



# New Modified Exponentiated Weibull Distribution: A Survival Analysis

A M. Rangoli <sup>1</sup>, A S. Talawar <sup>2</sup>, R P. Agadi <sup>3</sup>, Vijaya Sorganvi <sup>1</sup>

Received 12/07/2024

Review began 12/26/2024

Review ended 01/05/2025

Published 01/12/2025

© Copyright 2025

Rangoli et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.77347

1. Department of Community Medicine, Shri B M Patil Medical College Hospital and Research Centre, Vijayapura, IND  
2. Department of Statistics, Karnatak University Dharwad, Dharwad, IND 3. Department of Community Medicine, Shridevi Institute of Medical Sciences and Research Hospital, Tumakuru, IND

Corresponding author: A M. Rangoli, ajmr1008@gmail.com

## Abstract

### Introduction

In survival analysis, various lifetime distributions are used to model hazard and survival functions. This study introduces a modified Weibull distribution capable of exhibiting increasing, decreasing, constant, and bathtub-shaped hazard rates. The distribution's flexibility allows it to better capture different density patterns like bimodal observed in real-world data, especially in medical settings.

### Methodology

The proposed distribution's properties, including its hazard and survival functions, are explored in detail. Data from hospital records was used to validate the model. Parameters are estimated via the expectation-maximization (EM) algorithm, with standard errors and confidence intervals calculated. A comparison is drawn with other modified Weibull models to assess the performance.

### Results

The model demonstrates a good fit for the hospital dataset, providing a robust estimation of survival probabilities across different time periods. The EM algorithm ensures precise parameter estimation, and the results show the model's capability to capture varying hazard patterns effectively. Kaplan-Meier survival curves are plotted and compared with the survival curve from the proposed model, showing strong alignment.

### Conclusion

The modified Weibull distribution introduced in this study offers a versatile tool for modeling different hazard rate patterns. The model's strong performance, validated through real hospital data, suggests it could be a valuable addition to survival analysis, outperforming other modified Weibull models in terms of fit and flexibility.

**Categories:** Other, Public Health, Epidemiology/Public Health

**Keywords:** em algorithm, hazard, information criterion, kaplan-meier survival curves, modified weibull distribution, survival

## Introduction

In recent years, there has been a surge in interest in survival analysis, driven by the increasing complexity of survival data in clinical trials, public health studies, and industrial reliability assessments. These trends emphasize the need for accurate models of survival probabilities, which may decline over time in certain contexts. By incorporating time-to-event data and censoring, survival analysis offers a robust framework for analyzing factors that influence survival duration across populations.

Accurate modeling of failure rates is crucial, requiring life distributions that capture the characteristics of these rates. Over the years, various life distributions have been developed or modified to account for different hazard rate patterns. Among these, the Weibull distribution (WD) is widely recognized in reliability analysis due to its flexibility in modeling diverse failure rate behaviors. However, to better accommodate the complexities of real-world data, researchers have introduced extensions, exponentiated versions, and modifications of the Weibull distribution. These modifications introduce additional parameters to capture a broader range of failure behaviors, as seen in studies [1-11].

In this paper, we propose a new modified Weibull distribution with four parameters and estimate its hazard and survival functions. This new distribution is particularly useful for modeling bimodal densities and non-monotonic hazard functions.

#### How to cite this article

Rangoli A M, Talawar A S, Agadi R P, et al. (January 12, 2025) New Modified Exponentiated Weibull Distribution: A Survival Analysis. Cureus 17(1): e77347. DOI 10.7759/cureus.77347

## Materials And Methods

### Data description

Data was collected from Shri B. M. Patil Medical Hospital and Research Center in Vijayapura, India, detailing the failure times of patients admitted during the study period. The data spans from January 1, 2023, to December 31, 2023, and includes 936 patients who died due to various causes during this time frame.

#### Model Formulation

Let  $T_1, T_2, \dots, T_n$  are failure times, we assume that  $T_i$  to follows modified Weibull random variables with parameters  $(\alpha, \beta, \gamma, \lambda)$ ,  $i=1, 2, \dots, n$ .

Let  $f(t)$ ,  $F(t)$ ,  $S(t)$  and  $\lambda(t)$  be density, cumulative distribution, survival and hazard functions of the new modified Weibull distribution.

Let the cumulative distribution function (CDF) is,  $F(t) = 1 - e^{-\left(\frac{\alpha}{\beta}\right)^{\lambda} t^{\beta}} \left( e^{t^{\beta}} - 1 \right)^{\gamma} + \gamma t^{\lambda} \right)^{\alpha}$  where  $\alpha > 0, \beta > 0, \gamma \geq 0, \lambda > 0, t \geq 0, \beta$  and  $\lambda$  are shape parameters,  $\alpha$  and  $\gamma$  are scale parameters.

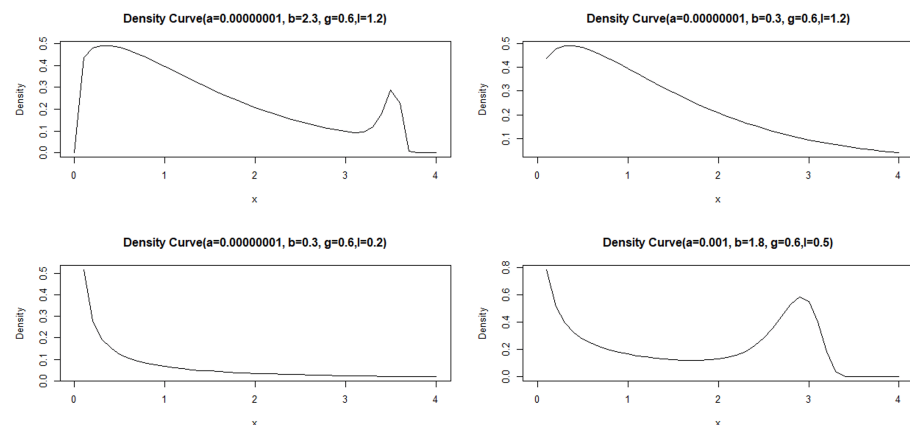
The probability density function (PDF) is,  $f(t) = \left( \frac{\alpha}{\beta} \right)^{\lambda} t^{\beta-1} e^{-\left(\frac{\alpha}{\beta}\right)^{\lambda} t^{\beta}} \left( e^{t^{\beta}} - 1 \right)^{\gamma} + \gamma t^{\lambda} \right)^{\alpha-1} e^{-\left(\frac{\alpha}{\beta}\right)^{\lambda} t^{\beta}} \left( e^{t^{\beta}} - 1 \right)^{\gamma} + \gamma t^{\lambda} \right)^{\alpha} \left( e^{t^{\beta}} - 1 \right)^{\gamma-1} e^{t^{\beta}} + \gamma t^{\lambda} \right)^{\alpha}$

Hazard function is,  $\lambda(t) = \left( \frac{\alpha}{\beta} \right)^{\lambda} t^{\beta-1} e^{-\left(\frac{\alpha}{\beta}\right)^{\lambda} t^{\beta}} \left( e^{t^{\beta}} - 1 \right)^{\gamma} + \gamma t^{\lambda} \right)^{\alpha-1} e^{-\left(\frac{\alpha}{\beta}\right)^{\lambda} t^{\beta}} \left( e^{t^{\beta}} - 1 \right)^{\gamma} + \gamma t^{\lambda} \right)^{\alpha} \left( e^{t^{\beta}} - 1 \right)^{\gamma-1} e^{t^{\beta}} + \gamma t^{\lambda} \right)^{\alpha}$  Survival function is,  $S(t) = e^{-\left(\frac{\alpha}{\beta}\right)^{\lambda} t^{\beta}} \left( e^{t^{\beta}} - 1 \right)^{\gamma} + \gamma t^{\lambda} \right)^{\alpha}$

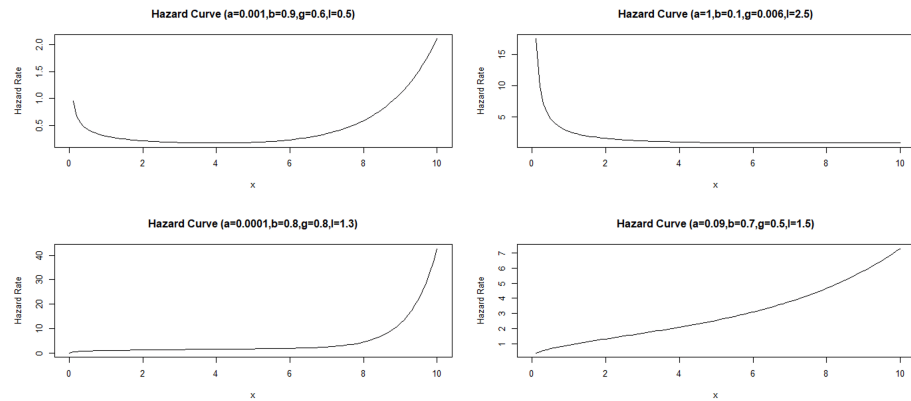
This modified Weibull distribution is widely used in various fields such as reliability analysis, human mortality studies, and survival analysis. Its versatility in modeling different types of failure patterns and hazard rates makes it an invaluable tool for understanding time-to-event data in both industrial and healthcare contexts.

#### Motivation for the Modified Weibull Distribution

Here are two key motivations for the proposed modified Weibull distribution with four parameters: the proposed distribution offers a wide range of shapes, including both unimodal and bimodal density patterns (Figure 1), making it highly applicable to real-life scenarios; the new probabilistic model exhibits various hazard rate function shapes, such as bathtub, decreasing, constant-increasing, and increasing hazard rates (Figure 2). This flexibility makes the modified Weibull distribution particularly useful for fitting diverse datasets with different hazard rate shapes.



**FIGURE 1: Density curves for different values of parameters.**



**FIGURE 2: Hazard curves for different values of parameters.**

## Results

### Some properties

#### Mean

The mean of the random variable  $T$  with pdf  $f(t)$  is given by,

$$E(T) = \int_0^{\infty} t f(t) dt$$

$$E(T) = \int_0^{\infty} t \left( \frac{\alpha}{\beta} \left( \frac{\alpha}{\beta} \right)^{\beta-1} \left( \frac{\alpha}{\beta} \right)^{\beta-1} e^{-\left( \frac{\alpha}{\beta} \right)^{\beta} t^{\beta}} + \gamma \left( \frac{\alpha}{\beta} \right)^{\beta-1} e^{-\left( \frac{\alpha}{\beta} \right)^{\beta} t^{\beta}} \right) dt$$

This integral does not have a simple closed form, so numerical method should be used to estimate mean. The mean represents the average time until the event occurs.

#### Variance

The variance measures the spread of the time to failure.

$$V(T) = E(T^2) - (E(T))^2$$

$$E(T^2) = \int_0^{\infty} t^2 f(t) dt$$

$$E(T^2) = \int_0^{\infty} t^2 \left( \frac{\alpha}{\beta} \left( \frac{\alpha}{\beta} \right)^{\beta-1} \left( \frac{\alpha}{\beta} \right)^{\beta-1} e^{-\left( \frac{\alpha}{\beta} \right)^{\beta} t^{\beta}} + \gamma \left( \frac{\alpha}{\beta} \right)^{\beta-1} e^{-\left( \frac{\alpha}{\beta} \right)^{\beta} t^{\beta}} \right) dt$$

This is also computed using numerical method.

#### Moment Generating Function (MGF)

The moment is given by,  $E(T^r)$  that is,  $E(T^r) = \int_0^{\infty} t^r f(t) dt$

The MGF is defined as,  $M_T(s) = E(e^{sT}) = \int_0^{\infty} e^{st} f(t) dt$

$$M_T(s) = \int_0^{\infty} e^{st} \left( \frac{\alpha}{\beta} \left( \frac{\alpha}{\beta} \right)^{\beta-1} \left( \frac{\alpha}{\beta} \right)^{\beta-1} e^{-\left( \frac{\alpha}{\beta} \right)^{\beta} t^{\beta}} + \gamma \left( \frac{\alpha}{\beta} \right)^{\beta-1} e^{-\left( \frac{\alpha}{\beta} \right)^{\beta} t^{\beta}} \right) dt$$

The MGF provides a way to compute all moments of the distribution by taking derivatives,

$$E(T^r) = \left. \frac{d^r M_T(s)}{ds^r} \right|_{s=0}$$

Here if  $r=1$  we get first moment (Mean,  $E(T)$ ), and  $r=2$  we get second moment ( $E(T^2)$ ).

### Cumulative Hazard Function

The cumulative hazard function is given by,

$$H(t) = \int_0^t \lambda(u) du = \int_0^t \frac{\alpha}{\beta} u^{\beta-1} e^{u^\beta} + \gamma \lambda(u^\beta) du$$

$$H(t) = \frac{\alpha}{\beta} u^\beta + \gamma \int_0^t u^{\beta-1} e^{u^\beta} du + \gamma \lambda(u^\beta) \int_0^t u^{\beta-1} du$$

$$H(t) = \frac{\alpha}{\beta} u^\beta + \gamma (e^{t^\beta} - 1) + \gamma t^\beta \lambda(t^\beta) \text{ So, we write } S(t) = e^{-H(t)}.$$

### Quantile Function

The quantile function is the inverse of the CDF. For a given probability p the quantile function Q(p) gives the value at t such that probability of failure by that time is p,

$$Q(p) = F^{-1}(p)$$

$$F(t) = p$$

$$1 - e^{-\left(\frac{\alpha}{\beta} u^\beta + \gamma (e^{u^\beta} - 1) + \gamma t^\beta \lambda(t^\beta)\right)} = p$$

$$\frac{\alpha}{\beta} u^\beta + \gamma (e^{u^\beta} - 1) + \gamma t^\beta \lambda(t^\beta) = -\ln(1 - p)$$

This equation can be solved using numerical method. If we substitute p=0.25, 0.5, 0.75, we get first, second (median) and third quartiles respectively.

### Order Statistics

Let  $T_1, T_2, \dots, T_n$  be a random sample of size n from modified Weibull distribution with PDF and CDF defined in (2) and (1) respectively. We give the density of  $r^{th}$  order statistic as,  $f_r(t) = \frac{n!}{(r-1)!(n-r)!} \left(1 - e^{-\left(\frac{\alpha}{\beta} t^\beta + \gamma (e^{t^\beta} - 1) + \gamma t^\beta \lambda(t^\beta)\right)}\right)^{r-1} \left(e^{-\left(\frac{\alpha}{\beta} t^\beta + \gamma (e^{t^\beta} - 1) + \gamma t^\beta \lambda(t^\beta)\right)}\right)^{n-r+1} \left(\frac{\alpha}{\beta} t^{\beta-1} e^{t^\beta} + \gamma \lambda(t^\beta)\right)$

The density of  $1^{st}$  order statistic is given by,  $f_1(t) = n \left(1 - e^{-\left(\frac{\alpha}{\beta} t^\beta + \gamma (e^{t^\beta} - 1) + \gamma t^\beta \lambda(t^\beta)\right)}\right)^{n-1} \left(e^{-\left(\frac{\alpha}{\beta} t^\beta + \gamma (e^{t^\beta} - 1) + \gamma t^\beta \lambda(t^\beta)\right)}\right) \left(\frac{\alpha}{\beta} t^{\beta-1} e^{t^\beta} + \gamma \lambda(t^\beta)\right)$

And the density of  $n^{th}$  order statistic is given by,  $f_n(t) = n \left(1 - e^{-\left(\frac{\alpha}{\beta} t^\beta + \gamma (e^{t^\beta} - 1) + \gamma t^\beta \lambda(t^\beta)\right)}\right)^{n-1} \left(e^{-\left(\frac{\alpha}{\beta} t^\beta + \gamma (e^{t^\beta} - 1) + \gamma t^\beta \lambda(t^\beta)\right)}\right) \left(\frac{\alpha}{\beta} t^{\beta-1} e^{t^\beta} + \gamma \lambda(t^\beta)\right)$

## Parameter estimation

### Maximum Likelihood Estimation

The maximum likelihood estimation (MLE) is widely used among the statistical inference because of its desirable properties like consistency, asymptotic efficiency. Use of maximum likelihood estimation is illustrated below.

The likelihood function is given by,  $L = \prod_{i=1}^n f(t_i)$  The log likelihood can be written as,

$$\log L = \sum_{i=1}^n \log f(t_i)$$

$$\log L = \sum_{i=1}^n \left\{ \log \left( \frac{\alpha}{\beta} t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda(t_i^\beta) \right) - \sum_{i=1}^n \left( \frac{\alpha}{\beta} t_i^\beta + \gamma (e^{t_i^\beta} - 1) + \gamma t_i^\beta \lambda(t_i^\beta) \right) \right\}$$

By partially differentiating equation (6) with respect to each parameter and setting the resulting expressions to zero, we obtain the parameter estimates. The first-order partial derivatives are provided in Appendix 1. However, since the partial differentiation does not yield an explicit solution, we consider using either a numerical method, such as the Newton-Raphson method, or the expectation-maximization (EM) algorithm.

#### Expectation-Maximization (EM) Algorithm

The expectation-maximization (EM) algorithm is a widely used iterative method for finding maximum likelihood estimates of parameters in probabilistic models, especially when the data is incomplete or has missing values [12]. Here's how the EM algorithm works:

#### Steps of the EM Algorithm

Initialization: Start with initial guesses for the parameters  $\hat{\theta}^{(0)}$

Expectation step (E-step): Compute the expected value of the log-likelihood function, with respect to the unknown (latent) variables, using the current estimate of the parameters  $\hat{\theta}^{(t)}$ . This step calculates the "responsibilities" or the probability of the latent variables, given the observed data and current parameters.

Formally, this is the expected value of the complete-data log-likelihood

$$Q(\theta \mid \hat{\theta}^{(t)}) = \mathbb{E}[\log p(X, Z \mid \theta) \mid X, \hat{\theta}^{(t)}]$$

where  $X$  represents the observed data,  $Z$  represents the unobserved data (latent variables or missing data), and  $\theta$  are the parameters.

Maximization step (M-step): Maximize the expected log-likelihood function obtained from the E-step with respect to the parameters  $\theta$ . This gives the updated parameter estimates.

Formally

$$\hat{\theta}^{(t+1)} = \arg \max_{\theta} Q(\theta \mid \hat{\theta}^{(t)})$$

The goal is to find the parameters that maximize the expected log-likelihood.

Iteration: Repeat the E-step and M-step until convergence, i.e., until the parameters  $\hat{\theta}^{(t)}$  stabilize and the improvement in the likelihood becomes negligible.

#### Asymptotic Confidence Bounds

When maximum likelihood estimates (MLEs) do not have a closed form, making it difficult to determine the parameter distribution directly, we rely on the asymptotic distribution of the MLEs to calculate confidence intervals [13]. The asymptotic distribution of the MLE  $\hat{\theta}$  is given by,  $\left( \hat{\theta} - \theta \right) \rightarrow N_{(4)}(0, I^{-1}(\theta))$  Where  $I(\theta) \rightarrow$  Fisher information matrix of the unknown parameters  $\theta = (\alpha, \beta, \gamma, \lambda)$ .

The elements of the 4 X 4 matrix of  $I(\cdot)$ , are approximated by

$$I_{ij} = - \frac{\partial^2 \log l(\hat{\theta})}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, 2, 3, 4,$$

where,

$$I_{ij} = - \frac{\partial^2 \log l(\hat{\theta})}{\partial \theta_i \partial \theta_j} = - \frac{\partial^2 \log l(\hat{\theta})}{\partial \alpha \partial \alpha} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \alpha \partial \beta} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \alpha \partial \gamma} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \alpha \partial \lambda} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \beta \partial \alpha} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \beta \partial \beta} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \beta \partial \gamma} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \beta \partial \lambda} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \gamma \partial \alpha} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \gamma \partial \beta} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \gamma \partial \gamma} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \gamma \partial \lambda} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \lambda \partial \alpha} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \lambda \partial \beta} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \lambda \partial \gamma} - \frac{\partial^2 \log l(\hat{\theta})}{\partial \lambda \partial \lambda}$$

$$\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\lambda}) \text{ estimated parameters.}$$

Now information matrix can be written as,

$$I(\hat{\theta}) = \begin{bmatrix} \frac{\partial^2 \log l(\hat{\theta})}{\partial \alpha^2} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \alpha \partial \beta} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \alpha \partial \gamma} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \alpha \partial \lambda} \\ \frac{\partial^2 \log l(\hat{\theta})}{\partial \beta \partial \alpha} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \beta^2} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \beta \partial \gamma} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \beta \partial \lambda} \\ \frac{\partial^2 \log l(\hat{\theta})}{\partial \gamma \partial \alpha} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \gamma \partial \beta} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \gamma^2} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \gamma \partial \lambda} \\ \frac{\partial^2 \log l(\hat{\theta})}{\partial \lambda \partial \alpha} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \lambda \partial \beta} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \lambda \partial \gamma} & \frac{\partial^2 \log l(\hat{\theta})}{\partial \lambda^2} \end{bmatrix}$$

$$\begin{aligned} & \frac{\partial^2 \log L}{\partial \alpha_i \partial \alpha_j} + \frac{\partial^2 \log L}{\partial \alpha_i \partial \beta_j} + \frac{\partial^2 \log L}{\partial \alpha_i \partial \gamma_j} + \frac{\partial^2 \log L}{\partial \alpha_i \partial \lambda_j} \\ & + \frac{\partial^2 \log L}{\partial \beta_i \partial \beta_j} + \frac{\partial^2 \log L}{\partial \beta_i \partial \gamma_j} + \frac{\partial^2 \log L}{\partial \beta_i \partial \lambda_j} \\ & + \frac{\partial^2 \log L}{\partial \gamma_i \partial \gamma_j} + \frac{\partial^2 \log L}{\partial \gamma_i \partial \lambda_j} \\ & + \frac{\partial^2 \log L}{\partial \lambda_i \partial \lambda_j} \end{aligned}$$

The elements of the fisher information matrix are given in Appendix 1. Therefore, the approximate 100(1- $\gamma$ )% two-sided, confidence interval for  $\theta$  is given by  $\widehat{\theta} \pm Z_{\gamma/2} \sqrt{I^{-1}(\widehat{\theta})}$ . Here  $Z_{\gamma/2}$  is the upper  $\gamma/2$  th percentile of a standard normal distribution.

#### Information Criterion

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are commonly used to assess which of several candidate distributions provides the best fit to the data [13]. A distribution is considered to fit the data well if it has lower AIC and BIC values compared to the others. Both criteria balance goodness of fit with model complexity, penalizing overfitting by including terms related to the number of parameters in the model.

$$AIC = 2K - 2 \ln L$$

$$BIC = K \ln(n) - 2 \ln L$$

Where  $L$  is the likelihood function,  $n$  is the sample size and  $K$  is the number of parameters estimated.

#### Kaplan-Meier (K-M) Estimator

The Kaplan-Meier estimator of the survival function is a non-parametric method used to estimate the survival function from lifetime data [13]. It is particularly useful for handling censored data, where the exact survival time is unknown for some subjects. The Kaplan-Meier estimator is given by,  $\widehat{S}(t) = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i})$ . Where  $t_i$  is the failure time,  $d_i$  is the number of events that occurs at time  $t_i$  and  $n_i$  is the number individuals at risk of experiencing the event immediately prior to  $t_i$ .

### Validation of the model

To validate the modified Weibull model, we analyzed the failure times of patients admitted to BLDE Hospital, Vijayapura. Data were collected from January 1, 2023, to December 31, 2023, comprising 936 patients who passed away during this period.

Table 1 presents the estimated parameters of the model, along with their corresponding standard errors, lower confidence limits (LCL), and upper confidence limits (UCL).

Parameter/Causes	Values	Standard Error	LCL	UCL
$\alpha$	0.001	0.00001377435	0.000973	0.001027
$\beta$	1.500999	0.03391981	1.434517	1.567481
$\gamma$	1.565522	0.05214207	1.463325	1.667719
$\lambda$	0.8138159	0.02118096	0.772302	0.85533

**TABLE 1: Estimated values for the real life dataset.**

UCL: upper confidence limit; LCL: lower confidence limit

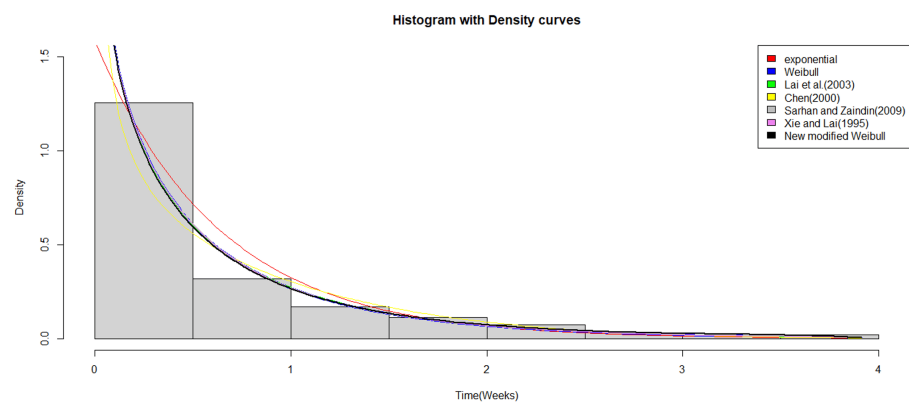
Table 2 provides the estimated AIC and BIC values for various modified forms of the Weibull model. Among all the distributions considered, the newly proposed modified Weibull distribution provided the best fit.

Distributions	logl	AIC	BIC
Exponential	-505.0891	1012.178	1017.02
Weibull	-476.1138	956.2276	965.9108
Chen (2000) [5]	-512.9805	1029.961	1039.644
Sarhan and Zaindin (2009) [9]	-476.1168	958.2335	972.7584
Lai et al. (2003) [6]	-476.2299	958.4597	972.6846
Xie et al. (2002) [10]	-470.4547	1012.178	1017.02
New modified Weibull	-468.0408	944.0816	963.448

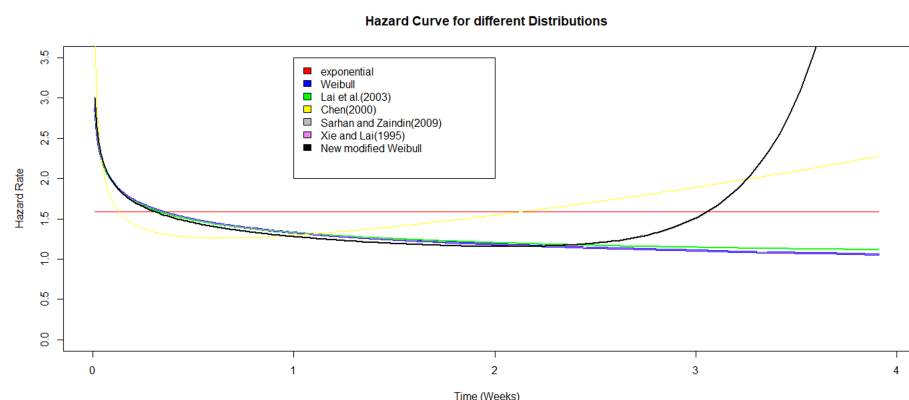
**TABLE 2: AIC and BIC values for the data.**

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion

Figure 3 illustrates the histogram of the data along with the fitted density curves for each distribution. Although Figure 3 may suggest that the data follows an exponential distribution, implying a constant hazard rate, this is not the case. By modifying the distribution, we achieve a more accurate hazard curve, as shown in Figure 4, which is the bathtub.



**FIGURE 3: Histogram and fitted density curves for the real life dataset.**

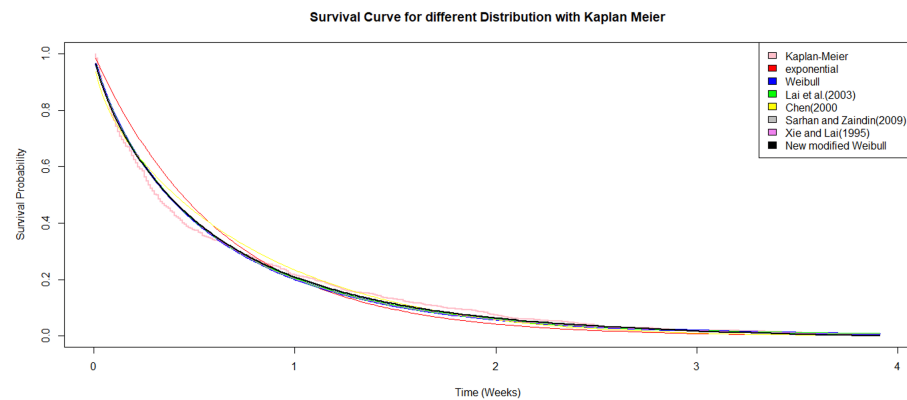


**FIGURE 4: Hazard curves for all distribution using real life dataset.**

In Figure 4, the hazard curves for all distributions are presented, with the modified Weibull distribution displaying a characteristic bathtub-shaped hazard function. Initially, the hazard decreases, indicating a

reduction in risk. After reaching 0.5 weeks, the hazard stabilizes, maintaining a steady state until about 2.5 weeks, after which the risk begins to increase again.

Figure 5 compares the survival curves for all distributions, including the Kaplan-Meier survival curve. The modified Weibull distribution provides a better fit to the observed data, with 50% of the patients having died within 0.5 weeks.



**FIGURE 5: Comparison of Kaplan-Meier survival curve with all distribution survival curves using real life dataset.**

## Discussion

The proposed four-parameter modified Weibull distribution is notable for its ability to model complex failure time data characteristics, such as bimodal density and non-monotonic hazard rates. Unlike the three-parameter model introduced by Lai et al. [6], which is less suited for datasets with bimodal behavior and generally shows a decreasing hazard function for the given dataset, the newly introduced model offers greater flexibility. This flexibility is evidenced by its characteristic bathtub-shaped hazard curve (Figure 4), allowing it to more accurately capture a wider range of failure patterns.

Similarly, the four-parameter additive Weibull model proposed by Xie and Lai [10], while more flexible than the three-parameter version, still tends to show a decreasing hazard function, making it less effective in modeling complex patterns such as bimodal data. In contrast, the newly modified Weibull distribution demonstrates superior performance, as reflected in its lower AIC and BIC values (Table 2), indicating a better fit to the data.

Although Figure 3 suggests that the data may follow an exponential distribution, characterized by a constant hazard rate, the modified Weibull model provides a more precise representation of the data. Its key advantage is its versatility, which allows it to represent a wide range of hazard function shapes, including decreasing, constant, increasing, and bathtub-shaped curves. This flexibility makes the model particularly suitable for datasets with intricate failure patterns, such as those exhibiting unimodal or bimodal densities (Figure 1). The model's ability to capture diverse hazard function shapes, ranging from simple monotonic to more complex bathtub-shaped curves demonstrates its potential in accurately modeling real-world failure time data, whereas traditional models may fall short in capturing the true underlying behavior.

However, a notable limitation of this model is the complexity involved in estimating its parameters. Standard estimation methods may not be effective, requiring simulation techniques to estimate the parameters accurately. Additionally, the moments of this distribution are not directly estimable and also require simulation methods to be computed. This complexity introduces challenges in practical applications, as these simulation-based approaches may be computationally intensive and time-consuming.

## Conclusions

The proposed four-parameter modified Weibull distribution is an effective model for analyzing failure time data, particularly in cases with bimodal density and non-monotonic hazard rates. Its flexibility allows it to capture various hazard shapes, including increasing, decreasing, constant, and, notably, the bathtub-shaped hazard, which is highly relevant in survival analysis.

Validation using hospital records on patient failure times demonstrated a best fit compared to other modified Weibull distributions proposed by different authors. This was confirmed by AIC and BIC values. The estimated standard errors, along with lower and upper confidence limits for the parameters, were provided



to ensure the robustness of the parameter estimates.

Additionally, the hazard curves for multiple distributions were plotted, where this model exhibited the characteristic bathtub-shaped hazard. The Kaplan-Meier survival curves were also plotted alongside these distributions for visual comparison of survival functions. The findings suggest that the four-parameter modified Weibull distribution is a flexible and reliable tool for modeling complex survival data with varying hazard rate behaviors, making it a strong candidate for practical applications in survival analysis.

## Appendices

### Appendix 1

Using the log likelihood function:

$$\log L = \sum_{i=1}^n \log \left( \frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1} \right) - \sum_{i=1}^n \left( \frac{\alpha}{\beta^\alpha} (e^{t_i^\beta} - 1) + \gamma t_i^\lambda \right)$$

$$deno = \frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}$$

$$nume\alpha = \beta t_i^{\beta-1} e^{t_i^\beta} \left( \frac{1 - \alpha \log \beta}{\beta^\alpha} \right)$$

$$nume\beta = \alpha \frac{(t_i^{\beta-1} e^{t_i^\beta} (2 \beta \log t_i + 1) - \alpha t_i^{\beta-1} e^{t_i^\beta})}{\beta^\alpha}$$

$$nume\lambda = \gamma t_i^{\lambda-1} (\lambda \log t_i + 1)$$

The first order partial derivatives with respect to  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\lambda$  were given below.

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^n \frac{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} - \frac{\alpha}{\beta^\alpha} (e^{t_i^\beta} - 1)}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}} - \sum_{i=1}^n \frac{e^{t_i^\beta} - 1}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}}$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^n \frac{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} \left( \frac{1 - \alpha \log \beta}{\beta^\alpha} \right) - \frac{\alpha}{\beta^\alpha} (e^{t_i^\beta} - 1) \frac{\alpha}{\beta^\alpha}}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}} - \sum_{i=1}^n \frac{e^{t_i^\beta} - 1}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}}$$

$$\frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^n \frac{\gamma \lambda t_i^{\lambda-1} - \gamma t_i^\lambda}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}} - \sum_{i=1}^n \frac{t_i^\lambda}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}}$$

$$\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^n \frac{\gamma t_i^{\lambda-1} (\lambda \log t_i + 1) - \gamma t_i^\lambda}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}} - \sum_{i=1}^n \frac{t_i^\lambda}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}}$$

The second order partial derivatives with respect to  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\lambda$  were given below.

$$\frac{\partial^2 \log L}{\partial \alpha^2} = \sum_{i=1}^n \frac{\left( \frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} - \frac{\alpha}{\beta^\alpha} (e^{t_i^\beta} - 1) \right)^2}{\left( \frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1} \right)^2} - \sum_{i=1}^n \frac{e^{t_i^\beta} - 1}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}} \frac{\partial}{\partial \alpha} \left( \frac{1 - \alpha \log \beta}{\beta^\alpha} \right)$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = \sum_{i=1}^n \frac{\left( \frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} \left( \frac{1 - \alpha \log \beta}{\beta^\alpha} \right) - \frac{\alpha}{\beta^\alpha} (e^{t_i^\beta} - 1) \frac{\alpha}{\beta^\alpha} \right)^2}{\left( \frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1} \right)^2} - \sum_{i=1}^n \frac{\partial}{\partial \beta} \left( \frac{e^{t_i^\beta} - 1}{\frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1}} \right)$$

$$\frac{\partial^2 \log L}{\partial \gamma^2} = \sum_{i=1}^n \frac{\gamma \lambda t_i^{\lambda-1} - \gamma t_i^\lambda}{\left( \frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1} \right)^2} - \sum_{i=1}^n \frac{t_i^\lambda}{\left( \frac{\alpha}{\beta^\alpha} \beta t_i^{\beta-1} e^{t_i^\beta} + \gamma \lambda t_i^{\lambda-1} \right)^2}$$

$$\frac{\partial^2 \log L}{\partial \lambda^2} = \sum_{i=1}^n \frac{(deno * \frac{\partial}{\partial \lambda}(nume\lambda)) - (nume\lambda)^2}{deno^2} - \sum_{i=1}^n \lambda t_i^\lambda (\log t_i)^2$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial \beta} = \frac{\partial^2 \log L}{\partial \beta \partial \alpha} = \sum_{i=1}^n \frac{(deno * \frac{\partial}{\partial \beta}(nume\alpha)) - (nume\alpha * nume\beta)}{deno^2} - \sum_{i=1}^n \frac{\partial}{\partial \beta} (e^{t_i^\beta} - 1) \left( \frac{1 - \alpha \log \beta}{\beta^\alpha} \right)$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial \gamma} = \frac{\partial^2 \log L}{\partial \gamma \partial \alpha} = - \sum_{i=1}^n \frac{\lambda t_i^{\lambda-1} nume\alpha}{deno^2}$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial \lambda} = \frac{\partial^2 \log L}{\partial \lambda \partial \alpha} = - \sum_{i=1}^n \frac{nume\lambda * nume\alpha}{deno^2}$$

$$\frac{\partial^2 \log L}{\partial \beta \partial \gamma} = \frac{\partial^2 \log L}{\partial \gamma \partial \beta} = - \sum_{i=1}^n \frac{\lambda t_i^{\lambda-1} * nume\beta}{deno^2}$$

$$\frac{\partial^2 \log L}{\partial \beta \partial \lambda} = \frac{\partial^2 \log L}{\partial \lambda \partial \beta} = - \sum_{i=1}^n \frac{nume\beta * nume\lambda}{deno^2}$$

$$\frac{\partial^2 \log L}{\partial \gamma \partial \lambda} = \frac{\partial^2 \log L}{\partial \lambda \partial \gamma} = \sum_{i=1}^n \frac{(deno * t_i^{\lambda-1} * (\lambda \log t_i + 1)) - (\lambda t_i^{\lambda-1} * nume\lambda)}{deno^2} - \sum_{i=1}^n t_i^\lambda \log t_i$$

## Additional Information

### Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

**Concept and design:** A M. Rangoli, A S. Talawar, R P. Agadi, Vijaya Sorganvi

**Acquisition, analysis, or interpretation of data:** A M. Rangoli, A S. Talawar, R P. Agadi, Vijaya Sorganvi

**Drafting of the manuscript:** A M. Rangoli, A S. Talawar, R P. Agadi, Vijaya Sorganvi

**Critical review of the manuscript for important intellectual content:** A M. Rangoli, A S. Talawar, R P. Agadi, Vijaya Sorganvi

**Supervision:** A M. Rangoli, A S. Talawar, R P. Agadi, Vijaya Sorganvi

### Disclosures

**Human subjects:** All authors have confirmed that this study did not involve human participants or tissue.

**Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue.

**Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

### Acknowledgements

Thanks to Shri B M Patil Medical College Hospital and Research Centre, Vijayapura for providing necessary data for the research work.

## References

1. Almalki SJ, Nadarajah S: A new discrete modified Weibull distribution . IEEE Trans. Reliab. 2014, 63:68-80. [10.1109/TR.2014.2299691](https://doi.org/10.1109/TR.2014.2299691)
2. Almalki SJ, Yuan J: A new modified Weibull distribution . Reliability Engineering & System Safety. 2013, 111:164-70. [10.1016/j.res.2012.10.018](https://doi.org/10.1016/j.res.2012.10.018)
3. Bebbington M, Lai CD, Zitikis R: A flexible Weibull extension. Reliability Engineering & System Safety. 2007, 92:719-26. [10.1016/j.res.2006.03.004](https://doi.org/10.1016/j.res.2006.03.004)
4. Carrasco JM, Ortega EM, Cordeiro GM: A generalized modified Weibull distribution for lifetime modeling .



- Computational Statistics & Data Analysis. 2008, 53-450. [10.1016/j.csda.2008.08.023](https://doi.org/10.1016/j.csda.2008.08.023)
5. Chen Z: A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function . Statistics & Probability Letters. 2000, 49:155-61. [10.1016/S0167-7152\(00\)00044-4](https://doi.org/10.1016/S0167-7152(00)00044-4)
6. Lai CD, Xie M, Murthy DNP: A modified Weibull distribution . IEEE Trans Reliab. 2003, 52:33-7. [10.1109/TR.2002.805788](https://doi.org/10.1109/TR.2002.805788)
7. Mudholkar GS, Hutson AD: The exponentiated Weibull family: some properties and a flood data application . Communications in Statistics - Theory and Methods. 1996, 25:3059-83. [10.1080/03610929608831886](https://doi.org/10.1080/03610929608831886)
8. Mudholkar GS, Srivastava DK, Freimer M: The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. Technometrics. 1995, 37:436-45. [10.1080/00401706.1995.10484376](https://doi.org/10.1080/00401706.1995.10484376)
9. Sarhan AM, Zaindin M: Modified Weibull distribution . Appl Sci. 2009, 11:123-36.
10. Xie M, Lai CD: Reliability analysis using an additive Weibull model with bathtub-shaped . Reliability Engineering & System Safety. 1996, 52:87-93. [10.1016/0951-8320\(95\)00149-2](https://doi.org/10.1016/0951-8320(95)00149-2)
11. Xie M, Tang Y, Goh TN: A modified Weibull extension with bathtub-shaped . Reliability Engineering & System Safety. 2002, 76:279-85. [10.1016/S0951-8320\(02\)00022-4](https://doi.org/10.1016/S0951-8320(02)00022-4)
12. A Review of Monte Carlo-Based Versions of the EM Algorithm . (2024). <https://arxiv.org/abs/2401.00945>.
13. Lawless JF: Statistical Models and Methods for Lifetime Data . John Wiley & Sons, New York; 2002.