)	−0.0402	0.9374	Not significant – does not affect survival
π_7 (x_7: AFP level-2)	1.0778	0.0609	Marginally significant – possibly increases survival.
π_8 (x_8: Stage-1 tumor size)	−1.9897	1.46e-07	Larger size → significantly lower survival
π_9 (x_9: Stage-2 tumor size)	−1.6974	2.4e-09	Also significantly lowers survival
π_{10} (x_{10}: Stage-3 tumor size)	−0.9835	0.0041	Reduced survival – significant

From Table 4, we deduced that:

i. Age is a strong prognostic factor: younger patients (especially < 35) have significantly shorter survival.
ii. Tumor stage (size) is also a critical factor: all stages show reduced survival, most notably in Stages 1 and 2.

iii. AFP level-1 is not significant, but AFP level-2 shows potential relevance.
iv. Gender (π_4 and π_5): Both coefficients are significant, but since both males and females have positive values, female is the reference category.

Cox-Snell Residuals Analysis for Liver Cancer

The Cox-Snell residual plots for liver cancer survival regression models (log-NGOF-Et-WSR and log-WSR) are shown respectively in Figures 7 and 8. The M.L.E of the parameters for the two fitted survival regression models to the liver cancer data's Cox-Snell residuals are presented in Table 5.

Table 5: MLE of the Parameters for the three Fitted Survival Regression Models to the Cox-Snell Residuals for the Liver Cancer Dataset

	Log-NGOFEtWSR		Log-WSR	
Parameters	Parameter Estimates	P-value	Parameter Estimates	P-value
α	0.0418	0.1485	-	-
β	2.2223	4.59e-06	-	-
γ	4.2800	0.0002	-	-
δ	1.0110	0.0002	-	-
ρ	6.8916	0.1641	1.0406	7.1e-11
π_0	1.2531	0.1110	4.6931	2.2e-16
π_1	-2.8001	1.50e-13	-5.3898	2.2e-16
π_2	-1.1889	0.0003	-3.8642	2.2e-16
π_3	-2.7696	1.72e-13	-5.0189	2.2e-16
π_4	3.3782	1.96e-14	0.6885	0.0689
π_5	3.7034	1.00e-13	-0.4194	0.3462
π_6	0.4331	0.3120	-0.2907	0.4100
π_7	0.9709	0.0447	-0.2347	0.5649
π_8	-1.1538	0.0020	-0.6140	0.0795
π_9	-0.6081	0.0731	-0.5250	0.0633
π_{10}	-1.4052	0.0003	-0.2972	0.4229
-2LL	98.62261		118.5995	
AIC	130.6226		142.5995	
BIC	157.6447		162.8661	
Rank	1		2	

Table 6: Model Fit Comparison (Based on Residuals)

Criterion	**Log-NGOFEtWSR**	**Log-WSR**	**Interpretation**
−2LL	98.62	118.60	Lower is better → Log-NGOFEtWSR is better
AIC	130.62	142.60	Log-NGOFEtWSR preferred
BIC	157.64	162.87	Log-NGOFEtWSR better
Rank	1	2	Log-NGOFEtWSR remains top-ranked

Table 6 revealed that, even based on **Cox-Snell residuals**, the **Log-NGOFEtWSR** model has a superior fit, reinforcing its reliability and robustness in capturing survival dynamics.

Table 7: Distribution Parameters Interpretation (Log-NGOFEtWSR)

Parameter	Estimate	P-value	Interpretation
A	0.0418	0.1485	Not significant
B	2.2223	4.59e-06	Significant – affects the shape of the hazard.
Γ	4.2800	0.0002	Significant–tail behaviour
Δ	1.0110	0.0002	Significant
P	6.8916	0.1641	Not significant

Table 7, the shape parameters β, γ, and δ are highly significant, suggesting they are essential in capturing the complex distribution of Cox-Snell residuals. This again demonstrates the flexibility and suitability of the Log-NGOFEtWSR model.

Table 8: Covariate Effects from Residual Model (Log-NGOFEtWSR)

Covariate	Estimate	P-value	Interpretation
π_0 (Intercept)	1.2531	0.1110	Not significant
π_1 (x_1: Age < 35)	−2.8001	1.50e-13	Significantly reduces residual survival
π_2 (x_2: Age 35–50)	−1.1889	0.0003	Reduces survival – significant
π_3 (x_3: Age > 50)	−2.7696	1.72e-13	Strong negative effect
π_4 (x_4: Male)	3.3782	1.96e-14	Strong positive effect on survival
π_5 (x_5: Female)	3.7034	1.00e-13	Also strong positive impact
π_6 (x_6: AFP level-1)	0.4331	0.3120	Not significant
π_7 (x_7: AFP level-2)	0.9709	0.0447	Marginally significant
π_8 (x_8: Stage-1 tumor size)	−1.1538	0.0020	Reduces survival significantly
π_9 (x_9: Stage-2 tumor size)	−0.6081	0.0731	Marginally significant negative effect
π_{10} (x_{10}: Stage-3 tumor size)	−1.4052	0.0003	Significant negative effect

From Table 8, we deduced that:

i. Age remains a significant predictor: younger age groups have worse survival.
ii. The tumour stage (x_8–x_{10}) shows a clear and consistent negative impact on survival.
iii. AFP level-2 (x_7) might be mildly protective, but AFP level-1 (x_6) is not significant.
iv. Gender (x_4, x_5) again shows a strong positive effect, supporting previous findings.

Based on both the original model fit and the Cox-Snell residuals, the Log-NGOFEtWSR survival regression model outperforms the Log-WSR model. It not only provides a superior fit as indicated by AIC, BIC, and −2LL but also captures significant effects from multiple covariates, including age, tumour stage, and gender. The shape parameters β, γ, and δ are significant, confirming the model's flexibility in handling complex survival patterns observed in liver cancer patients.

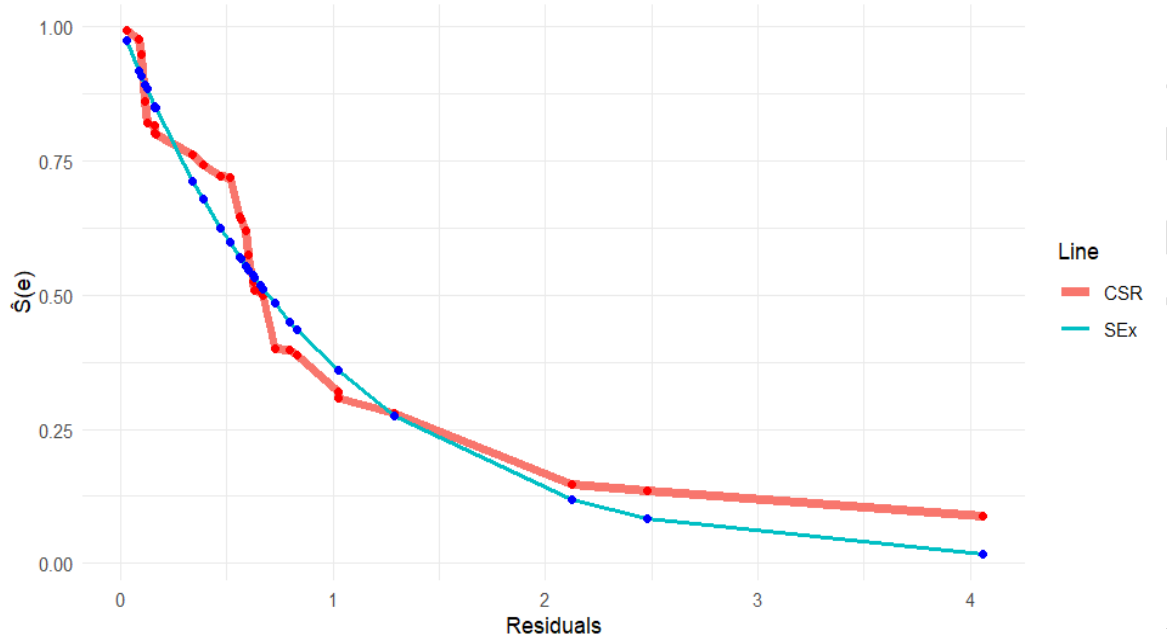

Figure 7: Cox-Snell Residuals Plots for Liver Cancer using Log-NGOF-Et-WSR

Using the Kolmogrov-Smirnov one sample test, the D_n statistic returns a value of 0.121 for Figure 7 which is less than the table value of 0.253 at the 5% level of significance indicating that the Cox-Snell residuals for the Log-NGOF-Et-WSR model follow the standard exponential distribution.

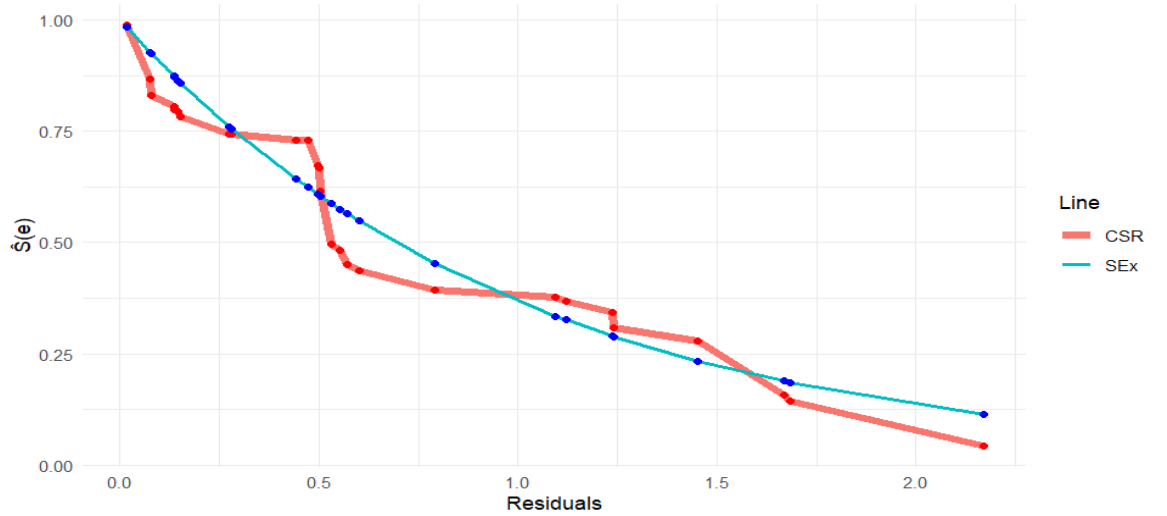

Figure 8: Cox-Snell Residuals Plots for Liver Cancer using Log-WSR

Also, Using the Kolmogrov-Smirnov one sample test, the D_n statistic returns a value of 0.1209 for Figure 8 which is less than the table value of 0.253 at the 5% level of significance indicating that the Cox-Snell residuals for the Log-WSR model follow the standard exponential distribution.

Figures 7 and 8 show the Cox-Snell residuals plots for the Log-NGOF-Et-WSR and Log-WSR models. Based on these plots, the estimated standard exponential curve is closer to the theoretical survival curve for the Log-NGOF-Et-WSR than the Log-WSR models. Even though the computed D_n statistic for the two models indicated that they adequately fit the liver cancer data, the best model (Log-NGOF-Et-WSR) should be used for inference.

Liver Cancer Kaplan-Meier Survival Probability

The results of the Kaplan-Meier survival probability analysis for the liver cancer dataset are presented in Table 9, while the 95% confidence interval of the survival probability is presented in Figure 9. The best-fitted survival regression model is also plotted together with the non-parametric Kaplan-Meier survival probability in Figure 10.

Table 11: Kaplan-Meier Survival Probability for Liver Cancer Dataset

Time (in Weeks)	Number at Risk	Number of Events	Survival Prob.	Standard Error	95% CI Lower	95% CI Upper
2	40	2	0.950	0.0345	0.8848	1.000
6	38	4	0.850	0.0565	0.7462	0.968
7	34	2	0.800	0.0632	0.6852	0.934
9	31	1	0.774	0.0663	0.6546	0.916
11	30	3	0.697	0.0732	0.5672	0.856
13	25	1	0.669	0.0754	0.5364	0.834
14	23	1	0.640	0.0775	0.5046	0.811
15	22	1	0.611	0.0792	0.4736	0.788
16	21	1	0.582	0.0806	0.4433	0.763
17	19	2	0.520	0.083	0.3808	0.711
18	17	1	0.490	0.0835	0.3506	0.684
19	16	2	0.429	0.0836	0.2925	0.628
37	12	1	0.393	0.0839	0.2585	0.597
41	11	2	0.321	0.0824	0.1944	0.531
51	8	1	0.281	0.0813	0.1596	0.496
52	7	1	0.241	0.079	0.1268	0.458
67	4	2	0.121	0.0721	0.0373	0.389
80	1	1	0.000	-	-	-

Table 9 shows the Kaplan-Meier survival probabilities for liver cancer patients. The survival probability decreased steadily over time, with a notable decline from 0.950 at week 2 to 0.241 by week 52. The median survival time was estimated to occur between 17 and 18 weeks, with survival falling below 50% at this point. Confidence intervals widened in later weeks, reflecting the reduced number at risk. This pattern indicates the severity of liver cancer and highlights the need for prompt intervention strategies.

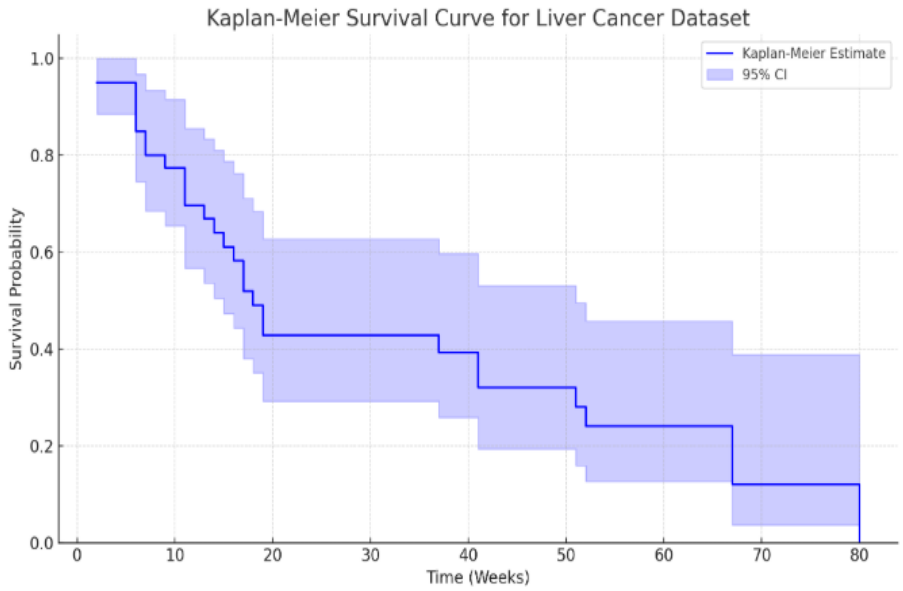

Figure 9: Liver Cancer Kaplan-Meier Survival Curve

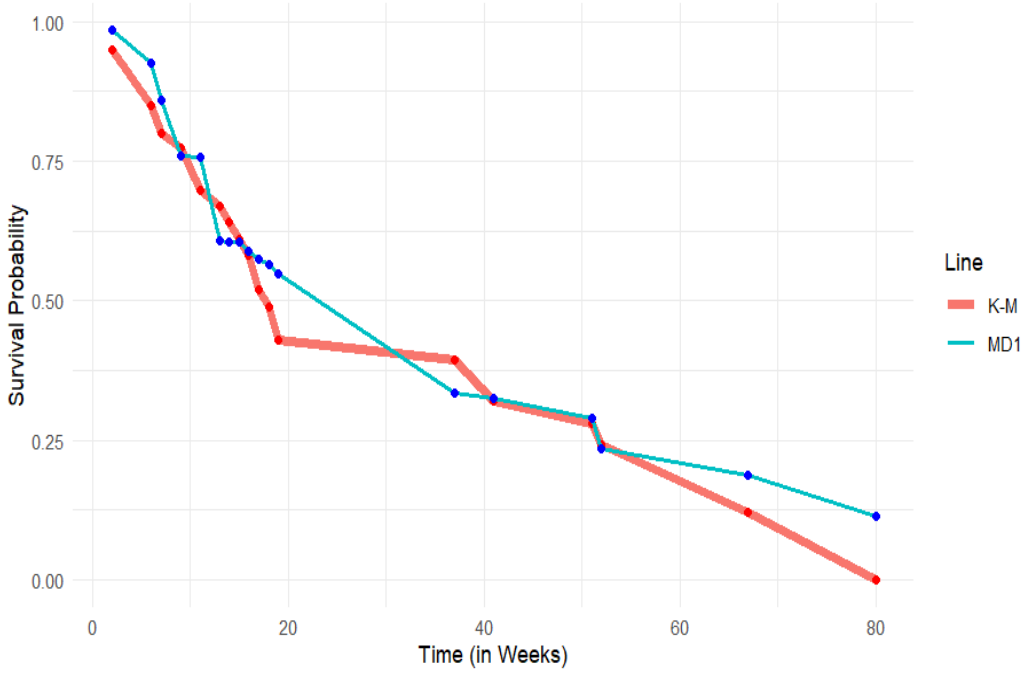

Figure 10: Liver Cancer Kaplan-Meier and Log-NGOF-Et-WSR Survival Curve

The plot in Figure 10 shows the best-fitted survival regression model together with the non-parametric Kaplan-Meier survival probability. The effects of the covariates play a significant role in describing the distribution of the event times and the existence of the association between the explanatory variables and

the event-time distribution. This made the best-fitted model more flexible compared to the Kaplan-Meier survival probability model.

4. Conclusion

The results of this study demonstrate that the Log-NGOF-Et-Weibull Survival Regression (Log-NGOFEtWSR) model is a highly effective tool for analyzing liver cancer survival data. Compared to the conventional Log-Weibull Survival Regression (Log-WSR) model, the proposed model provided a better fit to the data, as evidenced by lower values of -2 log-likelihood (105.18 vs. 128.65), Akaike Information Criterion (137.18 vs. 140.65), and Bayesian Information Criterion (164.20 vs. 174.92). Furthermore, the model adequately captured the variability in survival times and identified key prognostic factors, including age, gender, and tumor stage, as statistically significant predictors of liver cancer patient survival. Cox-Snell residual diagnostics confirmed the model's validity, with superior residual performance metrics (AIC = 130.62, BIC = 157.64) supporting the robustness of the Log-NGOFEtWSR model. The survival curve generated from the fitted model closely matched the non-parametric Kaplan-Meier survival estimates, indicating strong agreement between model-based and empirical survival probabilities. The flexibility of the NGOF-Et-Weibull distribution in modeling complex hazard structures, such as increasing, decreasing, and bathtub-shaped hazards, makes it especially suitable for medical survival data like liver cancer, where heterogeneous patient risk profiles exist. In summary, the Log-NGOF-Et-Weibull model not only enhances predictive accuracy but also provides deeper insights into patient risk stratification, offering a valuable framework for clinicians, researchers, and policy-makers in cancer prognosis and treatment planning. Future studies may explore its applicability to larger datasets or extend its use to other types of cancers and chronic diseases.

References

Ahmad, I., Gulzar, M., Shahzad, F., Yaqub, M., & Zhoor, T. (2013). Quality assessment of yoghurt produced at large (industrial) and small scale. *The Journal of Animal and Plant Sciences*, 23(1), 58-61.

Ahmed, S. A., Mostafa, A. H. M., El-Sherbini, M., & Abdelkhalek, A. (2022). Assessment of Microbial Safety and Quality of Market Raw Milk and Pasteurised Milk Sold in Dakahlia Governorate, Egypt. *Journal of Advanced Veterinary Research*, 12(4), 456-461.

Akor, A., Sadiq, I. A., Usman, A., Doguwa, S. I., & Akor, L. O. (2025). Advanced Survival Modeling of Tuberculosis Patients: Insights from Exponential and Weibull AFT Models. *Fudma Journal of Sciences*, *9*(3), 169-182. https://doi.org/10.33003/fjs-2025-0903-3350

Anwanwan, D., Singh, S. K., Singh, S., Saikam, V., & Singh, R. (2020). Challenges in liver cancer and possible treatment approaches. *Biochimica et biophysica acta (BBA)-Reviews on cancer*, *1873*(1), 188314.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. https://doi.org/10.3322/caac.21660

Charles, Z. (2023). https://real-statistics.com/survival-analysis/cox-regression/real-statistics-capabilities-for-cox-regression/survival-analysis-study-data/ 30/04/2024

Collett, D. (2015). *Modelling Survival Data in Medical Research* (3rd ed.). Chapman and Hall/CRC.

Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., ... & Balasubramanian, S. (2022). Early detection of cancer. *Science*, *375*(6586), eaay9040.

Hong, M. K., & Ding, D. C. (2025). Early diagnosis of ovarian cancer: A comprehensive review of the advances, challenges, and future directions. *Diagnostics*, *15*(4), 406.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data* (2nd ed.). Springer.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd ed.). Wiley-Interscience.

Li, Q., Ding, C., Cao, M., Yang, F., Yan, X., He, S., ... & Chen, W. (2024). Global epidemiology of liver cancer 2022: An emphasis on geographic disparities. *Chinese medical journal*, *137*(19), 2334-2342.

Mamun, T. I., Younus, S., & Rahman, M. H. (2024). Gastric cancer-Epidemiology, modifiable and non-modifiable risk factors, challenges and opportunities: an updated review. *Cancer Treatment and Research Communications*, *41*, 100845.

Marrero, J. A., Fontana, R. J., Su, G. L., Conjeevaram, H. S., Emick, D. M., Lok, A. S. F. (2005). Alpha-fetoprotein, Des-γ-carboxy prothrombin, and Lectin-bound AFP in Early Detection of Hepatocellular Carcinoma. *Gastroenterology*, 129(1), 145–150. https://doi.org/10.1053/j.gastro.2005.04.008

Mathur, R., Sharma, R. S., Saini, G., Shukla, S. P., & Jha, A. K. (2025). Advancing Oral Cancer Detection and Prevention: Integrating Innovative Screening Tools, Molecular Insights, and Targeted Therapies. *Asian Pacific Journal of Cancer Biology*, *10*(2), 507-514.

Murphy, K. P. (2007). *Conjugate Bayesian analysis of the Gaussian distribution*. Technical Report, MIT.

Sadiq, I. A, Doguwa, S. I. S., Yahaya, A., & Usman, A. (2023). Development of New Generalized Odd Fréchet-Exponentiated-G Family of Distribution. *UMYU Scientifica*, *2*(4), 169-178. https://doi.org/10.56919/usci.2324.021

Sadiq, I. A, Doguwa, S. I. S., Yahaya, A., & Usman, A. (2023). Development of Fréchet Families of Distributions: Survival Regression Models and Applications. PhD. Thesis Department of Statistics, ABU Zaria.

Sadiq, I. A., Zakari, Y., Doguwa, S. I., Isiya, M., Suleiman, S. Y., Ajayi, A. H., ... & Samuel, W. (2025). Survival Analysis of Acute Myocardial Infarction Patients Using the Kumaraswamy-Logistic Model and Kaplan-Meier Estimation. *Iraqi Statisticians journal*, 15-32. https://doi.org/10.62933/32m8tn21

Semary, H., Sadiq, I. A., Doguwa, S. I. S., Ishaq, A. I., Suleiman, A. A., Daud, H., & Abd Elgawad, M. A. (2025). Advancing survival regression using the NGOF exponentiated Weibull distribution

for vesicovaginal fistula and radiation data applications. *Journal of Radiation Research and Applied Sciences*, *18*(2), 101497.

Usman , . A. ., Doguwa, S. I. ., Sadiq, I. A. ., & Akor, A. (2025). Exploring Accelerated Failure Time Models for Tuberculosis Survival: Loglogistic and Weibull Survival Regression Model. *Journal of Science Research and Reviews*, *2*(1), 27-36. https://doi.org/10.70882/josrar.2025.v2i1.27

Usman, A., Doguwa, S. I., Sadiq, I. A., & Akor, A. (2025). Exploring Accelerated Failure Time Models for Tuberculosis Survival: Loglogistic and Weibull Survival Regression Model. *Journal of Science Research and Reviews*, *2*(1), 27-36.