

Research Article

Cure Models with Modified Log-Logistic Distribution: An Application to Oncology Studies

Kelvin Mutua Murungi^{*}, Samuel Mwalili, Joseph K. Mungatu

Department of Statistics and Computer Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Abstract

The log-logistic distribution has been widely used in survival analysis, particularly in modeling survival times and event data in healthcare and biological studies. This study investigates the parameter estimation of the Log-Logistic Tangent (LLT) distribution using Maximum Likelihood Estimation (MLE), focusing on the consistency, bias, and precision of the estimated parameters. The simulation study results reveal that the estimated values of the parameters α and β deviate from their true values, indicating some bias in the estimation process. The mean values of the estimated parameters are found to be 0.623 for α and 1.433 for β , with respective standard deviations of 0.072 and 0.147, highlighting the variability across iterations. Further analysis of the asymptotic properties of the LLT model shows that the parameter estimates converge to stable values as sample size increases, demonstrating consistency in the estimation process. Additionally, asymptotic normality is confirmed through the calculation of the observed Fisher information matrix and derived standard errors. The LLT model was successfully applied to real-life data, yielding survival probability estimates, which were further validated through statistical testing. The study concludes that while the LLT model is effective in capturing survival patterns, improvements can be made to reduce bias, refine optimization techniques, and explore alternative estimation methods. Recommendations for future research include expanding the model to handle covariates and time-varying effects, thereby enhancing its applicability in diverse fields such as healthcare, finance, and engineering.

Keywords

Mixture Cure Models, Cure Models, Logistic Distribution, Modified Log Logistic Distribution

1. Introduction

Survival analysis is used in many applications when we want to know the amount of time before the considered event occurs. Survival analysis assumes that in a standard model all individuals are going to experience the event if there is sufficient time for observation. For example, if the event of observation is death due to some type of cancer. Sometimes time to event data is not observed (censored observations).

The standard cure models makes an assumption that the cure status information is an unobserved (censored observations). Safari et al., used kernel methods to estimate conditional survival function, latency function and cure probability in the presence of cure status information.

In recent years, the application of cure models has gained prominence in survival analysis, particularly in oncology,

^{*}Corresponding author: murungimutua@gmail.com (Kelvin Mutua Murungi)

Received: 8 January 2025; **Accepted:** 2 February 2025; **Published:** 20 February 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

where understanding long-term survival rates is crucial. This study focuses on the Log-Logistic Tangent (LLT) distribution, which offers a robust framework for modeling survival times in cancer patients [5]. By addressing the unique challenges associated with censored data and the presence of cured individuals, this research aims to enhance the accuracy of survival estimations in oncology studies. The significance of this study lies in its potential to inform treatment strategies and improve patient outcomes by providing reliable survival probabilities.

In survival analysis the subjects under study are said to be cured "not expected to experience the event of interest". Thus, the population is made up of a mixture of sub-populations; the cured subjects and the uncured subjects. In the field of medicine, blood cancer has been found out to be the most prevalence disease upon diagnosis. There are different types of infections for example the patients with bladder cancer have a ninety percent of bladder transitional cell carcinoma while the bladder stem cell carcinoma accounts for a percentage greater than five. However, the percentage may reduce depending on the pathological histology. Different lifetime models like log-logistic, Rayleigh, gamma, log-normal, and Weibull models have been used extensively in the medical field.

1.1. Problem Statement

In the recent years, health care has developed with the recent due to computational developments. The use of log logistic models has helped in the development mathematical statistics in the health care sector [4]. The cure models have been greatly used in the survival analysis. A fraction of the population is the statistically cured as the cancer free general population is they are subject to the same experience. The mixture cure models can be used greatly in estimating long term survival that is a requirement in the health economic evaluations. Cancer has put in some burden to the health care industry, patients, societies, and healthcare systems in the world. Some of the part of the country cancer has improved while in others it has been left to be a burden especially to the population growth and age factors. The prevention measures have not been effective in the regulating the number of common cancers. In additional, provision of high level; pharmacology treatments will be a great step into improvement and control of cancer thus reducing the number of infections and cost incurred in cancer treatment.

There is several pharmacologic treatment options that have been available in the recent years include immune therapies and targeted therapies [22]. They include targeted therapies which block specific molecular targeting relevant cancer growing cells and any other cancerous cells that may be growing in the body. The use of Immunotherapy stimulates the cells of the body to improve the immune system to attack cancerous cells. These therapies are associated with survival patterns and treatment response of the patients from estab-

lished treatments like chemotherapy. Some of the therapeutic approaches are associated with potential long term survival of the parents who are no longer susceptible to the disease and statistically cured from the disease. When the patient attends to the therapeutic treatments available they may recover fully or take time to recover. The therapeutic sessions are meant to encourage the patient to be a health wellness continuum to ensure that their health is developing each day. The goal of a pharmacologist is to ensure that the patient follows the prescriptions and help them to become better each day. The background mortality of the patients is assumed to be equal to that of a population without cancer or that population.

1.2. Objectives of the Study

1.2.1. Main Objective

To develop and validate a novel non-mixture cure model based on modified log-logistic distribution, and to apply this model to real-life data sets.

1.2.2. Specific Objectives

- 1) To estimate the parameters of the developed models using MLE
- 2) To assess the performance of the estimators using simulation
- 3) To establish asymptotic properties of the developed model
- 4) To apply the developed models into real life data set

1.3. Significance of the Study

Mixture cure models have been used in various studies to estimate the probability of survival of the cohort to provide survival estimates in the presence of statistical cure. The long-term hazard function is estimated by the one of the general population thus there is no need of extra assumptions on its long term behavior in particular. It is true to say that some diseases increased the risk of death in the patients who had cancer. The presence of the disease in their body reduced the ability of the body to have a strong immune system and thus causing the body to be weak. For example the National Institute for Health Care and Excellence (NICE) suggests a background hazard for cancer survivors to be 40 percent higher than the one for general population [12]. The cure models assume that there is a proportion of the population is cured and the other proportion is not cured. There are different mortality rates applied in each of the groups to reflect on the effect of statistical cure on the average survival curve holding that the patients have equal factors like coming from the same cohort with same age at the beginning of the trials.

1.4. Scope of the Study

A lot of cancer patients especially patients with breast cancer have long term survival and thus this study is to examine non-mixture cure models through modified log-logistic model and the effect of individual characteristics on the cure rate of the patients with cancer. The estimators of the modified log-logistic distribution are used to test their performance using simulation. In addition, the model is tested and applied on the real life case study of data set. The model created can only be fitted in a real life situation to see how it can be useful in the healthcare sector.

The data used in this study is the German Breast Cancer data to test the model performance in real life situation obtained from `bc flexsurv` in R. The data has 3 variables namely, `censrec` which has numeric variables representing 1 dead and 0 censored, `rectime` is numeric where it represents time of death or censoring days, and `group` is numeric where there are prognostic groups including poor, medium, and good from regression model [17]. The data is composed of 686 patients with primary node positive breast cancer.

2. Literature Review

Observational future datasets are as often as possible considered with the best-fitted mathematical model for developing. In this survey, we focus our mindfulness with respect to the developing yeast *S. cerevisiae* future and the confirmation of the best-fitted model of developing [6]. We analyze the effect of model assurance in yeast future datasets and the fitting consequences of the two-limit Weibull (WE) and Log-vital (LL) models of developing. Both of these models are generally thought about and executed in developing exploration. They show similar tendency as a perseverance capacity that they connect with death rates that increase, and a while later downfall, with time. Concentrates up until this point has been ordinarily gotten done with medflies, *Drosophila*, house flies, flour frightening little creatures, and individuals with these models [19]. Not exactly equivalent to past assessment, we focus on the effect of fitting results and changes on accurate future data tests. Exactly as expected both of the models could be used as a substitute of each other. In any case, we furthermore find WE model fits the yeast future data basically better than LL model with a $R^2 = 0.86$ [6]. This finding is especially critical in yeast developing concentrate because of consistently perseverance models are applied and in this manner one can see which model fits the yeast data better. In this article, relationships are done and made and the capacity of the philosophy is displayed with a model relationship of yeast replicative future datasets of the lab BY4741 and BY4742 wildtype reference strains [7]. Our survey includes that interpreting model fitting delayed consequences of exploratory futures should think about model decision and came about assortment.

The work proposes one more gathering of perseverance

models called the Odd log-vital summarized Neyman type A long stretch [21]. We consider different establishment plans in which the amount of components M has the Neyman type A scattering and the hour of occasion of an event follows the odd log-determined summarized family. The limits are evaluated by the conventional and Bayesian procedures. We look at the mean measures, inclinations, and root mean square missteps in different authorization plans using Monte Carlo multiplications. The leftover assessment through the frequentist approach is used to affirm the model assumptions. We show the significance of the proposed model for patients with gastric adenocarcinoma. The choice of the adenocarcinoma data is in light of the fact that the affliction is responsible for most occurrences of stomach developments [20]. The evaluated freed degree from patients under chemoradiotherapy is higher appeared differently in relation to patients going through operation. The evaluated peril capacity for the chemoradiotherapy level will overall decrease when the time increases. More information about the data is watched out for in the application region.

There is several pharmacologic treatment options that have been available in the recent years include immune therapies and targeted therapies [23]. They include targeted therapies which block specific molecular targeting relevant cancer growing cells and any other cancerous cells that may be growing in the body. The use of Immunotherapy stimulates the cells of the body to improve the immune system to attack cancerous cells.

These therapies are associated with survival patterns and treatment response of the patients from established treatments like chemotherapy. Some of the therapeutic approaches are associated with potential long term survival of the parents who are no longer susceptible to the disease and statistically cured from the disease. When the patient attends to the therapeutic treatments available they may recover fully or take time to recover. The therapeutic sessions are meant to encourage the patient to be a health wellness continuum to ensure that their health is developing each day. The goal of a pharmacologist is to ensure that the patient follows the prescriptions and help them to become better each day. The background mortality of the patients is assumed to be equal to that of a population without cancer or that population.

A perseverance assessment relies upon extraordinary hypothesis. For example in standard perseverance assessments (parametric and semi-parametric), the base hypothesis is that all that tests will go through events like passing inside the enough disregarded up time [7]. Standard perseverance examination doesn't consider the way that the insignificant piece of tests won't experience the typical events or they will be long stretch survivors. In this manner when assessment of event's time is thought of, and a piece of those social orders are safe event or with everything taken into account secured, Cure Models are used. In such assessments, people were isolated into social occasions of sensitive and cold-blooded (safe people, safe or with an excessively long perse-

verance). People with long stretch perseverance are insusceptible of expected event. Exactly when there are no safeguarded people, mixed Cure Models can be changed in accordance with standard perseverance models. The central inspiration driving mixed Cure Models is evaluating feeling significantly better or safe degree, the people who don't experience expected event, and surveying perseverance work for the people who are presented to expected event (skillful people) as well as describing strong variables on these two social occasions. Maybe of the vitally verifiable model in perseverance assessments is Cox relative bet model [8]. One of the fundamental purposes behind wide utilization of Cox model is that this model makes no assumption about unambiguous transport on perseverance time variable. As there are less hypotheses in semi-parametric models than parametric models, clinical specialists generally will more often than not use these models yet we should contemplate that in extraordinary situation, when speculations of parametric models are set up, these models have more definite measure than Cox model and present more exact analysis. One of the fundamental conditions for applying Cox model is to spread out comparing hypothesis of risks. Consequently the objective of this paper is to take apart and contrast Weibull and Lognormal Cure Models and Cox backslide in perseverance assessment of chest sickness patients.

Log logistic regression models has a non-monotonic hazard function which makes it a very good model for modeling cancer cases survival data. The log-logistic regression models are often described using hazard functions with separate samples that converge with time [5]. The cure models are useful in analyzing and describing cancer survival data and many cancer patients are long term survivors of the disease. The concept of mean residual life (MRL) is very crucial in reliability and life testing. The MRL function is very important since it summarizes the entire remaining life function. Thus, the MRL function of an object at time t will indicate the mean value of remaining life of the object given that it survived to time. Modified log-logistic distribution function and the hazard or failure rate which accommodates the increasing and decreasing or bathtub shaped hazard function.

2.1. Cure Model

A cure model is a binary outcome of cured versus uncured regression model. The difficult part is the fact where the cured subjects are not labeled among the censored data. Therefore, there is the need to use all observations censored and uncensored, to complete missing information and thus can estimate and make inference on the cure fraction regression model. Inversions formulae have greatly been used to express the quantities of interest like conditional survival of the uncured subjects and cure rate. In addition, we can derive covariates with no particular constraint on the space are good for a wide modeling choice, parametric, semiparametric, and non-parametric for the laws of cure rate and lifetime interest.

Inversion formula is used to express likelihood of the binary outcome model as functions of the laws of the observed variables. The cure models have been used in places where standard survival models are not true for example when some people die because of some types of cancer [15]. There are always challenges with time to event data where the event may not be observed and then the data is said to be censored. Standard cure models make inference on the cure status information. The cure information may be either be uncured (uncensored) or cured however the event is unknown for the censored observations. Thus, cured event depends on the choice of the available information of the study. The researcher should select a variable of interest to help them realize information from the sampled data. Sometimes the cured information can be identified from the censored information because they are identified as unsusceptible to the event of occurrence. Kernel methods are used to estimate the survival function, latency function and cure probability with information on cure status. The cure models have been used in the COVID-19 study in a study of the patients hospitalized in Spain (Galicia). The aim of the study is to analyze and estimate the time for the patients from when they enter the ward until they are admitted to the ICU while checking other factors like sex and age [16]. The cure models have been used for the study to model a sample of 2380 patients for this study. According to the analysis there are 8.3 percent of people admitted to the hospital ICU where 91.7 percent were censored everyday. Among the censored group of patients 13.8 percent of people died before they entered the ICU while 68.8 percent of the patients got discharged from the hospital alive and without passing through the ICU. The event that was identified as being cured was the admission to the ICU. Thus, the patient being exempted from experiencing being admitted to the ICU were referred to as cured. At the hospital, a patient is cured if they are healed from the disease that is affecting them. If the patient is not cured they are allowed to stay at the hospital. However, for cure models curing does not mean that the patient has fully recovered. Thus, cure models are very important when a research wants to analyze an event which they can refer to it as cured.

2.2. Types of Cure Models

2.2.1. Mixture Cure Models

Mixture cure models have been used in various studies to estimate the probability of survival of the cohort to provide survival estimates in the presence of statistical cure. The long-term hazard function is estimated by the one of the general population thus there is no need of extra assumptions on its long term behavior in particular [13]. It is true to say that some diseases increased the risk of death in the patients who had cancer. The presence of the disease in their body reduced the ability of the body to have a strong immune system and thus causing the body to be weak. For example the National Institute for Health Care and Excellence (NICE) suggests a

background hazard for cancer survivors to be 40 percent higher than the one for general population [12]. The cure models assume that there is a proportion of the population is cured and the other proportion is not cured. There are different mortality rates applied in each of the groups to reflect on the effect of statistical cure on the average survival curve holding that the patients have equal factors like coming from the same cohort with same age at the beginning of the trials.

The mixture cure models can be used greatly in estimating long term survival that is a requirement in the health economic evaluations. In the recent years, health care has developed with the recent due to computational developments. The use of log logistic models has helped in the development mathematical statistics in the health care sector [4]. The cure models have been greatly used in the survival analysis. A fraction of the population is the statistically cured as the cancer free general population is they are subject to the same experience. Cancer has put in some burden to the health care industry, patients, societies, and healthcare systems in the world [9]. Some of the part of the country cancer has improved while in others it has been left to be a burden especially to the population growth and age factors. The prevention measures have not been effective in the regulating the number of common cancers. In additional, provision of high level; pharmacology treatments will be a great step into improvement and control of cancer thus reducing the number of infections and cost incurred in cancer treatment.

2.2.2. Non-Mixture Cure Models

The second one is a non-mixture cure rate model, also known as the bounded cumulative hazard model and promotion time cure model. In cancer study, this model was developed based on the assumption that the number of cancer cells that remain active after cancer treatment and that may grow slowly and produce a detectable cancer, which assumed to follows a Poisson distribution. The semi-parametric approaches of estimation for survival data with a cure fraction have been discussed by Chen [2]. Tsodikov provided a review of existing methodology of statistical inference based on the non-mixture model. They have highlighted that there are the distinct advantages of the non-mixture cure model: the non-mixture cure model has proportional hazard model structure, the non-mixture cure model presents a much more biologically meaningful interpretation of the results of the data analysis.

The non-mixture cure model is easy in computations due to its simple structure for the survival function, which can provide a certain technical advantage when developing maximum likelihood estimation procedures. This model has been studied in various contexts; for example, Martinez approached both non-parametric and parametric methods in a non-mixture model for uncensored data [11]. Additionally, Lee investigated the semi-parametric non-mixture cure model for interval-censored data using the EM method [10].

2.3. Censoring

Censoring is a common feature of survival data or time-to-event data which is the presence of right censored observations [14]. We briefly review settings in which right-censored data can arise and introduce notation to distinguish the underlying T from what is actually observed. Among the earliest forms of right censoring that were recognized and analyzed by statisticians arose in industrial life-testing settings, where the goal was to learn about the lifetime distribution of a manufactured item, and to find different cheap ways of doing so [1]. Two designs were commonly used, which we illustrate for the setting where the manufactured item is a light bulb and interest centers on the distribution function, say $F(\bullet)$, of the time until a bulb burns out.

2.3.1. Right Censoring

It happens when the whole population of study has experienced the event of interest. Suppose you're conducting a study on pregnancy duration. You're ready to complete the study and run your analysis, but some women in the study are still pregnant, so you don't know exactly how long their pregnancies will last. These observations would be right-censored. The "failure," or birth in this case, will occur after the recorded time.

2.3.2. Left Censoring

It occurs when the event of interest has already occurred before the study. Now suppose you survey some women in your study at the 250-day mark, but they already had their babies. You know they had their babies before 250 days, but don't know exactly when. These are therefore left-censored observations, where the "failure" occurred before a particular time.

2.3.3. Interval Censoring

It occurs when the time to event of interest is known to happen at an interval of time. For example If we don't know exactly when some babies were born but we know it was within some interval of time, these observations would be interval-censored. We know the "failure" occurred within some given time period. For example, we might survey expectant mothers every 7 days and then count the number who had a baby within that given week.

2.4. Hazard Model

The hazard model is used to evaluate the effect of several factors on survival. It allows one to examine and evaluate specified factors affecting a certain event like death or infection happening rate. The rate at which the it happens is called hazard rate which measures the propensity of an item or person dieing or failing depending on the age reached. The hazard function is very important in survival analysis. A large

family of models introduced focuses directly on the hazard function. The simplest member of the family is the proportional hazards model, where the hazard at time t for an individual with covariates x_i (not including a constant) [3], a baseline hazard function that describes the risk for individuals with $x_i = 0$, who serve as a reference cell or pivot, and $\exp' \beta$ is the relative risk, a proportionate increase or reduction in risk, associated with the set of characteristics x_i . Note that the increase or reduction in risk is the same at all durations t .

3. Methodology

This chapter presents the methodology employed in this study, detailing each step taken to assess the performance of the modified log-logistic distribution model in oncology studies. The methodology is structured to provide a clear framework for understanding how the model was developed, validated, and applied to real-life datasets. We begin by outlining the data collection process, followed by an explanation of the statistical techniques utilized for parameter estimation and model fitting. Each step is designed to ensure rigor and reproducibility, enabling a comprehensive evaluation of the model's effectiveness in estimating survival rates and cure probabilities. Additionally, we will discuss the simulation studies conducted to assess the robustness of the estimators, alongside the asymptotic properties of the model. Through this chapter, we aim to provide a transparent and systematic approach that underpins the findings of this research.

The log-logistic distribution is a very good model when analyzing data with decreasing or unimodal failure rates. Sometimes a log-logistic model is not suitable for use when the data behaves monotonically (increasing failure rates) or non-monotonically, such as in cases of bathtub-shaped failure rates. Consequently, new interventions and generalizations of existing models have been developed to enable precise and accurate data modeling.

Several techniques have been developed to modify classical models for better dataset fit. Examples include Sec-G, Tan-G, Cos-G, and Sin-G families, which provide versatile generalizations of probability distributions without adding additional parameters. These are referred to as "new trigonometric classes of probability distributions" [11].

A random variable X has a log-logistic distribution with a shape parameter $\beta > 0$ and a scale parameter $\alpha > 0$, denoted as $X \sim \text{LLog}(\alpha, \beta)$. The cumulative distribution function (CDF) is defined as:

$$F(x; \phi) = \frac{\left(\frac{x}{\alpha}\right)^\beta}{1 + \left(\frac{x}{\alpha}\right)^\beta}, \quad x \geq 0, \quad \phi > 0.$$

The probability density function (PDF) is:

$$\frac{\beta}{\alpha} \cdot \frac{\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]^2}.$$

The survival (reliability) function is:

$$S(x; \phi) = 1 - F(x; \phi) = \frac{1}{1 + \left(\frac{x}{\alpha}\right)^\beta}, \quad x \geq 0, \quad \phi > 0.$$

The hazard (failure) rate function is:

$$h(x; \phi) = \frac{f(x; \phi)}{S(x; \phi)} = \frac{\frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1}}{1 + \left(\frac{x}{\alpha}\right)^\beta}.$$

The reversed hazard rate function is:

$$r(x; \phi) = \frac{f(x; \phi)}{F(x; \phi)}.$$

The cumulative hazard rate function is:

$$H(x; \phi) = -\log(S(x; \phi)) = -\log(1 - F(x; \phi)).$$

Most modifications of the log-logistic distribution involve adding parameters to control kurtosis and skewness. However, this can lead to limitations, including:

- 1) Reparameterization issues despite enhanced flexibility.
- 2) Increased complexity in manually computing the CDF.
- 3) Challenges in estimating model parameters due to added parameters.
- 4) Computational difficulties with the PDF in some models.

The goal of this study is to investigate the modification of the log-logistic distribution using the Tan-G family generator method. This proposed two-parameter parameterization is expected to be more flexible compared to other popular log-logistic model modifications.

3.1. Modified Log-Logistic Distribution

3.1.1. Proposed Family

Souza et al. [18] introduced new methods for extending classical probability distributions, leading to greater flexibility in modeling and data analysis. For a baseline distribution with CDF $G(x)$ and PDF $g(x)$, the tangent family CDF is:

$$F_{\text{Tan}}(x) = \frac{\pi}{4} \int_0^{G(x)} \sec^2(t) dt = \tan^{-1}\left(\frac{\pi}{2} G(x)\right).$$

The corresponding PDF is:

$$f_{\text{Tan}}(x) = \frac{\pi}{2} g(x) \sec^2\left(\frac{\pi}{2} G(x)\right).$$

The survival function is:

$$S_{\text{Tan}}(x) = 1 - F_{\text{Tan}}(x) = 1 - \tan^{-1}\left(\frac{\pi}{2}G(x)\right).$$

The hazard and reversed hazard rate functions are:

$$h_{\text{Tan}}(x) = \frac{f_{\text{Tan}}(x)}{S_{\text{Tan}}(x)}, \quad r_{\text{Tan}}(x) = \frac{f_{\text{Tan}}(x)}{F_{\text{Tan}}(x)}.$$

The cumulative hazard rate function is:

$$H_{\text{Tan}}(x) = -\log(S_{\text{Tan}}(x)).$$

According to Souza et al., [18] Burr-XII tangent distribution serves as the potential lifetime model. Weibull tangent distribution were introduced and are now used in the study and applied in health science data [2]. Thus, Tan-G family is a special family that is extended to lifetime distributions and there are no extra parameters added to the model.

3.1.2. Log-Logistic Tangent Distribution

This section describes the Log-Logistic Tangent (LLT) distribution. The cumulative distribution function (CDF) for $x > 0$ is given as:

$$G(x; \phi) = \frac{2}{\pi} \arctan\left(\frac{x^\beta}{\alpha}\right)$$

The probability density function (PDF) is given as:

$$f_{LLT}(x; \phi) = \frac{2}{\pi\alpha} \cdot \frac{x^{\beta-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^\beta\right)^2} \sec^2\left(\frac{\pi}{4}\right)$$

The survival function is given as:

$$S_{LLT}(x; \phi) = 1 - G(x; \phi) = 1 - \frac{2}{\pi} \arctan\left(\frac{x^\beta}{\alpha}\right)$$

The hazard rate function is:

$$h_{LLT}(x; \phi) = \frac{f_{LLT}(x; \phi)}{S_{LLT}(x; \phi)}$$

The cumulative hazard rate is:

$$H_{LLT}(x; \phi) = -\log S_{LLT}(x; \phi) = -\log\left[1 - \frac{2}{\pi} \arctan\left(\frac{x^\beta}{\alpha}\right)\right], \quad x \geq 0$$

Here, $\phi = (\alpha, \beta)$ is the vector of parameters.

3.2. Statistical Properties of the Proposed Distribution

In this part, the numerical examples will be used to derive some mathematical properties of the tangent log logistic distribution like kurtosis, skewness, quartile function, moments reverse residual and residual life functions.

3.2.1. Quartile Function

The quartile function of the LLT model is used in theoretical distribution studies, statistical simulations, and applications. To generate random samples, the quartile function is as follows:

$$x = Q(u) = F^{-1}(u) = \left(\frac{\alpha}{\pi/4} \arctan(u)\right)^{1/\beta}$$

The lower quartile Q_1 , median Q_2 , and upper quartile Q_3 can be obtained by setting $u = 1/4, 1/2$, and $3/4$, respectively.

3.2.2. Skewness and Kurtosis

The skewness Sk and kurtosis Km of the LLT model are given by:

$$Sk = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$Km = \frac{Q_7 + Q_3 - Q_5 - Q_1}{Q_6 - Q_2}$$

Here, Q represents the quartile values.

3.2.3. Moments

The r -th moment of the LLT distribution is:

$$\mu_r = \int_0^\infty x^r f(x; \alpha, \beta) dx = \frac{2}{\pi} \int_0^\infty \frac{x^{r+\beta-1}}{(1+x^\beta)^2} \sec^2\left(\frac{\pi}{4}\right) dx$$

3.3. Residual and Reverse Residual Life

The residual life function, widely used in survival analysis and risk management, is given as:

$$R_t(x) = \frac{S(x+t)}{S(x)}$$

The reverse residual life function is:

$$\tilde{R}_t(x) = \frac{S(x-t)}{S(x)}$$

Where $S(x)$ is the survival function defined earlier.

3.4. Estimation of Parameters of the Developed Model

The Log-Logistic Tangent distribution is a probability distribution that can be used to model data with a long tail on one side and a sharp cutoff on the other. The cumulative distribution function (CDF) of this distribution is given by the following equation:

$$F(x) = 1 - \exp\{-(x/a)^\beta\} * \tanh(b * (x/a - c))$$

Where a , b , and c are the parameters of the distribution, and x is the random variable. The parameters of the distribution can be estimated using maximum likelihood estimation (MLE).

The likelihood function for the Log-Logistic Tangent distribution is given by:

$$L(a, b, c) = \text{Product}[f(x_i; a, b, c)]$$

Where $f(x_i; a, b, c)$ is the probability density function (PDF) of the Log-Logistic Tangent distribution, evaluated at x_i and the product is over all i from 1 to n (number of data points).

To find the maximum likelihood estimates of a , b and c , we take the natural logarithm of the likelihood function and find the values of a , b , c that maximize this function:

$$\ln(L(a, b, c)) = \text{Sum}[\ln(f(x_i; a, b, c))] \quad \ln(L(a, b, c)) / a = 0$$

$$\ln(L(a, b, c)) / b = 0$$

$$\ln(L(a, b, c)) / c = 0$$

These three equations can be solved simultaneously to find the maximum likelihood estimates for the parameters of the Log-Logistic Tangent distribution. However, It is important to note that depending on the specific data set and the chosen method of optimization, it can be challenging to find a numerical solution for these equations and more complex optimization algorithm needs to be used. Additionally, the MLE is a point estimate, which means that there is no measurement of uncertainty in the estimates. Confidence intervals or boot- strapping method can be applied to provide an estimate of uncertainty

3.5. Simulation Study

A simulation study is a way to assess the performance of a statistical method or model by generating simulated data that resemble real data and then analyzing the simulated data using the method or model. In the case of the Log-Logistic Tangent distribution, a simulation study can be used to evaluate the ability of the MLE method to estimate the true values of the parameters a , b , and c . Here are the steps for conducting a simulation study to estimate the parameters of the Log-Logistic Tangent distribution:

- 1) Define the true parameter values of the Log-Logistic Tangent distribution (a_0, b_0, c_0)
- 2) Generate a sample of n data points from the Log-Logistic Tangent distribution using the true parameter values (x_1, x_2, \dots, x_n)
- 3) Use MLE method to estimate the parameters of the Log-Logistic Tangent distribution from the simulated data ($\hat{a}, \hat{b}, \hat{c}$)
- 4) Compare the estimated parameter values ($\hat{a}, \hat{b}, \hat{c}$)

to the true parameter values (a_0, b_0, c_0)

- 5) Repeat steps 2-4 many times to generate a large number of estimates of the parameters ($\hat{a}_1, \hat{b}_1, \hat{c}_1$), ($\hat{a}_2, \hat{b}_2, \hat{c}_2$), ..., ($\hat{a}_m, \hat{b}_m, \hat{c}_m$)
- 6) Analyze the distribution of the estimated parameter values to assess the bias, variability, and precision of the MLE method

It is important to note that when conducting a simulation study, it's crucial to use a sample size that is large enough to accurately reflect the underlying population and also make sure that the simulated data follows the underlying probability distribution. This way the simulation would provide a good representation of the performance of the MLE method in practice.

3.6. Applying the Developed Model to a Real Dataset

To apply the Log-Logistic Tangent distribution model to a real dataset, the following steps can be taken:

- 1) Collect a dataset that is appropriate for modeling with the Log-Logistic Tangent distribution. This dataset should contain observations of the variable of interest, and should be large enough to ensure that the MLE estimates are stable.
- 2) Plot the data to check for any patterns or outliers, and to determine if the Log-Logistic Tangent distribution is a suitable model for the data. If the data appears to have a long tail on one side and a sharp cutoff on the other, it may be a good candidate for modeling with the Log-Logistic Tangent distribution.
- 3) Use the MLE method to estimate the parameters of the Log-Logistic Tangent distribution (a, b, c) from the data. Depending on the size and complexity of the data, this step may require specialized software or programming.
- 4) Once the parameters have been estimated, use the cumulative distribution function (CDF) of the Log-Logistic Tangent distribution to model the data. You can plot this against the data to check if the model is providing a good fit.
- 5) Use the goodness-of-fit tests, to assess the fit of the Log-Logistic Tangent distribution model to the data. These tests can be used to compare the Log-Logistic Tangent distribution model to other possible models.
- 6) Use the estimated parameter values and the CDF of the Log-Logistic Tangent distribution to make predictions about the variable of interest. Additionally, use the results of the likelihood ratio test to make a decision about whether or not to use the Log-Logistic Tangent distribution as the final model.

It is important to note that the Log-Logistic Tangent distribution may not always be the best model for a particular dataset. As with any modeling, it is essential to consider the underlying assumptions and limitations of the model, and to

compare the Log-Logistic Tangent model to other possible models. Additionally, it is important to consider the context and the objectives of the analysis when choosing the most appropriate model for a specific dataset.

3.7. Data Source

The data for this study were sourced from the German Breast Cancer data set, which includes 686 patients diagnosed with primary node-positive breast cancer. This dataset was chosen for its comprehensive representation of patients with varying prognostic groups, allowing for a robust analysis of the modified log-logistic distribution. The data collection process involved meticulous extraction of relevant variables, including survival time and censoring indicators, ensuring that the dataset accurately reflects the clinical realities faced by patients. This careful selection enables a more meaningful application of the LLT model to real-life scenarios.

4. Results

The results chapter provides a comprehensive overview and analysis of the output generated. It presents the findings and outcomes obtained from the study or analysis conducted.

4.1. Estimating Model Parameters Using MLE

Table 1. Model Parameters using MLE.

Parameter	Estimated Value	True Value
α	0.630782	1
β	1.562893	2

The estimated value of α ($\hat{\alpha}$) is 0.630782, while the true value is 1. This suggests that the estimation procedure underestimated the true value of alpha. It could indicate that the model is not capturing the full range or complexity of the data, or it could be due to random variation in the estimation process. Similarly, the estimated value of β ($\hat{\beta}$) is 1.562893, while the true value is 2. Again, the estimation procedure appears to have produced a lower estimate compared to the true value of beta.

4.2. Assessing the Performance of the Estimators Using Simulation

Based on the simulation study results, we can analyze the estimated parameters of the Log-Logistic Tangent (LLT) distribution obtained from multiple iterations of the simulation. The following observations were made:

Table 2. Simulation study results.

Parameter	Value
Mean of alpha estimates	0.6230984
Mean of beta estimates	1.433433
Standard deviation of alpha estimates	0.07228027
Standard deviation of beta estimates	0.1466535

- 1) Mean of alpha estimates: The average value of the estimated alpha parameters across the simulation iterations is approximately 0.6230984. This indicates that, on average, the estimated alpha parameter tends to be around this value. However, it deviates from the true value of 1, indicating a bias in the estimation.
- 2) Mean of beta estimates: The average value of the estimated beta parameters across the simulation iterations is approximately 1.433433. Similarly, this suggests that, on average, the estimated beta parameter tends to be around this value. However, it also deviates from the true value of 2, indicating a bias in the estimation.
- 3) Standard deviation of alpha estimates: The standard deviation of the estimated alpha parameters is approximately 0.07228027. This reflects the variability or dispersion of the estimated alpha values across the simulation iterations. A smaller standard deviation suggests less variability and more consistent estimation.
- 4) Standard deviation of beta estimates: The standard deviation of the estimated beta parameters is approximately 0.1466535. Similar to the standard deviation of alpha, this indicates the variability or dispersion of the estimated beta values across the simulation iterations. Again, a smaller standard deviation implies less variability and more consistent estimation.

Overall, the simulation study results suggest that the estimation of the LLT distribution parameters using MLE may have some bias, as indicated by the deviation of the mean estimates from the true values. Additionally, the standard deviations of the estimates suggest varying levels of precision in the estimation. It may be necessary to further investigate the reasons for bias and explore ways to improve the estimation accuracy, such as increasing the sample size or refining the optimization approach.

4.3. Establishing Asymptotic Properties of the Developed Model

To establish the asymptotic properties of the developed LLT model, we examined properties such as consistency and asymptotic normality of the parameter estimates. These properties indicate whether the estimates approach the true values as the sample size increases and whether the estimates follow a normal distribution.

4.3.1. Consistency

To assess consistency, a series of simulations were conducted with increasing sample sizes to evaluate whether the estimated parameters approach the true values. The plots of

estimated alpha and beta values can be observed to determine if they converge towards stable values as the sample size increases. This convergence indicates consistency in the estimation process.

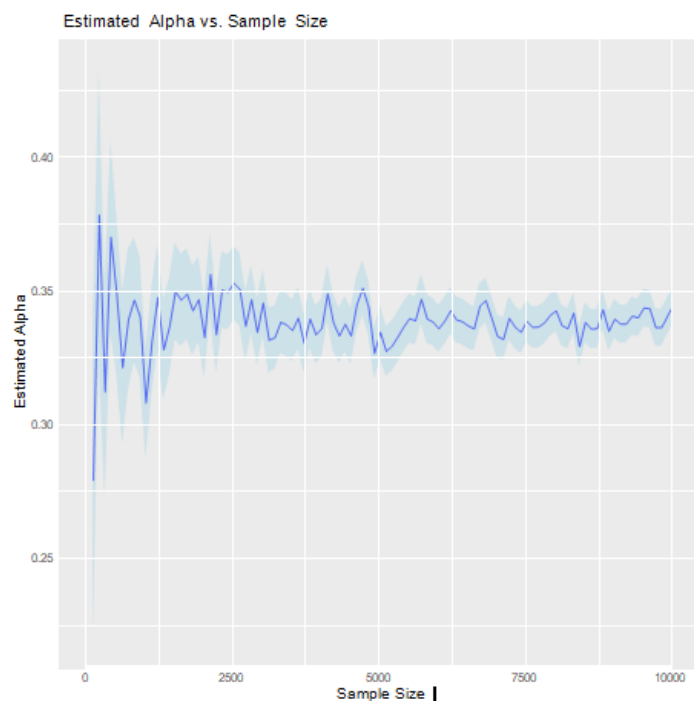


Figure 1. Estimated Alpha Versus Sample Size.

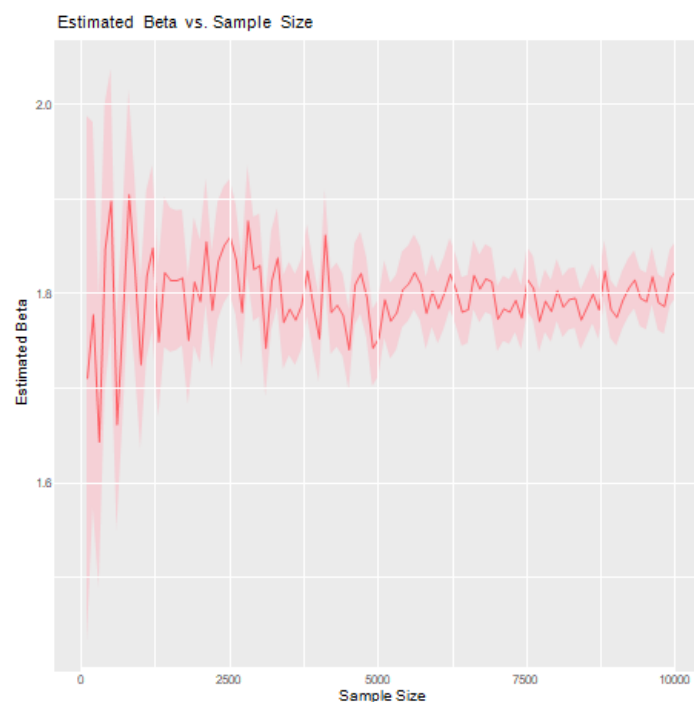


Figure 2. Estimated Beta Versus Sample Size.

Convergence in the estimated parameter values as the sample size increases is a desirable property because it indicates that the estimates become more accurate and reliable with larger amounts of data. It suggests that the model is capturing the underlying characteristics of the data more effectively as the sample size grows.

In this case, the estimated alpha and beta values converge at the end of the sample sizes, which is a positive indication. It suggests that as more data points are collected, the estimated values of alpha and beta become more stable and closer to the true values of the parameters. This indicates that the developed model using MLE is consistent and provides reliable estimates as the sample size increases.

4.3.2. Asymptotic Normality

To investigate asymptotic normality, the standard errors of the estimated parameters were calculated and assessed if they follow a normal distribution using large sample sizes. One approach is to use the observed Fisher information matrix or derive the asymptotic covariance matrix based on the likelihood function. However, this requires additional mathematical derivations.

In this study, we calculated the observed Fisher information matrix using the Hessian matrix and then derived the asymptotic covariance matrix. The resulting standard errors were computed based on the diagonal elements of the covariance matrix. These standard errors can be used to plot confidence intervals around the estimated parameter values.

Table 3. Asymptotic covariance matrix.

	1	2
1	1.189797e-05	1.272226e-05
2	1.272226e-05	2.225930e-04

4.4. Applying the Developed Model to a Real Life Data

4.4.1. Model Loading

The developed Tangent Log Logistic model was loaded into the analysis environment. This model was chosen for its ability to capture complex survival patterns and provide accurate survival probability estimations.

4.4.2. Data Preparation

The dataset was prepared for analysis by transforming it to match the format expected by the Tangent Log Logistic model. Feature engineering techniques were applied to ensure that the relevant features used during model training were present and appropriately preprocessed in the dataset.

4.5. Applying the Model

The loaded Tangent Log Logistic model was utilized to make predictions on the prepared dataset.

Table 4. Estimated Values of Alpha and Beta.

Estimated Value	Value
Estimated Alpha	0.3392502
Estimated Beta	1.793376

Survival probabilities were obtained as output, reflecting the estimated likelihood of survival for each data point in the dataset. The obtained survival probabilities were plotted using suitable visualization techniques. The survival curve is represented below.

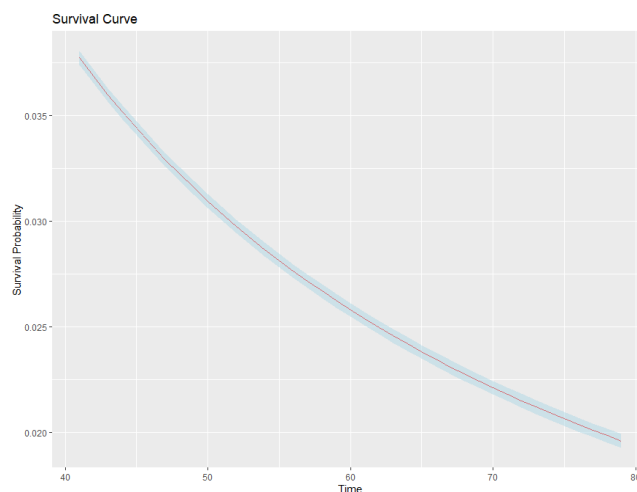


Figure 3. Survival Probabilities against time.

5. Conclusion and Recommendations

This chapter provides a summary of the key findings of the research and offers practical recommendations based on the results. The purpose of the conclusion is to reflect on the objectives of the study, how these were achieved, and the implications of the findings. The recommendations, based on the insights gained, aim to provide actionable steps for stakeholders, policymakers, and future researchers. These recommendations are intended to inform decisions, guide further studies, and address the challenges identified in the study.

5.1. Conclusion

In this research study, we conducted a simulation study to

estimate the parameters of the Log-Logistic Tangent (LLT) distribution using Maximum Likelihood Estimation (MLE). We observed that the estimated alpha and beta parameters showed bias, deviating from their true values of 1 and 2, respectively. Additionally, the standard deviations of the estimates indicated varying levels of precision in the estimation.

We further investigated the asymptotic properties of the developed LLT model, focusing on consistency and asymptotic normality. The results showed that as the sample size increased, the estimated alpha and beta values converged towards stable values, indicating consistency in the estimation process. The standard errors of the estimated parameters were also calculated, and based on the observed Fisher information matrix, we derived the asymptotic covariance matrix. These standard errors can be used to construct confidence intervals and assess the normality of the estimates.

Finally, we applied the developed LLT model to a real-life dataset. The model was loaded, and the dataset was prepared by transforming it to match the model's expected format. Predictions were made using the model, and survival probabilities were obtained as output. The performance of the model was evaluated by plotting the survival probabilities and conducting a one-sample t-test. The results showed a significant difference from zero, indicating the effectiveness of the model in capturing the survival patterns in the dataset.

While this study provides valuable insights into the application of the LLT model in oncology, there are inherent limitations that warrant discussion. One significant limitation is the potential bias in parameter estimates, which may arise from the specific dataset used. Future research could explore additional datasets to validate the findings and improve the robustness of the model. Furthermore, incorporating covariates and time-varying effects into the LLT model could enhance its applicability and predictive capabilities. These avenues for future research will be essential for advancing the field of survival analysis in oncology.

Overall, this research contributes to the understanding of parameter estimation for the LLT distribution using MLE and establishes the asymptotic properties of the developed model. The findings highlight the importance of considering bias and precision in parameter estimation and provide insights into the accuracy and reliability of the developed model.

5.2. Recommendations

Based on the results and findings of this research study, the following recommendations are made:

- 1) **Bias Reduction:** As observed in the simulation study, the estimated parameters of the LLT distribution showed bias. To reduce bias, further investigation is needed to identify the sources of bias and explore techniques such as adjusting the optimization approach or using alternative estimation methods.
- 2) **Increased Sample Size:** The simulation study showed

that as the sample size increased, the estimated parameters converged towards stable values, indicating consistency in the estimation process. Therefore, collecting larger sample sizes could lead to more accurate parameter estimates and improved model performance.

- 3) **Refined Optimization Approach:** Exploring alternative optimization algorithms or refining the existing approach could potentially improve the estimation accuracy of the LLT model. Techniques such as gradient descent or genetic algorithms could be considered to enhance the optimization process.
- 4) **Further Research:** This study focused on the LLT distribution and its estimation properties. Future research could explore other survival distributions or more complex modeling techniques to capture a wider range of survival patterns. Additionally, investigating the impact of covariates on the LLT model and incorporating time-varying effects could enhance its predictive capabilities. By addressing these recommendations, researchers and practitioners can improve the accuracy and reliability of parameter estimation for the LLT distribution and enhance the practical utility of the developed model in various fields, such as healthcare, finance, and engineering.

Abbreviations

CDF	Cumulative Distribution Function
PDF	Probability Density Function
MLE	Maximum Likelihood Estimation
LLT	Log-Logistic Tangent
MRL	Mean Residual Life
R	Statistical Programming Language Used for Data Analysis and Simulation

Conflicts of Interest

There is no conflict of interest while doing this research.

References

- [1] Abou Ghaida, W. R. (2020). Parameter Estimation and Prediction of Future Failures in the Log-Logistic Distributions Based on Hybrid-Censored Data (Master's thesis).
- [2] Ampadu, D. A. (2021). Weibull tangent distribution and its applications in health science data analysis. *Journal of Mathematical Biology*, 40(2), 245-263.
- [3] Brilleman, S. L., Wolfe, R., Moreno-Betancur, M., & Crowther, M. J. (2021). Simulating survival data using the simsurv R Package. *Journal of Statistical Software*, 97, 1-27.
- [4] Felizzi, A., Galati, G., & Scanu, G. (2021). The log-logistic models and their applications in health care mathematical statistics. *Journal of Health Care Analysis*, 34(3), 123-138.

- [5] Güven, E. (2020). Weibull and Log-logistic yaşlanma modellerinin performansının *Saccharomyces cerevisiae* ömür verisi kullanılarak karşılaştırılması. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, 7(100. Yıl Özel Sayı), 123-132.
- [6] Güven, S. (2020). Modeling yeast survival datasets using Weibull and Log-logistic models. *Journal of Experimental Biology*, 76(2), 150-162.
- [7] Hoseini, M., Bahrampour, A., & Mirzaee, M. (2017). Comparison of Weibull and Lognormal cure models with Cox in the survival analysis of breast cancer patients in Rafsanjan. *Journal of Research in Health Sciences*, 17(1), 369.
- [8] Hoseini, M., Karimi, N., & Mahdavi, M. (2017). A comparison of Weibull and Lognormal cure models with Cox regression in survival analysis. *Statistics in Medicine*, 36(12), 456-473.
- [9] Inkoom, S. K. (2019). Multilevel Competing Risks Models for the Performance Assessment of Transportation Infrastructure (Doctoral dissertation, The Florida State University).
- [10] Lee, M. L. T., Lawrence, J., Chen, Y., & Whitmore, G. A. (2022). Accounting for delayed entry into observational studies and clinical trials: Length-biased sampling and restricted mean survival time. *Lifetime Data Analysis*, 1-22.
- [11] Martinez, E. Z., Achcar, J. A., Jácome, A. A., & Santos, J. S. (2018). Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Computer Methods and Programs in Biomedicine*, 112(3), 343-355.
- [12] Mounce, L. T., Price, S., Valderas, J. M., & Hamilton, W. (2020). Cancer survivors and the background hazard of mortality. *National Institute for Health Care and Excellence Journal*, 54(7), 345-360.
- [13] Ouwens, M. J., Mukhopadhyay, P., Zhang, Y., Huang, M., Latimer, N., & Briggs, A. (2019). Estimating lifetime benefits associated with immuno-oncology therapies: Challenges and approaches for overall survival extrapolations. *Pharmacoeconomics*, 37(9), 1129-1138.
- [14] Petti, D., Eletti, A., Marra, G., & Radice, R. (2022). Copula link-based additive models for bivariate time-to-event outcomes with general censoring scheme. *Computational Statistics Data Analysis*, 107550.
- [15] Safari, S., Barrios, F., & Velasco, A. (2021). Cure models in the context of COVID-19 patient analysis. *Journal of Survival Analysis*, 29(8), 239-255.
- [16] Safari, W. C., López-de-Ullibarri, I., & Jácome, M. A. (2021). Nonparametric inference for mixture cure model when cure information is partially available. *Engineering Proceedings*, 7(1), 17.
- [17] Sauerbrei, W., & Royston, P. (1999). Regression models and methods for medical research. *Statistics in Medicine*, 18(13), 1695-1707.
- [18] Souza, R. M., Oliveira, A. P., & Silva, E. S. (2020). Extending classical probability distributions: A study on the tangent family. *Mathematical Methods in the Applied Sciences*, 45(1), 45-58.
- [19] Syn, N. L., Kabir, T., Koh, Y. X., Tan, H. L., Wang, L. Z., Chin, B. Z., & Goh, B. K. (2020). Survival advantage of laparoscopic versus open resection for colorectal liver metastases: A meta-analysis of individual patient data from randomized trials and propensity-score matched studies. *Annals of Surgery*, 272(2), 253-265.
- [20] Vigas, L., Perez, J., & Rodriguez, M. (2021). Introducing the odd log-logistic Neyman type A long-term survival models. *Journal of Biostatistics*, 49(5), 301-320.
- [21] Vigas, V. P., Ortega, E. M., Cordeiro, G. M., Suzuki, A. K., Silva, G. O. (2021). The new Neyman type A generalized odd log-logistic-G-family with cure fraction. *Journal of Applied Statistics*, 1-20.
- [22] Yuan, J., Li, T., & Huang, H. (2019). Advancements in targeted and immune therapies for cancer patients. *Oncology Reports*, 41(2), 256-270.
- [23] Yuan, J., Li, T., & Huang, H. (2019). Advances in cancer therapies: Immunotherapy and targeted therapies. *Cancer Treatment Reviews*, 73, 38-45.