Stochastic Models in Probability and Statistics
Vol. -, No. -, 2025, pp. -. Research Article

Log transformed transmuted exponential distribution: an increasing hazard rate model to deal with cancer patients data

Md. Tahir*, Sanjay K. Singh and Abhimanyu Singh Yadav

Department of Statistics, Banaras Hindu University, Varanasi, India

Email(s): tahirmansori@gmail.com, singhsk64@gmail.com, asybhu10@gmail.com

Abstract. In this paper, we introduce a novel distribution called the log transformed transmuted exponential (LTTE), which is derived by applying a log transformation to the transmuted exponential distribution as the baseline model. We derive several key mathematical and statistical properties of the LTTE distribution, including its moments, quantile function, skewness, kurtosis, reliability function, and hazard rate, along with their respective shapes. The maximum likelihood estimation method is used to estimate the parameters of the distribution. The practical applicability of the LTTE distribution is demonstrated by fitting it to three real-life datasets related to cancer patients. The results indicate that the LTTE distribution offers a superior fit, as evidenced by better values of AIC, BIC, and the Kolmogorov–Smirnov (KS) statistic, when compared to other existing lifetime models.

Keywords: Maximum likelihood estimation, Moments, Transmuted exponential, Log transformation.

1 Introduction

It is an indisputable fact that the world is facing an epidemic of noncommunicable diseases, with cancer cases continuing to grow at an alarming rate. Cancer is currently ranked as the second leading cause of death, following cardiovascular diseases; see Jemal et al. (2008). The GLOBOCAN 2018 report recorded 18.1 million new cancer cases and 9.6 million cancer-related deaths globally. Emerging challenges such as rapid urbanization, population aging, unhealthy lifestyles, and indoor and outdoor air pollution are contributing to the growing cancer burden worldwide, particularly in middle and low-income countries, for example, India. According to the WHO 2020 ranking on cancer burden, India ranks third in terms of new yearly cancer cases,

*Corresponding author

Received: / Accepted:
DOI: 10.22067/smps.2025.91480.1041

following China and United States. The Indian Council of Medical Research's National Cancer Registry Program (ICMR-NCRP), an initiative by the government of India to estimate cancer incidence in the country, reported 1.39 million new cancer cases in 2020 and 1.46 million in 2022. The GLOBOCAN projection estimates that cancer cases in India will rise to 2.08 million by 2040, representing a nearly 50% increase from 2020, Sathish Kumar et al. (2022). Among all the cancer cases, Breast cancer is the leading cause of cancer incidence and mortality in India, accounting for 13.5% of new cases and 10.6% of cancer-related deaths in 2020, Mehrotra and Yadav (2022). Urban factors such as sedentary lifestyle, high obesity rates, delayed marriage, and childbirth, and minimal breastfeeding contribute to its higher burden in urban areas. Another type of hazardous cancer is bladder cancer which ranks as the ninth most common cancer, representing 3.9% of all cancer cases in India, Prakash et al. (2019). It is primarily linked to tobacco use and exposure to industrial chemicals. Also, leukemia, a cancer of the blood-forming tissues, compromises the body's ability to fight infections. It accounts for 27% to 52% of childhood cancers in males and 19% to 52% in females across various population-based registries, Bhutani et al. (2004).

With such a large number of cancer cases being reported, there is a vast amount of data available to perform statistical analyses to identify root causes and develop better treatments for cancer patients. This large-scale data can be effectively analyzed with the help of statistical models, which play a critical role in data interpretation. Statistical models help in quickly and accurately gaining information about the population, often at a lower cost. Once a model is identified, inferences can be drawn from the sample data to understand the broader population trends. Therefore, the development and construction of suitable models are essential for solving complex real-world problems, such as cancer data analysis. Many statisticians have developed various models to address the data from different types of cancer, such as breast cancer, bladder cancer, and leukemia, namely, Al-Kadim and Mahdi (2018) developed the exponentiated transmuted exponential model for analyzing breast cancer survival times, outperforming models like log normal, log logistic, and exponential. Khan et al. (2013) proposed the transmuted inverse Weibull model for bladder cancer data, showing strong performance compared to other models. Kumar et al. (2015) introduced the DUS exponential model, which surpassed the transmuted inverse Weibull and other models. Elbatal et al. (2013) developed the transmuted generalized linear exponential model for leukemia, enhancing the understanding of survival patterns in leukemia patients.

In this discussion, we will explore some of the statistical models developed for analyzing various types of cancer data and their applications. However, it is important to note that not every model is always perfectly suited to real-life phenomena. This is because real-life situations are dynamic and subject to change over time and several inherent factors may influence the outcomes. These factors may include shifts in environmental conditions, advances in medical treatments or changes in patient demographics and lifestyles. As a result, existing models may not always capture the complexities or evolving trends inherent in these phenomena. Given these challenges, there is an ongoing need to update or refine existing models to ensure they accurately reflect current realities. In some cases, this might involve modifying an existing model to account for new factors or changes in the underlying data. In other cases, the development of entirely new models may be necessary to address the limitations of current models and improve their predictive accuracy and applicability. With this motivation in mind, the present study aims to

develop a new statistical model that can effectively analyze data from multiple types of cancer, namely, breast cancer, bladder cancer, and leukemia. The goal is to create a model that not only fits the unique characteristics of these cancer datasets but also outperforms many existing models in terms of different model comparison criteria (e.g., AIC, BIC). By doing so, this study seeks to contribute to more effective analyses of cancer data, leading to better insights into the survival patterns, risk factors, and treatment outcomes for cancer patients across different types of cancer.

In the field of statistical modeling, numerous methodologies have been proposed to create new distributions by modifying or extending an existing baseline distribution. These approaches often involve introducing additional parameters or applying transformations to the baseline distribution, allowing for greater flexibility and adaptability to a wide range of real-world data scenarios. The motivation behind these methods is to capture the complexities and nuances of different types of data that cannot be adequately modeled by standard distributions alone. Some common techniques include applying a parameterized transformation to the cumulative distribution function (CDF) or probability density function (PDF) of the baseline distribution, compounding it with another distribution or incorporating additional shape or scale parameters. Few of them are discussed here; namely, Gupta et al. (1998) proposed a method for generalizing the existing distribution by taking power of the CDF of any baseline probability distribution. Verma et al. (2024) has also proposed a new distribution using the generalization technique.

In recent years, various transformation techniques have been introduced to develop new probability models. The quadratic rank transmutation map (QRTM) technique is widely used but often increases computational complexity by adding parameters; see Shaw and Buckley (2009). In contrast, the DUS transformation technique, introduced by Kumar et al. (2015), enhances baseline distribution flexibility while remaining parsimonious in parameters, reducing estimation complexity. Similarly, the log transformation technique proposed by Maurya et al. (2016) combines parameter parsimony with increased distributional flexibility. These advancements simplify parameter estimation while maintaining robust modeling capabilities.

The primary objective of this article is to introduce a novel probability distribution, termed the log transformed transmuted exponential (LTTE) distribution. This new model is derived using the log transformation technique, which has been recognized for its ability to enhance the flexibility of baseline distributions while maintaining parameter parsimony. Specifically, the LTTE distribution is developed by applying the log transformation to the transmuted exponential distribution, which serves as the baseline. The transmuted exponential distribution (see Owoloko et al. (2015)) is a generalized version of the standard exponential distribution, introduced to provide greater flexibility in modeling data. This distribution is obtained by applying the QRTM to the exponential distribution, thereby adding a single parameter that enhances its ability to capture diverse data behaviors. The CDF and PDF of this distribution are given by

$$G(x;\lambda,\alpha) = (1-e^{-\lambda x})\left(1+\alpha e^{-\lambda x}\right),$$

and

$$g(x;\lambda,\alpha) = \lambda e^{-\lambda x}\left(1-\alpha+2\alpha e^{-\lambda x}\right), \quad x \geq 0, \lambda > 0, |\alpha| \leq 1,$$

respectively.

Now by considering the above defined base line distributions, the logarithmic transformed transmuted exponential distribution is given with the following CDF and PDF

$$F(x) = 1 - \frac{1}{\log 2} \log\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right],$$

and

$$f(x) = \frac{\lambda e^{-\lambda x}(1 - \alpha + 2\alpha e^{-\lambda x})}{[2 - (1 - e^{-\lambda x})(1 + \alpha e^{-\lambda x})]\log 2}, \quad x \geq 0, \lambda > 0, |\alpha| \leq 1,$$

respectively. The above proposed distribution is denoted by LTTE(x; α, λ), where α and λ are

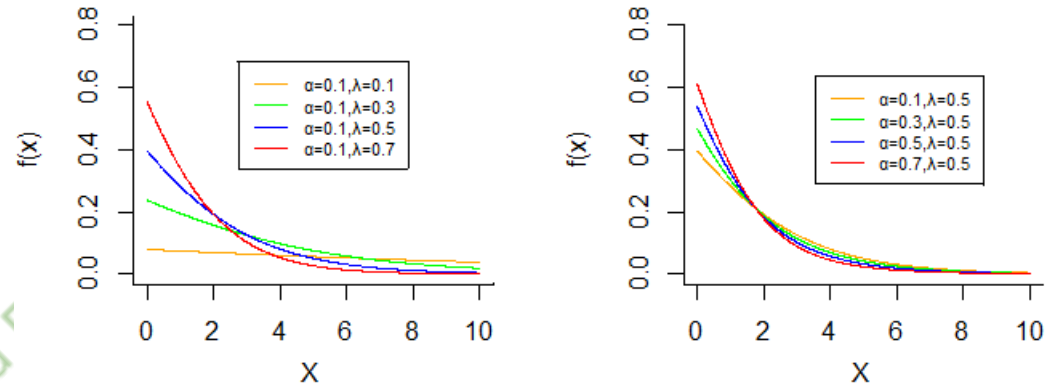

Figure 1: PDF of LTTE for some selected values of λ and α.

the shape and scale parameters, respectively. The shape of the PDF of the proposed distribution is presented in Figure 1. By leveraging the properties of log transformation, the proposed distribution aims to address limitations in existing models, offering improved adaptability to various data sets and practical applications. This approach not only expands the family of transmuted distributions but also contributes to the growing repertoire of tools for statistical modeling and analysis.

The structure of the paper is as follows: Section 1 provides an introduction to the study. Section 2, along with its subsections, explores the distributional properties of the proposed model in detail. Section 3 discusses the parameter estimation using the maximum likelihood estimation (MLE) technique. Section 4 presents simulation studies to evaluate the performance of the estimators. In Section 5, the applicability of the proposed model is demonstrated using three real datasets related to cancer patients. Finally, Section 6 concludes the paper with a summary of the findings and key conclusions.

2 Distributional properties

A new probability distribution is characterized by considering its associated properties. Each of the properties of PDF provides valuable insights and behavior of the random variable it represents. Thus in this section, different distributional properties have been derived for the proposed probability distribution.

2.1 Survival characteristics

In this section, we will discuss the survival function and hazard rate of the proposed model.

- The survival function $S(x)$ is the probability that an equipment/item survived at least time x and it is defined as

$$S(x) = P(X > x) = \frac{1}{\log 2} \log\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right].$$

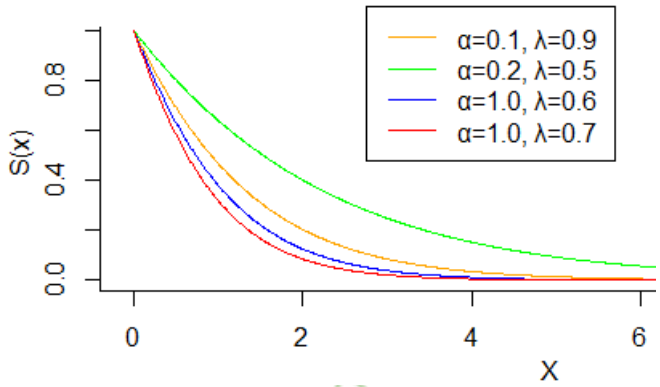

Figure 2: Survival function of LTTE for varying values of α and λ.

- The hazard rate function $h(x)$ is the instantaneous failure rate and it is defined by

$$h(x) = \frac{\lambda e^{-\lambda x}\left(1 - \alpha + 2\alpha e^{-\lambda x}\right)}{\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right] \log\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right]}.$$

- The reverse hazard rate $\tilde{h}(x)$ is obtained as

$$\tilde{h}(x) = \frac{\frac{\lambda e^{-\lambda x}\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]\log 2}}{1 - \frac{\log\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]}{\log 2}}.$$

- The cumulative hazard function for LTTE is given by

$$H(x) = -\log S(x) = -\log\left[\frac{\log\left\{2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right\}}{\log 2}\right].$$

The graphical representation of the $S(x)$ and $h(x)$ of the proposed model for varying values of model parameters α and λ are presented in the Figures 2 and 3, respectively. We have examined the nature of hazard for different combinations of model parameters α and λ, and from the graph, it is evident that the model is of increasing hazard nature (except the parameters combination where values of α and λ are very high and it showing nonmonotone hazard rate).

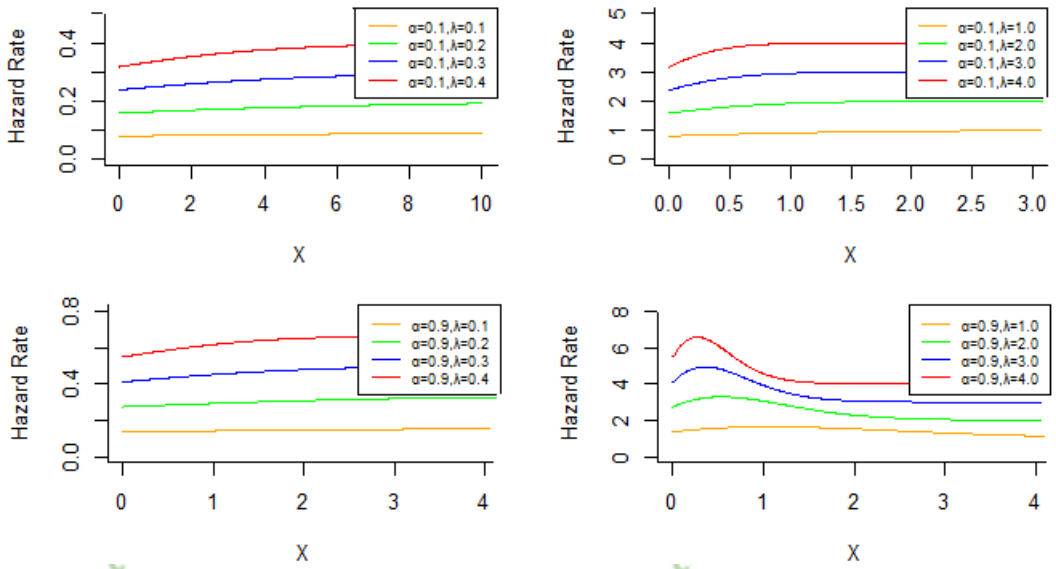

Figure 3: Hazard rate of LTTE for varying values of α and λ.

2.2 Moments

Moments are fundamental properties of any distribution and are widely used to analyze its features and characteristics. The rth moment about the origin for the proposed distribution is expressed as

$$\mu_r' = E(X^r) = \int_0^\infty x^r \frac{\lambda e^{-\lambda x}\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]\log 2} dx$$
$$= \frac{\lambda}{\log 2}\sum_{i=0}^{\infty}(-1)^i \frac{\lambda^i}{i!}\int_0^\infty x^{i+r} \frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]} dx.$$

The respective moments are obtained by putting the values of r.

2.3 Quantile function

The pth quantile function denoted by $Q(p)$ of LTTE $(x;\ \alpha,\ \lambda)$ is obtained by solving

$$F\left[Q(p)\right] = p;$$

and after simplification, the expression for quantile function is given by

$$Q(p) = -\frac{1}{\lambda}\log\left[\sqrt{\left\{\left(\frac{1-\alpha}{2\alpha}\right)^2 - \left(\frac{1-2^{1-p}}{\alpha}\right)\right\}} - \left(\frac{1-\alpha}{2\alpha}\right)\right]. \quad (1)$$

The respective values of quantile can be obtained by putting different values of $p \in (0,1)$ in the above expression for known values of model parameters.

2.4 Skewness and kurtosis

The skewness and kurtosis are commonly employed to analyze the asymmetry and sharpness of a probability distribution. However, their computation often relies on moments, which may not exist for certain distributions. To address these limitations, alternative measures based on the quantile function have been proposed. Notably, Bowley (1920) and Moors (1988) introduced coefficients of skewness and kurtosis that depend on quantile. Bowley's coefficient of skewness, in particular, is defined as follows:

$$B = \frac{Q(3/4) + Q(1/4) - 2Q(1/2)}{Q(3/4) - Q(1/4)}.$$

Similarly, the Moors' coefficient of kurtosis is given by

$$M = \frac{Q(3/8) - Q(1/8) + Q(7/8) - Q(5/8)}{Q(6/8) - Q(2/8)}.$$

Using (1), the coefficients of skewness and kurtosis can be calculated.

2.5 Order statistics

The order statistics are very crucial in statistical analysis as in this case, the analysis of data is performed in ascending or descending order. They are very helpful specially when dealing with extreme observations (e.g., minimum, maximum). Here we will find the expressions of PDFs for 1st, rth, and nth order statistics when the sample follows the proposed distribution.

Suppose that $X_1, X_2, \ldots, X_n$ are a random sample of size n from the proposed distribution and their corresponding order statistics are $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$. The PDF of the rth order statistic $X_{r:n}$, say $f_{r:n}(x)$, for the proposed distribution is given by

$$f_{r:n}(x) = \frac{n!}{(r-1)!\,(n-r)!\,[\log 2]^n} \left[\log\left\{\frac{2}{[2-(1-e^{-\lambda x})\,(1+\alpha e^{-\lambda x})]}\right\}\right]^{r-1} \times \lambda e^{-\lambda x} \frac{\left[\log\left\{2-(1-e^{-\lambda x})(1+\alpha e^{-\lambda x})\right\}\right]^{n-r}(1-\alpha+2\alpha e^{-\lambda x})}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]}.$$

For $r = 1$ and $r = n$, it simplifies as, respectively,

$$f_{1:n}(x) = \frac{n\lambda e^{-\lambda x}}{[\log 2]^n} \left[\log\left\{2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right\}\right]^{n-1} \frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]}$$

and

$$f_{n:n}(x) = \frac{n\lambda e^{-\lambda x}}{[\log 2]^n} \left[\log\left\{\frac{2}{\left\{2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right\}}\right\}\right]^{n-1} \frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]}.$$

2.6 Entropy

Entropy is a measure of the uncertainty or randomness associated with a random variable X characterized by its PDF $f(x)$. It is a significant concept with applications in various fields, including communication, physics and reliability. Among the various measures of entropy, the Rényi entropy, introduced by Rényi (1961), is one of the most widely used. For the proposed distribution, the Rényi entropy is defined as

$$R_\rho = \frac{1}{(1-\rho)} \log \int_{-\infty}^{+\infty} [f(x)]^\rho \, dx \qquad \rho \neq 1.$$

After simplification, the expression for R_ρ for proposed distribution is given by

$$R_\rho = (-1)^i \frac{1}{1-\rho} \log \left\{ \sum_{i=0}^{\infty} \int_0^{\infty} \frac{(x\rho)^i \lambda^{i+\rho}}{i!} \left[\frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\log 2 \left\{ 2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right) \right\}} \right]^\rho dx \right\}, \qquad \rho \neq 1,$$

where ρ is the order of entropy.

2.7 Bonferroni and Lorenz curve

The Lorenz curve and the Bonferroni curve are graphical tools used to analyze inequality in the distribution of resources, such as income or wealth. The Lorenz curve plots the cumulative share of a quantity held by the bottom $x\%$ of the population, highlighting overall inequality, with greater deviations from the diagonal line indicating higher inequality, for more detail see Lorenz (1905). In contrast, the Bonferroni curve focuses on the proportional share of resources held by the lower part of the population, making it more sensitive to changes at the lower end of the distribution; see for more detail Bonferroni (1941). Both curves complement each other in understanding inequality, with the Lorenz curve offering a broader perspective and the Bonferroni curve providing detailed insights into the distribution among the less advantaged. The expression for Lorenz and Bonferroni curves for the proposed distribution are given by

$$L(p) = \frac{1}{\mu} \int_0^p x f(x) \, dx = \frac{\lambda}{\mu \log 2} \int_0^p x e^{-\lambda x} \frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]} dx$$

and

$$B(p) = \frac{1}{\mu p} \int_0^p x f(x) \, dx = \frac{\lambda}{\mu p \log 2} \int_0^p x e^{-\lambda x} \frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]} dx,$$

respectively.

3 Parameter estimation

Parameter estimation is the process of determining the unknown parameters of a statistical model based on observed data, with the goal of identifying the parameters that best describe the underlying distribution or process. Common methods include MLE, which maximizes the likelihood function; the method of moments, which matches sample moments with theoretical

moments; least squares estimation, which minimizes the sum of squared differences between observed and predicted values. The choice of method depends on the data, model and assumptions about the distribution. Here, the MLE method has been considered for the estimation of the model parameters.

3.1 MLE

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $LTTE(x;\alpha,\lambda)$. Then the likelihood function is given by

$$L(\alpha,\lambda|\mathbf{x}) = \frac{\lambda^n}{(\log 2)^n} e^{-\lambda\sum_{i=1}^n x_i} \frac{\prod_{i=1}^n \left(1-\alpha+2\alpha e^{-\lambda x_i}\right)}{\prod_{i=1}^n \left[2-\left(1-e^{-\lambda x_i}\right)\left(1+\alpha e^{-\lambda x_i}\right)\right]}.$$

Therefore, the log-likelihood function after ignoring the constant term is given by

$$\log L = n\log\lambda - \lambda\sum_{i=1}^n x_i + \sum_{i=1}^n \log\left(1-\alpha+2\alpha e^{-\lambda x_i}\right) - \sum_{i=1}^n \log\left[2-\left(1-e^{-\lambda x_i}\right)\left(1+\alpha e^{-\lambda x_i}\right)\right].$$

The maximum likelihood estimates $(\hat{\alpha},\hat{\lambda})$ of the parameter α, and λ can be obtained by differentiating the above equation with respect to α and λ, respectively, and equating them to zero. The following normal equations are obtained:

$$\sum_{i=1}^n \frac{2e^{-\lambda x_i}-1}{\left(1-\alpha+2\alpha e^{-\lambda x_i}\right)} + \sum_{i=1}^n e^{-\lambda x_i} \frac{1-e^{-\lambda x_i}}{\left[2-\left(1-e^{-\lambda x_i}\right)\left(1+\alpha e^{-\lambda x_i}\right)\right]} = 0 \tag{2}$$

and

$$\frac{n}{\lambda} - \sum_{i=1}^n x_i - 2\alpha\sum_{i=1}^n \frac{x_i e^{-\lambda x_i}}{\left(1-\alpha+2\alpha e^{-\lambda x_i}\right)} + \sum_{i=1}^n \frac{x_i e^{-\lambda x_i}\left(1-\alpha+2\alpha e^{-\lambda x_i}\right)}{\left[2-\left(1-e^{-\lambda x_i}\right)\left(1+\alpha e^{-\lambda x_i}\right)\right]} = 0, \tag{3}$$

respectively. The nonlinear equations (2) and (3) are challenging to solve analytically, as they cannot be expressed in closed form. Various methods have been proposed to address such equations, with the Newton–Raphson method being one of the most widely used techniques for iterative and numerical solutions. Using the Newton–Raphson method, the estimates of α and λ (denoted as $\hat{\alpha}$ and $\hat{\lambda}$, respectively) can be obtained by solving these nonlinear equations iteratively.

3.2 Asymptotic confidence interval

Since the explicit distributions of the ML estimators are not available in closed form, thus the asymptotic confidence intervals are constructed in this subsection; see Singh et al. (2014). To achieve this, the Fisher information matrix is derived to facilitate the computation of the asymptotic confidence intervals for the parameters α and λ. The resulting expressions for the Fisher information matrix are provided as follows:

$$I(\alpha,\lambda) = E\begin{bmatrix} -\frac{\delta^2 LogL}{\delta\alpha^2} & -\frac{\delta^2 LogL}{\delta\alpha\delta\lambda} \\ -\frac{\delta^2 LogL}{\delta\lambda\delta\alpha} & -\frac{\delta^2 LogL}{\delta\lambda^2} \end{bmatrix}.$$

All the above derivatives are evaluated at $(\hat{\alpha},\hat{\lambda})$. The asymptotic variance-covariance matrix of the maximum likelihood estimators is obtained by inverting the Fisher information matrix. The diagonal elements of $I^{-1}(\alpha,\lambda)$ provide the asymptotic variances of α and λ. Using large sample theory, a two-sided $100(1-\beta)\%$ asymptotic confidence intervals for the parameters α and λ are constructed as

$$\hat{\alpha} \mp Z_{\beta/2}\sqrt{\widehat{var}(\alpha)},$$

and

$$\hat{\lambda} \mp Z_{\beta/2}\sqrt{\widehat{var}(\lambda)},$$

respectively, where $Z_{\beta/2}$ is the tabulated value of standard normal distribution at $\beta/2\%$ level of significance. The width of a confidence interval reflects its precision, with narrower intervals indicating more precise estimates. The average width is the mean of all interval widths computed across simulations. Coverage probability measures how often the intervals contain the true parameter value, indicating their reliability. Ideally, intervals should have a small width for precision and a coverage probability close to the nominal level $(1-\beta)$ for accuracy.

3.3 MLE of survival function and hazard function

If $\hat{\alpha}$ and $\hat{\lambda}$ are the maximum likelihood estimates of the parameters α and λ, respectively, then by the invariance property of likelihood estimators, the estimates of the survival function and the hazard function for any mission time $t>0$ can also be obtained. According to this property, the survival function $S(t)$ and the hazard function $h(t)$, which are functions of α and λ, can be estimated by substituting their maximum likelihood estimates into the respective expressions. Thus, the estimated survival function and the estimated hazard function are given by

$$\hat{S}(x) = \frac{1}{\log 2}\log\left[2-\left(1-e^{-\hat{\lambda}x}\right)\left(1+\hat{\alpha}e^{-\hat{\lambda}x}\right)\right],$$

and

$$\hat{h}(x) = \frac{\hat{\lambda}e^{-\hat{\lambda}x}\left(1-\hat{\alpha}+2\hat{\alpha}e^{-\hat{\lambda}x}\right)}{\left[2-\left(1-e^{-\hat{\lambda}x}\right)\left(1+\hat{\alpha}e^{-\hat{\lambda}x}\right)\right]\log\left[2-\left(1-e^{-\hat{\lambda}x}\right)\left(1+\hat{\alpha}e^{-\hat{\lambda}x}\right)\right]},$$

respectively. These estimates provide practical insights into the reliability and risk associated with different mission times.

4 Simulation study

In this section, a Monte Carlo simulation is conducted to evaluate the performance of the proposed point estimators and interval estimates of the parameters. The simulation examines the behavior of the estimators under varying sample sizes and varying model parameter values. Specifically, the sample sizes considered are $n=30,60,90,120,150$, and the parameter combinations include $(\alpha=0.1,\lambda=0.9)$, $(\alpha=0.2,\lambda=0.8)$, $(\alpha=0.06,\lambda=0.7)$ and $(\alpha=0.1,\lambda=1.0)$. These parameter settings are chosen to cover a wide range of scenarios for assessing the performance of the estimators. The performance of the point estimators is assessed based on their

mean square errors (MSEs), while interval estimators are evaluated using average width (AW) and coverage probability (CP). 95% asymptotic confidence intervals for the parameters are constructed for the same variation, and their corresponding AWs and CPs are also reported. All computations are carried out using the open-source statistical R Software (2024) ensuring transparency and reproducibility of the study. Each combination of sample size and parameter values is examined through 10,000 replications and the results-comprising average estimates (AE), biases and MSEs which are summarized in tabular form (Table 1). The simulation results indicate that the MSEs of the estimators decreases to zero as the sample size n increases, accompanied by a reduction in the widths of the confidence intervals. This behavior demonstrates that the estimators become more precise and their estimated values converge to the true parameter values as the sample size grows. Furthermore, the negligible bias observed across all scenarios supports the conclusion that the proposed estimators are asymptotically unbiased.

Table 2 presents the computed estimates of the survival and hazard functions for varying values of model parameters, demonstrating how the reliability and risk change across different parameters and mission time scenarios.

Table 1: AE, bias, and MSEs along with AW and CPs for the parameters for varying values of α and λ with different sample sizes n.

n	$\alpha = 0.1$					$\lambda = 0.9$				
	AE	Bias	MSE	AW	CP	AE	Bias	MSE	AW	CP
30	-0.0395	-0.1395	0.2692	1.9696	0.8745	0.9852	0.0852	0.0859	1.0690	0.9151
60	0.0450	-0.0550	0.1650	1.5552	0.8680	0.9324	0.0324	0.0449	0.8076	0.9021
90	0.0785	-0.0215	0.1271	1.3574	0.8737	0.9141	0.0141	0.0337	0.6932	0.8939
120	0.0940	-0.0060	0.1077	1.2173	0.8740	0.9061	0.0061	0.0275	0.6169	0.8953
150	0.1037	0.0037	0.0951	1.1213	0.8788	0.9012	0.0012	0.0237	0.5656	0.8959
n	$\alpha = 0.2$					$\lambda = 0.8$				
30	0.0223	-0.1777	0.2758	1.9693	0.8742	0.8980	0.0980	0.0787	0.9997	0.9105
60	0.1140	-0.0860	0.1653	1.5746	0.8686	0.8449	0.0449	0.0404	0.7644	0.8993
90	0.1565	-0.0435	0.1289	1.3731	0.8681	0.8235	0.0235	0.0305	0.6562	0.8868
120	0.1745	-0.0255	0.1077	1.2537	0.8693	0.8147	0.0147	0.0246	0.5941	0.8878
150	0.1888	-0.0112	0.0960	1.1569	0.8696	0.8080	0.0080	0.0213	0.5455	0.8832
n	$\alpha = 0.06$					$\lambda = 0.7$				
30	-0.0663	-0.1263	0.2673	1.9719	0.8718	0.7597	0.0597	0.0490	0.8160	0.9177
60	0.0143	-0.0457	0.1652	1.5440	0.8680	0.7211	0.0211	0.0259	0.6121	0.9037
90	0.0464	-0.0136	0.1276	1.3383	0.8760	0.7078	0.0078	0.0195	0.5221	0.8965
120	0.0593	-0.0007	0.1070	1.2025	0.8770	0.7027	0.0027	0.0158	0.4653	0.8987
150	0.0667	0.0067	0.0933	1.1075	0.8849	0.6997	-0.0003	0.0134	0.4263	0.9020
n	$\alpha = 0.01$					$\lambda = 1.0$				
30	-0.0995	-0.1095	0.2661	1.9585	0.8708	1.0740	0.0740	0.0935	1.1327	0.9200
60	-0.0250	-0.0350	0.1658	1.5259	0.8687	1.0238	0.0238	0.0499	0.8453	0.9073
90	0.0030	-0.0070	0.1263	1.3209	0.8781	1.0074	0.0074	0.0373	0.7199	0.9027
120	0.0150	0.0050	0.1066	1.1766	0.8832	1.0008	0.0008	0.0305	0.6363	0.9033
150	0.0195	0.0095	0.0909	1.0781	0.8929	0.9982	-0.0018	0.0254	0.5802	0.9106

Table 2: Estimate of the reliability and hazard functions for varying values of t for different sample sizes n and model parameters α and λ.

n	t	$\alpha=0.1,\lambda=0.9$		$\alpha=0.2,\lambda=0.8$		$\alpha=0.06,\lambda=0.7$		$\alpha=0.1,\lambda=1.0$	
		$\hat{R}(t)$	$\hat{h}(t)$	$\hat{R}(t)$	$\hat{h}(t)$	$\hat{R}(t)$	$\hat{h}(t)$	$\hat{R}(t)$	$\hat{h}(t)$
30		0.4705	0.8723	0.4925	0.8068	0.5754	0.6357	0.4280	0.9858
60		0.4708	0.8384	0.4918	0.7775	0.5734	0.6155	0.4292	0.9451
90	1.00	0.4698	0.8278	0.4904	0.7689	0.5715	0.6099	0.4286	0.9319
120		0.4690	0.8230	0.4893	0.7650	0.5703	0.6074	0.4280	0.9257
150		0.4684	0.8200	0.4884	0.7627	0.5694	0.6060	0.4276	0.9218
30		0.3796	0.9037	0.4039	0.8331	0.4913	0.6605	0.3361	1.0200
60		0.3817	0.8641	0.4051	0.7980	0.4913	0.6356	0.3390	0.9734
90	1.25	0.3816	0.8509	0.4045	0.7867	0.4901	0.6279	0.3393	0.9574
120		0.3813	0.8446	0.4038	0.7812	0.4893	0.6243	0.3392	0.9498
150		0.3810	0.8406	0.4033	0.7779	0.4887	0.6221	0.3391	0.9448
30		0.3047	0.9280	0.3298	0.8537	0.4175	0.6812	0.2628	1.0455
60		0.3080	0.8845	0.3323	0.8142	0.4191	0.6527	0.2664	0.9950
90	1.50	0.3086	0.8693	0.3324	0.8008	0.4187	0.6433	0.2673	0.9771
120		0.3086	0.8621	0.3322	0.7941	0.4183	0.6389	0.2675	0.9685
150		0.3086	0.8574	0.3320	0.7899	0.4180	0.6361	0.2676	0.9629
30		0.2439	0.9465	0.2686	0.8695	0.3536	0.6983	0.2049	1.0640
60		0.2476	0.9004	0.2718	0.8269	0.3563	0.6672	0.2086	1.0112
90	1.75	0.2486	0.8838	0.2725	0.8119	0.3565	0.6565	0.2098	0.9920
120		0.2489	0.8759	0.2726	0.8042	0.3564	0.6514	0.2102	0.9828
150		0.2491	0.8707	0.2726	0.7994	0.3563	0.6482	0.2105	0.9766

5 Real data applications

In this section, three real datasets related to different types of cancer patients have been utilized to illustrate the practical applicability of the proposed log transformed transmuted exponential distribution (LTTED). The first step involved evaluating whether the selected cancer datasets were appropriate for modeling with the proposed distribution. This suitability assessment was carried out by comparing the performance of the LTTE model with some other widely used lifetime distributions. The comparisons were made using well-established model selection criteria, including the Akaike information criterion (AIC), the Bayesian information icriterion (BIC), the KS statistic and the corresponding p-value obtained from the KS test. The performance of the proposed LTTE model was compared against four competing lifetime distributions: the transmuted exponential distribution (TED), the exponentiated exponential distribution (EED), the Weibull distribution (WD), and the log-exponential distribution (LED). These distributions were chosen due to their popularity and applicability in modeling lifetime data. The model comparison was based on specific criteria: lower values of AIC, BIC, and KS statistic indicated a better fit to the data, while higher p-values from the KS test suggested greater conformity of the data sets to the theoretical distribution. In general, a model demonstrating smaller AIC,

BIC, and KS values alongside a larger p-value was considered the most suitable. To provide a clear understanding, the PDFs of the competing models are detailed below:

- Exponential distribution (see Owoloko et al. (2015)) with PDF

$$f(x,\lambda,\alpha) = \lambda e^{-\lambda x}\left(1-\alpha+2\alpha e^{-\lambda x}\right), \qquad x \geq 0, \alpha, \lambda > 0.$$

- Exponentiated exponential distribution with PDF

$$f(x,\lambda,\alpha) = \alpha\lambda e^{-\lambda x}\left(1-e^{-\lambda x}\right)^{\alpha-1}, \qquad x \geq 0, \alpha, \lambda > 0.$$

- Weibull distribution with PDF

$$f(x,\alpha,\lambda) = \alpha\lambda\left(\lambda x\right)^{\alpha-1} e^{-(\lambda x)^{\alpha}}, \qquad x \geq 0, \alpha, \lambda > 0.$$

- Log exponential distribution (Maurya et al. (2016)) with PDF

$$f(x,\lambda) = \frac{\lambda e^{-\lambda x}}{\left(1+e^{-\lambda x}\right)\log 2}, \quad x \geq 0, \lambda > 0.$$

5.1 Data set-I: Breast cancer data

The first dataset represents the survival times of 121 patients diagnosed with breast cancer, collected from a large hospital over the period 1929 to 1938 and is mentioned in Lee (1992). This dataset has been previously analyzed and discussed in studies by Al-Kadim and Mahdi (2018). The dataset is as follows:
0.3, 0.3, 4.0, 5.0, 5.6, 6.2, 6.3, 6.6, 6.8, 7.4, 7.5, 8.4, 8.4, 10.3, 11.0, 11.8, 12.2, 12.3, 13.5, 14.4, 14.4, 14.8, 15.5, 15.7, 16.2, 16.3, 16.5, 16.8, 17.2, 17.3, 17.5, 17.9, 19.8, 20.4, 20.9, 21.0, 21.0, 21.1, 23.0, 23.4, 23.6, 24.0, 24.0, 27.9, 28.2, 29.1, 30.0, 31.0, 31.0, 32.0, 35.0, 35.0, 37.0, 37.0, 37.0, 38.0, 38.0, 38.0, 39.0, 39.0, 40.0, 40.0, 40.0, 41.0, 41.0, 41.0, 42.0, 43.0, 43.0, 43.0, 44.0, 45.0, 45.0, 46.0, 46.0, 47.0, 48.0, 49.0, 51.0, 51.0, 51.0, 52.0, 54.0, 55.0, 56.0, 57.0, 58.0, 59.0, 60.0, 60.0, 60.0, 61.0, 62.0, 65.0, 65.0, 67.0, 67.0, 68.0, 69.0, 78.0, 80.0, 83.0, 88.0, 89.0, 90.0, 93.0, 96.0, 103.0, 105.0, 109.0, 109.0, 111.0, 115.0, 117.0, 125.0, 126.0, 127.0, 129.0, 129.0, 139.0, 154.0

The total time on test (TTT) plot for the considered real dataset is presented in Figure 4. The plot indicates that the data exhibits an increasing hazard rate, which aligns with the hazard rate of the proposed model, making it potentially suitable for modeling such data. Al-Kadim and Mahdi (2018) introduced the exponentiated transmuted exponential (ETE) distribution to analyze this dataset and demonstrated that ETE outperformed three other lifetime models namely lognormal (LN), log-logistic (LL) and exponential distribution (ED) based on its lower AIC (1169.63) and BIC (1178.02). In this study, we further compare the proposed LTTE model with above mentioned four other competing models and observed that the proposed LTTE model exhibits the lowest AIC, BIC, and KS statistic values among the fitted models, including the ETE model, see Table 3. This strongly suggests that the LTTE model provides the best fit for this dataset and can be considered the most appropriate model for analyzing the survival times of breast cancer patients in this context.

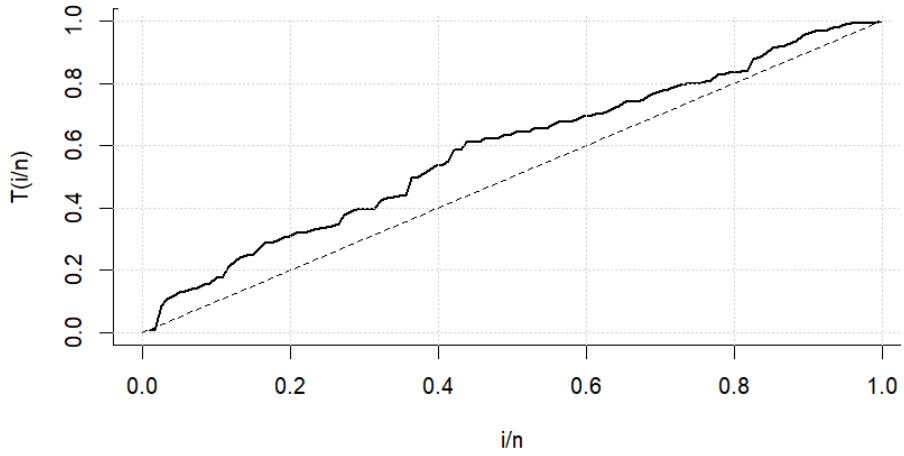

Figure 4: TTT plot of breast cancer dataset.

5.2 Data set-II: Bladder cancer data

The second dataset represents the remission times of 128 bladder cancer patients, extracted from Lee and Wang (2003) and is given by
0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.8,2 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34,,14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69

The TTT plot of the considered dataset (Figure 5) shows that the data is of nonmonotone hazard nature even though from Table 3, it can be seen that our proposed model fits the data very well. Khan et al. (2013) analyzed this dataset using the transmuted inverse Weibull (TIW) distribution, comparing it with the transmuted inverse Rayleigh (TIR), transmuted inverted exponential (TIE), and inverse Weibull (IW) distributions. Based on AIC, BIC, and KS values, they concluded that the TIW distribution provided the best fit. Kumar et al. (2015) proposed the DUS exponential distribution and demonstrated that it outperformed the TIW model, achieving lower AIC (834.044) and BIC (836.896) values. In this study, we extend the comparison to include the proposed LTTE distribution. Table 3 shows that the LTTE model achieves the lowest AIC and BIC values among all the models considered by Khan et al. (2013) and Kumar et al. (2015), establishing it as the best fit for this dataset.

Table 3: The values of the negative of log-likelihood $-\log L$, AIC, BIC, and KS value along with p-values for all the considered data sets.

Models	$-\log L$	AIC	BIC	KS	p-value
Data set-I: Breast Cancer Data					
LTTED	578.943	**1161.885**	1167.477	0.054	0.866
TED	578.978	1161.955	1167.547	0.068	0.638
EED	580.094	1164.187	1169.779	0.080	0.414
WD	579.024	1162.047	1167.639	0.060	0.779
LED	582.319	1166.639	1169.435	0.103	0.153
Data set-II: Bladder Cancer Data					
LTTED	310.064	**624.128**	627.506	0.080	0.382
TED	311.441	626.882	630.260	0.095	0.195
EED	310.156	624.311	627.689	0.073	0.511
WD	322.056	648.111	651.489	0.070	0.557
LED	318.921	639.843	641.532	0.078	0.411
Data set-III: Leukemia Dataset					
LTTED	412.784	**829.568**	835.272	0.185	0.113
TED	413.497	830.995	836.699	0.188	0.105
EED	413.078	830.155	835.859	0.165	0.201
WD	414.087	832.174	837.878	0.307	0.001
LED	414.962	831.923	834.775	0.293	0.002

5.3 Data set-III: Leukemia dataset

The dataset-III is taken from Abouammoh et al. (1994), which represents the ordered lifetimes of 40 patients suffering from leukemia, collected from one of the Ministry of Health hospitals in Saudi Arabia. The data set is
115 181 255 418 441 461 516 739 743 789 807 865 924 983 1024 1062 1063 1165 1191 1222 1222 1251 1277 1290 1357 1369 1408 1455 1478 1549 1578 1578 1599 1603 1605 1696 1735 1799 1815 1852

The TTT plot for this dataset is presented in Figure 6. The plot indicates that the data exhibits an increasing hazard rate, which suggests that models with an increasing hazard rate, such as the proposed LTTE model, may be particularly suitable for accurately representing the underlying data behavior. The results of this comparison, as summarized in the third part of Table 3, reveal that the LTTE model achieves the lowest AIC and BIC values among all the competing models. These lower values indicate that the LTTE model provides the best balance between model complexity and goodness-of-fit for the leukemia data set. Consequently, the LTTE model is determined to be the most appropriate and effective distribution for analyzing this dataset, reinforcing its potential as a robust tool for modeling lifetime data with an increasing hazard rate.

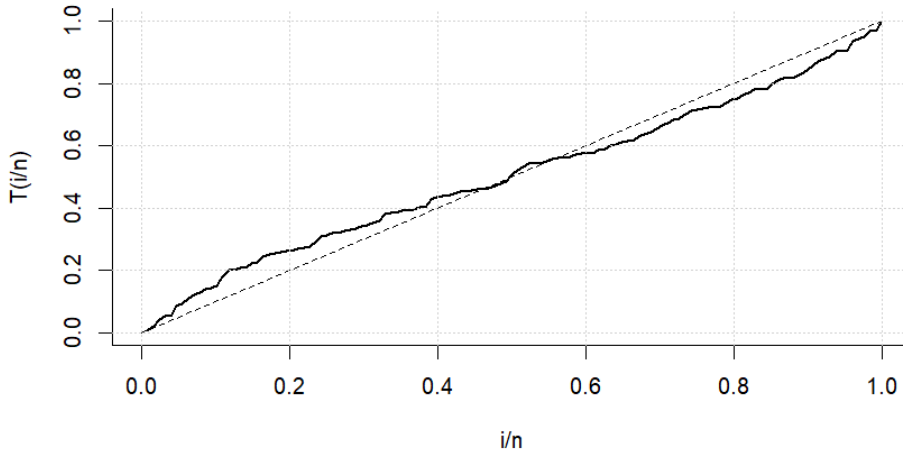

Figure 5: TTT plot of bladder cancer dataset.

6 Summary and conclusions

This research paper formulates a new lifetime probability model, named log transformed transmuted exponential by the extension of transmuted exponential via log transformation. The proposed distribution has an increasing hazard rate. Numerous significant properties of the new distribution are discussed including moments, skewness, kurtosis, order statistics, entropy, quantile function, reliability function and hazard rate. The maximum likelihood estimation procedure is employed to estimate the model parameters. Lastly, we considered three real-life datasets of cancer patients and four other distributions namely transmuted exponential, exponentiated exponential, log exponential and Weibull distributions. It is observed that the proposed LTTE distribution fits the considered datasets very well. The AIC, BIC, and KS-test values illustrate that proposed distribution is better than the above mentioned existing distributions and can be used as an alternate model for the cancer patients datasets.

Acknowledgments

We are grateful to the editors and reviewers for their valuable feedback.

References

Abouammoh, A. M., Abdulghani, S. A., and Qamber, I. S. (1994). On partial orderings and testing of new better than renewal used classes. *Reliability Engineering & System Safety*, **43**, 37–41.

Al-Kadim, K. A., and Mahdi, A. A. (2018). Exponentiated transmuted exponential distribution. *Journal of University of Babylon for Pure and Applied Sciences*, **26**, 78–90.

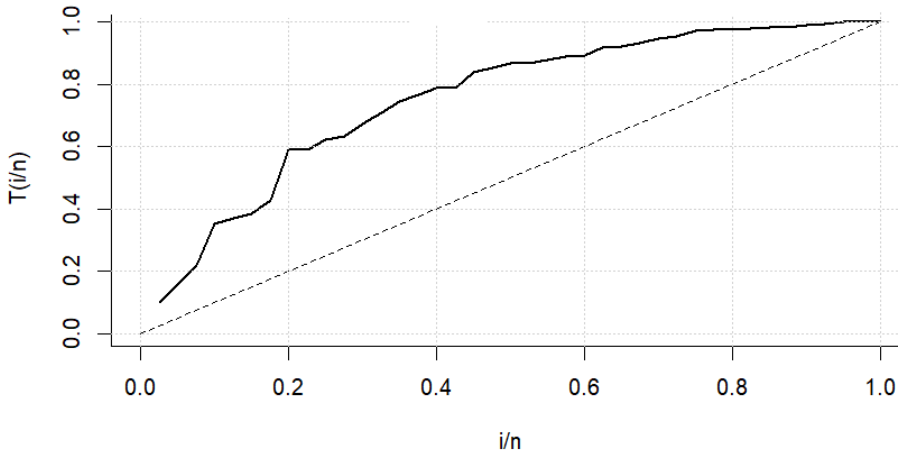

Figure 6: TTT plot of Leukemia dataset

Bhutani, M. A. N. I. S. H. A., Kochupillai, V. I. N. O. D., and Bakshi, S. (2004). Childhood acute lymphoblastic leukemia: Indian experience. *Indian J Med Paediatr Oncol*, **20**, 3–8.

Bonferroni, C. E. (1941). *Elementi di Statistica Generale*. Gili, Torino.

Bowley, A. L. (1926). *Elements of Statistics*. P. S. King & Son Ltd, London.

Elbatal, I., Diab, L. S., and Alim, N. A. (2013). Transmuted generalized linear exponential distribution. *International Journal of Computer Applications*, **83**, 29–37.

Gupta, R. C., Gupta, P. L., and Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics-Theory and Methods*, **27**, 887–904.

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., and Thun, M. J. (2007). Cancer statistics, 2007. *CA: A Cancer Journal for Clinicians*, **57**, 43–66.

Khan, M. S., King, R., and Hudson, I. (2013). Characterizations of the transmuted inverse Weibull distribution. *Anziam Journal*, **55**, C197–C217.

Kumar, D., Singh, U., and Singh, S. K. (2015). A method of proposing new distribution and its application to Bladder cancer patients data. *Journal of Statistics Applications and Probability Letters*, **2**, 235–245.

Lee, E. T. (1992) *Statistical Methods for Survival Data Analysis*, 2nd Edition, Wiley Interscience, New York.

Lee, E. T and Wang, J. (2003). *Statistical Methods for Survival Data Analysis*. John Wiley & Sons, New York.

Lorenz, M. O (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**, 209–219.

Maurya, S. K., Kaushik, A., Singh, R. K., Singh, S. K., and Singh, U. (2016). A new method of proposing distribution and its application to real data. *Imperial Journal of Interdisciplinary Research*, **2**, 1331–1338.

Mehrotra, R., and Yadav, K. (2022). Breast cancer in India: Present scenario and the challenges ahead. *World Journal of Clinical Oncology*, **13**, 209.

Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **37**, 25–32.

Owoloko, E. A., Oguntunde, P. E., and Adejumo, A. O. (2015). Performance rating of the transmuted exponential distribution: an analytical approach. *SpringerPlus*, **4**, 818.

Prakash, G., Pal, M., Odaiyappan, K., Shinde, R., Mishra, J., Jalde, D., ... & Bakshi, G. (2019). Bladder cancer demographics and outcome data from 2013 at a tertiary cancer hospital in India. *Indian Journal of Cancer*, **56**, 54–58.

Rényi, A. (1961). On measures of entropy and information. *In Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, Contributions to the Theory of Statistics*, **4**, 547–562.

R Core Team, (2024). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Sathishkumar, K., Chaturvedi, M., Das, P., Stephen, S., and Mathur, P. (2022). Cancer incidence estimates for 2022 & projection for 2025: result from National Cancer Registry Programme, India. *Indian Journal of Medical Research*, Medknow, **156(4& 5)**, 598–607.

Shaw, W. T. and Buckley, I. R. (2009). The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *arXiv preprint arXiv:0901.0434*.

Singh, S. K., Singh, U., and Yadav, A. S. (2014). Parameter estimation in Marshall-Olkin exponential distribution under type-I hybrid censoring scheme. *Journal of Statistics Applications & Probability*, **3**, 117–127.

Verma, E., Singh, S. K., and Yadav, S. (2025). A new generalized class of Kavya–Manoharan distributions: inferences and applications. *Life Cycle Reliability and Safety Engineering*, **14**, 79–91.