Journal of Probability and Statistical Science
20(1), 166-183, August 2022

New defective models based on the Topp-Leone generated family of distributions

Weizhong Tian [a]
Shenzhen Technology University
Yaoting Yang [b]
Xi'an University of Technology
Tingting Liu [b]
Xi'an University of Technology
Arjun K. Gupta [c]
Bowling Green State University

ABSTRACT
This paper proposes two new defective models based on the Topp-Leone (TL) generated family of distributions. Unlike most of the cure rate models, the advantage of the defective model is that the cure rate can be modeled without adding any additional parameters to the model. The maximum likelihood function (MLE) is discussed for these models, and the asymptotic property of MLE is verified by simulation. The applications of this method are illustrated by three real data sets.

Keywords: Topp-Leone generated family, Defective distribution, Cure rate, Gompertz distribution, Inverse Gaussian distribution.

1 Introduction

In today's society, with the development of society and the progress of science and technology, researchers pay more attention to the analysis survival data. Models for survival data typically assume that all individuals in the study population are susceptible to the interested event with long sufficient follow-up. The most popular model that takes these conditions into account and has been studied by several authors in recent years is the standard mixture model, which was introduced by Boag [2] and Berkson and Gage [1]. The population survival function for the standard mixture model is $S(t) = p_1 + (1 - p_1)S_0(t)$, where $p_1 \in (0,1)$ is referred to as the cured fraction, and $S_0(t)$ is a proper survival function for uncured patients. The most common choices for $S_0(t)$ have been log-logical, lognormal, Gompertz, exponential, and Weibull distributions. See Matller and Zhou [16] for detail.

□Received November 2021, revised December 2021, in final form January 2022.
□ Weizhong Tian is affiliated with *College of Big Data and Internet, Shenzhen Technology University, Shenzhen, Guangdong 518118, China.* Yaoting Yang, Tingting Liu, with are affiliated with *School of Science, Xi'an University of Technology, Xi'an, Shaanxi 710054, China.* Arjun K. Gupta is affilated with *Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, USA*
Correspondingauthor. tianweizhong@sztu.edu.cn

An alternative method for solving such problems is the defective distribution,which has been introduced by Balka et al. [3, 4]. The advantages of defective distribution are allowing for cure fraction without any additional parameters in the model and not needing to make assumptions about the existence of cure rates. In the literature, the most common defective distributions are the Gompertz and the inverse Gaussian distributions, considered by [15, 6, 24, 3, 4]. A modified version of the Gompertz distribution was considered by Cantor and Shuster [5] in a pediatric cancer data set, and Gieser et al. [9] extended it to include covariate effects. Rocha et al. [18] discussed two new defective distributions based on the Marshall-Olkin (MO) family, which showed that if the baseline distribution is defective, then the new distribution under the baseline family would be defective. Additionally, according to the similar idea, the defective distribution with the baseline distribution Kumaraswamy family, was studied by Rocha et al. [20]. Rocha et al. [19] used an extended weibull distribution family combined with the MO family to generate a new family of defective distribution, which including ten new defective distributions.

The aim of this paper is to propose two new defunctive distributions based on the Toppe Leone (TL) generated family of distribution, which was discussed by Rezaei et al. [21]. This paper is organized as follows. In section 2, the TL generated family of distribution and two defective forms of the Gompertz and inverse Gaussian distributions are introduced. In section 3, we study the two new defective distributions based on the TL generated family of distributions. In section 4, through the simulations, the performance of Maximum Likelihood estimation are investigated. In section 6, we present three applications in real data sets to illustrate the proposed approach. The conclusion is given in section 7.

2 Methodology

2.1 The Topp-Leone generated family of distributions

Topp and Leone [23] introduced the TL distribution of a random variable X with a cumulative distribution function (CDF),

$$F(x;\alpha) = x^{\alpha}(2-x)^{\alpha}, (0 < x < 1, \alpha > 0).$$

Many authors have studied the mathematical characteristics of the TL distribution. For example, Nadarajah and Kotz [17] studied moments; Ghitany et al. [10] discussed reliability measures and stochastic orderings, and Zhou et al. [26] studied the distributions of sums, products and ratios. The behavior of the kurtosis (Kotz and Seier [12]), record values (Zghoul [25]) and moments of order statistics (Genc [7]) was also studied. Besides this, Genc [8] provided the stress strength model, and Sindhu et al. [22] discussed the Bayesian estimation under trimmed samples.

For making TL distribution more useful to model lifetime data sets, Rezaei et al. [21] introduced the TL generated family, such that, the random variable X belongs the TL generated family, if its CDF has the form as

$$F(x;\alpha) = G(x)^{\alpha}(2-G(x))^{\alpha}, (\alpha > 0) \quad (2.1)$$

where $G(x)$ is a baseline CDF. Therefore, the probability density function (pdf) of X is

$$f(x;\alpha) = \alpha g(x) G(x)^{\alpha-1}[1-G(x)][2-G(x)]^{\alpha-1}. \quad (2.2)$$

2.2 The defective Gompertz and inverse Gaussian distributions

The Gompertz distribution is used to model survival data in a variety of knowledge domains, especially where there is approximate exponential risk. Its pdf, survival function and hazard function are, respectively,

$$f(t;a,b) = be^{at}e^{-\frac{b}{a}(e^{at}-1)}, \tag{2.3}$$

$$s(t;a,b) = e^{-\frac{b}{a}(e^{at}-1)}, \tag{2.4}$$

and

$$h(t;a,b) = be^{at},$$

for $t > 0$, $a > 0$ is the shape parameter and $b > 0$ is the location parameter.

The defective Gompertz distribution allows the negative values of the parameter a, and the fraction of cure, or the survival function limit, is

$$p_0 = \lim_{t\to\infty} s(t;\alpha,b) = \lim_{t\to\infty} e^{-\frac{b}{a}(e^{at}-1)} = e^{\frac{b}{a}} \in (0,1).$$

The inverse Gaussian is a skewed, two-parameter continuous distribution whose density is similar to the Gamma distribution with greater skewness and a sharper peak. Its pdf, survival function and hazard function are, respectively,

$$f(t;a,b) = \frac{1}{\sqrt{2b\pi t^3}} e^{-\frac{1}{2bt}(1-at)^2}, \tag{2.5}$$

$$s(t;a,b) = 1 - \left(\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + e^{\frac{2a}{b}}\Phi\left(\frac{-1+at}{\sqrt{bt}}\right)\right), \tag{2.6}$$

$$h(t;a,b) = \frac{\frac{1}{\sqrt{2b\pi t^3}} e^{-\frac{1}{2bt}(1-at)^2}}{1 - \left(\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + e^{\frac{2a}{b}}\Phi\left(\frac{-1+at}{\sqrt{bt}}\right)\right)},$$

where $t \in \Re$, $a > 0, b > 0$, and $\Phi(\cdot)$ is the cdf of a standard normal distribution.

The inverse Gaussian distribution can be defective when $a < 0$. The fraction of cure, or the survival function limit, is

$$p_0 = \lim_{t\to\infty} s(t;a,b) = \lim_{t\to\infty} \left[1 - \left(\Phi\left(\frac{-1+at}{\sqrt{bt}}\right) + e^{\frac{2a}{b}}\Phi\left(\frac{-1+at}{\sqrt{bt}}\right)\right)\right] = 1 - e^{\frac{2b}{a}} \in (0,1).$$

3 New defective distributions based on Topp-Leone generated family of distributions

3.1 The Topp-Leone Gompertz distribution

Using (2.2), (2.3) and (2.4), we obtain the pdf and survival function of TL Gompertz distribution.

$$f^*(t;\alpha,a,b) = 2\alpha be^{at}\ ^{\frac{2b}{a}(e^{at}-1)}\left[1 - e^{-\frac{2b}{a}(e^{at}-1)}\right]^{\alpha-1}, \tag{3.1}$$

$$s^*(t;\alpha,a,b) = 1 - \left(1 - e^{-\frac{2b}{a}(e^{at}-1)}\right)^{\alpha}, \tag{3.2}$$

where $t > 0$, $\alpha > 0$, $a > 0$ and $b > 0$.

If $a < 0$, the TL Gompertz distribution is defective and its cure fraction is

$$p = \lim_{t\to\infty} s^*(t;\alpha,a,b) = \lim_{t\to\infty} 1 - \left[1 - e^{-\frac{2b}{a}(e^{at}-1)}\right]^{\alpha} = 1 - \left(1 - e^{\frac{2b}{a}}\right)^{\alpha} \in (0,1).$$

The pdf, survival function and hazard function of the defective TL Gompertz distributions and defective Gompertz with different values of parameters are drawn in Figure 1, Figure 2 ,

Figure 3, respectively.

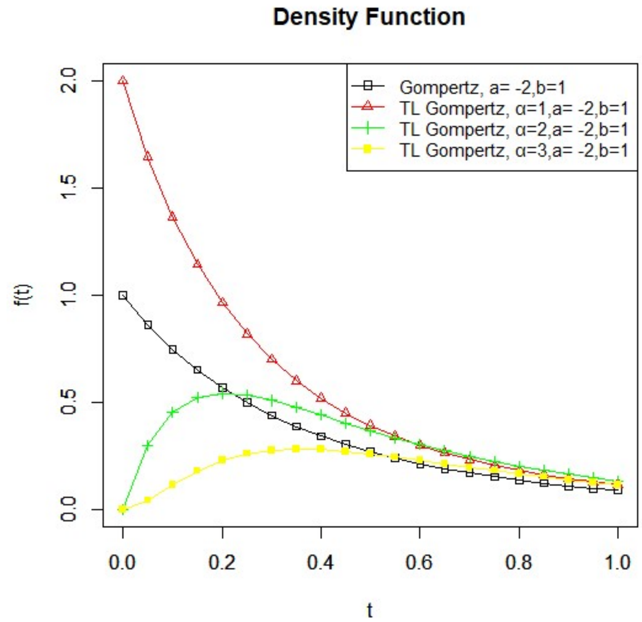

Figure 1: The pdfs of the defective Gompertz and defective TL Gompertz distributions.

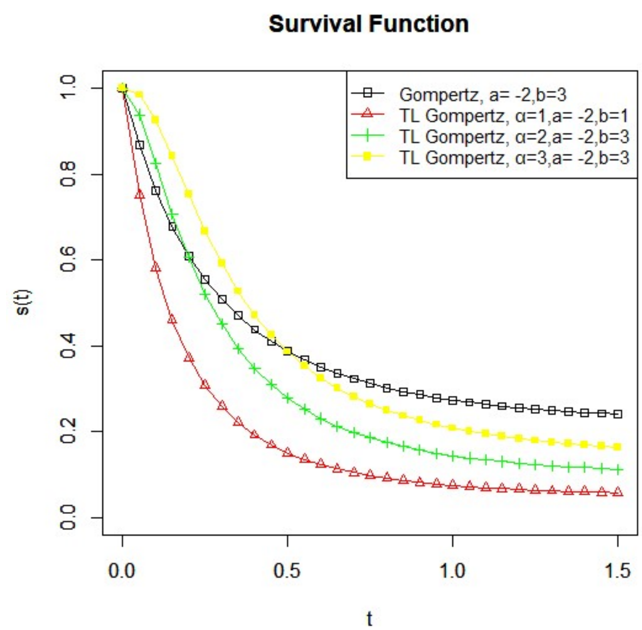

Figure 2: The survival functions of the defective Gompertz and defective TL Gompertz distributions.

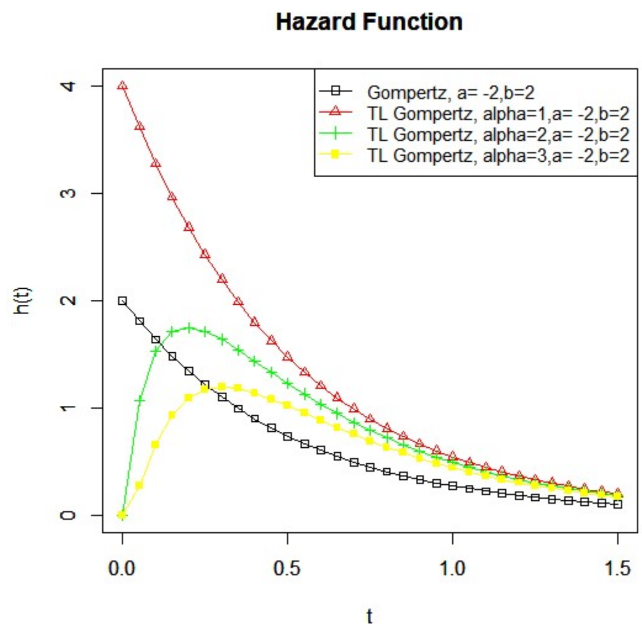

Figure 3: The hazard functions of the defective Gompertz and defective TL Gompertz distributions.

3.2 The Topp-Leone inverse Gaussian distribution

Using (2.2), (2.5) and (2.6), we obtain the pdf and survival function of the TL inverse Gaussian distribution as follows,

$$f^*(t;\alpha,a,b) = \frac{2\alpha}{\sqrt{2b\pi t^3}} e^{-\frac{1}{2bt}(1-at)^2}\left(1-\left(\Phi(\frac{-1+at}{\sqrt{bt}})+e^{\frac{2a}{b}}\Phi(\frac{-1-at}{\sqrt{bt}})\right)\right) \times\left[1-\left(1-\left(\Phi(\frac{-1+at}{\sqrt{bt}})+e^{\frac{2a}{b}}\Phi(\frac{-1-at}{\sqrt{bt}})\right)\right)^2\right]^{\alpha-1}, \quad (3.3)$$

$$s^*(t;\alpha,a,b) = 1-\left[1-\left(1-\left(\Phi(\frac{-1+at}{\sqrt{bt}})+e^{\frac{2a}{b}}\Phi(\frac{-1-at}{\sqrt{bt}})\right)\right)^2\right]^{\alpha}, \quad (3.4)$$

where $t>0$, $\alpha>0$, $a>0$ and $b>0$.

If $a<0$, the TL inverse Gaussian distribution is defective, and its cure fraction is

$$p = \lim_{t\to\infty} s^*(t;\alpha,a,b) = \lim_{t\to\infty} 1-\left[1-\left(1-\left(\Phi(\frac{-1+at}{\sqrt{bt}})+e^{\frac{2a}{b}}\Phi(\frac{-1-at}{\sqrt{bt}})\right)\right)^2\right]^{\alpha}$$

$$= 1-\left(2e^{\frac{2a}{b}}-e^{\frac{4a}{b}}\right)^{\alpha} \in (0,1).$$

The pdf, survival function and hazard function of the defective inverse Gaussian and TL inverse Gaussian distributions with different values of parameters are drawn in Figure 4, Figure 5, Figure 6, respectively.

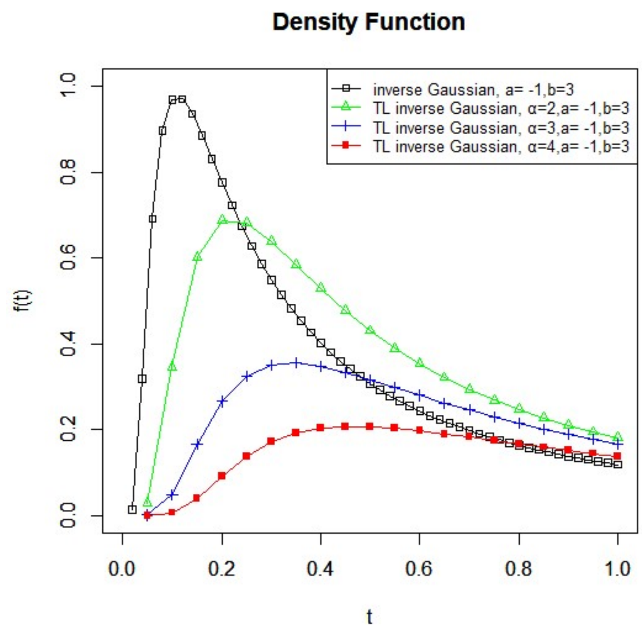

Figure 4: The pdfs of the defective inverse Gaussian and defective TL inverse Gaussian distributions.

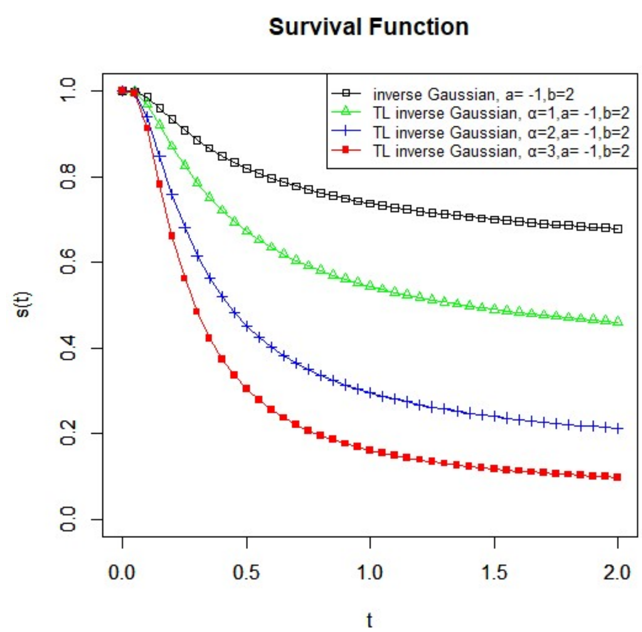

Figure 5: The survival functions of the defective inverse Gaussian and defective TL inverse Gaussian distributions.

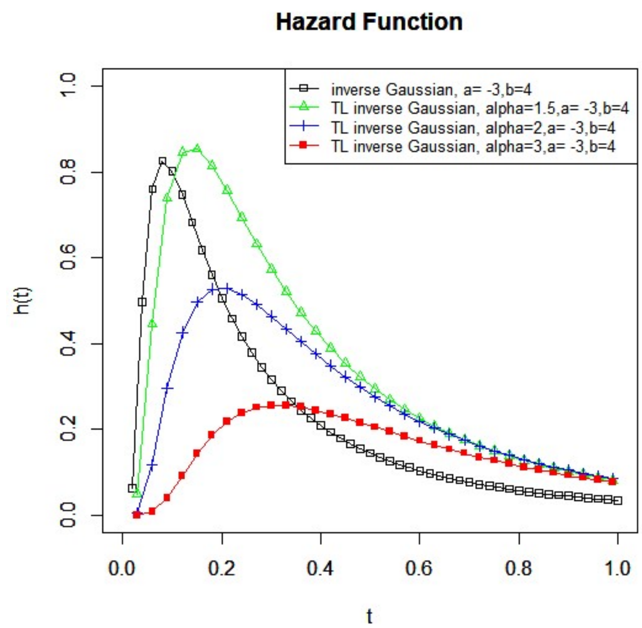

Figure 6: The hazard functions of the defective inverse Gaussian and defective TL inverse Gaussian distributions.

4 Inference analysis

We consider a data set $D = (t,\delta)$, where $t = (t_1, t_2, \cdots, t_n)'$ are the observed failure times and $\delta = (\delta_1, \delta_2, \cdots, \delta_n)'$ are the censored failure times. For $i = 1, \cdots, n$, δ_i is equal to 1 if a failure is observed and 0 otherwise.

Suppose that the data are independently and identically distributed from a distribution with pdf and survival functions by $f^*(\cdot,\theta)$ and $s^*(\cdot,\theta)$, respectively, where $\theta = (\theta_1, \cdots, \theta_q)'$ denotes a vector of parameters, and q denotes the number of parameters. The likelihood function of θ can be written as

$$L(\theta, D) = \prod_{i=1}^{n} \left[f^*(t_i, \theta)^{\delta_i} s^*(t_i, \theta)^{1-\delta_i} \right],$$

and the corresponding log-likelihood function is

$$logL(\theta, D) = const + \sum_{i=1}^{n} \delta_i log(f^*(t_i,\theta)) + \sum_{i=1}^{n} (1-\delta_i) log(s^*(t_i,\theta)).$$

For the TL Gompertz distribution given by (3.1) and (3.2), we have

$$\begin{aligned} logL(\theta, D) &= const + \sum_{i=1}^{n} \delta_i log\left[2\alpha b e^{at_i - \frac{2b}{a}(e^{at_i}-1)} \left(1 - e^{-\frac{2b}{a}(e^{at_i}-1)}\right)^{\alpha-1} \right] \\ &+ \sum_{i=1}^{n} (1-\delta_i) log\left[1 - \left(1 - e^{-\frac{2b}{a}(e^{at_i}-1)}\right)^{\alpha} \right], \end{aligned}$$

and

$$\frac{\partial logL(\theta,D)}{\partial\alpha}=\sum_{i=1}^{n}\left[\frac{1}{2a}+log\left(e^{-\frac{2b}{a}(e^{at_i}-1)}\right)\right]-\sum_{i=1}^{n}(1-\delta_i)\frac{\alpha\left(1-e^{-\frac{2b}{a}(e^{at_i}-1)}\right)^{\alpha-1}}{1-\left(1-e^{-\frac{2b}{a}(e^{at_i}-1)}\right)^{\alpha}},$$

$$\frac{\partial logL(\theta,D)}{\partial a}=\sum_{i=1}^{n}\delta_i\left[t_i+\frac{2b}{a^2}(e^{at_i}-1)-\frac{2bt_i}{a}e^{at_i}\right.$$
$$\left.+(\alpha-1)\frac{\left(\frac{2bt_i}{a}e^{at_i}-\frac{2b}{a^2}(e^{at_i}-1)\right)e^{-\frac{2b}{a}(e^{at_i}-1)}}{1-e^{-\frac{2b}{a}(e^{at_i}-1)}}\right]$$
$$+\sum_{i=1}^{n}(1-\delta_i)\frac{\alpha\left(1-e^{-\frac{2b}{a}(e^{at_i}-1)}\right)^{\alpha-1}e^{-\frac{2b}{a}(e^{at_i}-1)}\left(\frac{2b}{a^2}(e^{at_i}-1)-\frac{2bt_i}{a}e^{at_i}\right)}{1-\left(1-e^{-\frac{2b}{a}(e^{at_i}-1)}\right)^{\alpha}},$$

$$\frac{\partial logL(\theta,D)}{\partial b}=\sum_{i=1}^{n}\delta_i\left[\frac{1}{b}-\frac{2}{a}(e^{at_i}-1)+(\alpha-1)\frac{\frac{2}{a}(e^{at_i}-1)e^{-\frac{2b}{a}(e^{at_i}-1)}}{1-e^{-\frac{2b}{a}(e^{at_i}-1)}}\right]$$
$$+\sum_{i=1}^{n}(1-\delta_i)\frac{\frac{2\alpha}{a}(e^{at_i}-1)e^{-\frac{2b}{a}(e^{at_i}-1)}\left(1-e^{-\frac{2b}{a}(e^{at_i}-1)}\right)^{\alpha-1}}{1-\left(1-e^{-\frac{2b}{a}(e^{at_i}-1)}\right)^{\alpha}}.$$

For the TL inverse Gaussian distribution given by (3.3) and (3.4), we have

$$logL(\theta,D)\quad=const+\sum_{i=1}^{n}\delta_i log\left[\frac{2\alpha}{\sqrt{2b\pi t_i^3}}e^{-\frac{1}{2bt_i}(1-at_i)^2}W(1-W^2)^{\alpha-1}\right]$$
$$+\sum_{i=1}^{n}(1-\delta_i)log(1-(1-W^2)^{\alpha}),$$

and

$$\frac{\partial logL(\theta,D)}{\partial\alpha}=\sum_{i=1}^{n}\delta_i\left[\frac{1}{2\alpha}+log(1-W^2)\right]-\sum_{i=1}^{n}\alpha(1-\delta_i)\frac{(1-W^2)^{\alpha-1}}{1-(1-W^2)^{\alpha}},$$

$$\frac{\partial logL(\theta,D)}{\partial a}=\sum_{i=1}^{n}\delta_i\left[\frac{1-at_i}{b}-\frac{B}{W}+2(\alpha-1)\frac{WB}{1-W^2}\right]-\sum_{i=1}^{n}(1-\delta_i)\frac{2\alpha(1-W^2)^{\alpha-1}WB}{1-(1-W^2)^{\alpha}},$$

$$\frac{\partial logL(\theta,D)}{\partial b}=\sum_{i=1}^{n}\delta_i\left[\frac{(1-at_i)^2}{2b^2t_i}-\frac{1}{4b\pi}+\frac{A}{W}+\frac{2(\alpha-1)WA}{1-W^2}\right]$$
$$-\sum_{i=1}^{n}(1-\delta_i)\frac{2\alpha(1-W^2)^{\alpha-1}WA}{1-(1-W^2)^{\alpha}},$$

where $u_1 = \Phi(\frac{-1+at_i}{\sqrt{bt_i}})$, $u_2 = \Phi(\frac{-1-at_i}{\sqrt{bt_i}})$, $v_1 = \phi(\frac{-1+at_i}{\sqrt{bt_i}})$, $v_1 = \phi(\frac{-1-at_i}{\sqrt{bt_i}})$, $W = 1 - u_1 - e^{\frac{2a}{b}}u_2$, $A = \frac{1-at_i}{2}\sqrt{\frac{1}{b^3t_i}}v_1 - \frac{2a}{b^2}e^{\frac{2a}{b}}u_2 + \frac{1+at_i}{2}\sqrt{\frac{1}{b^3t_i}}v_2$, and $B = \sqrt{\frac{t_i}{b}}v_1 + \frac{2}{b}e^{\frac{2a}{b}}u_2 - \sqrt{\frac{t_i}{b}}e^{\frac{2a}{b}}v_2$.

The estimated value of the parameters, $\widehat{\boldsymbol{\theta}}$, can be obtained by solving the partial derivative of the likelihood function simultaneously. It is well known that the distribution of $\hat{\theta} - \theta$ can be approximated by a q-variate normal distribution with zero means and covariance matrix $I(\hat{\theta})^{-1}$, and $I(\theta)$ denotes the observed information matrix defined by

$$I(\theta) = \begin{pmatrix} \frac{\partial^2 logL}{\partial\theta_1^2} & \frac{\partial^2 logL}{\partial\theta_1\theta_2} & \cdots & \frac{\partial^2 logL}{\partial\theta_1\theta_q} \\ \frac{\partial^2 logL}{\partial\theta_2\theta_1} & \frac{\partial^2 logL}{\partial\theta_2^2} & \cdots & \frac{\partial^2 logL}{\partial\theta_2\theta_q} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 logL}{\partial\theta_q\theta_1} & \frac{\partial^2 logL}{\partial\theta_q\theta_2} & \cdots & \frac{\partial^2 logL}{\partial\theta_q^2} \end{pmatrix}.$$

5 Simulation

In this section, we discuss the performance of the MLE with respect to sample size to show, among other things, that the usual asymptotes of maximum likelihood estimators still hold for defective distributions. The description of data generation are described below. All computations were performed in R.

Suppose the time of occurrence of an event of interest has a cumulative distribution function $F(t)$. We want to simulate a random sample of size n containing real times, censored times and a cure fraction of p. An algorithm for this purpose is

1. Determine the desired parameter values, as well as the value of the cure fraction p;
2. For each $i = 1,2,\cdots,n$, generate a random variable $M_i \sim Bernoulli(1-p)$;
3. If $M_i = 0$ set $t'_i = \infty$. If $M_i = 1$ take t'_i as the root of $F(t) = u$, where $u \sim uniform(0, 1-p)$, then take $p' = max(t'_i)$, considering only the finite t'_i;
4. Generate $u'_i \sim uniform(0,\ p')$, for $i = 1,2,\cdots,n$;
5. Calculate $t_i = min(t'_i, u'_i))$. If $t_i < u'_i$ set $\delta_i = 1$, otherwise set $\delta_i = 0$.

In this first experiment, we take the sample size $n = 100, 200, 500, 1000$. Each sample was replicated 1000 times. The random samples came from the defective TL Gompertz distribution with parameters $(\alpha, a, b, p) = (3, -2, 4, 0.0540)$, $(4, -2, 4, 0.0713)$, $(3, -3, 4, 0.1943)$, $(4, -3, 4, 0.2503)$, $(3, -2, 2, 0.3535)$, $(4, -2, 2, 0.4410)$, $(3, -3, 2, 0.6007)$, and $(4, -2, 2, 0.7059)$, respectively. We use optim functions to estimate the numerical solution of each parameter, and get the estimator and standard deviations, which are shown in Table 1.

In this second experiment, we take the sample size $n = 100, 200, 500, 1000$. Each sample was replicated 1000 times. The random samples came from the defective TL inverse Gaussian distribution with parameters $(\alpha, a, b, p) = (2, -2, 5, 0.5145)$, $(4, -2, 5, 0.7643)$, $(2, -2, 7, 0.3430)$, $(4, -2, 7, 0.5684)$, $(2, -1.5, 5, 0.3657)$, $(4, -1.5, 5, 0.5977)$, $(2, -1.5, 7, 0.2282)$, and $(4, -1.5, 7, 0.4044)$, respectively. We use optim functions to estimate the numerical solution of each parameter, and get the estimator and standard deviations, which are shown in Table 2.

Table 1: Simulation of the MLE for parameter and standard deviation (SD) of the defective TL Gompertz distribution.

n	α	a	b	p	$\hat{\alpha}$(SD)	$\hat{a}$(SD)	$\hat{b}$(SD)	$\hat{p}$(SD)
100	3	-2	4	0.0540	3.2591(0.7782)	-2.0989(0.5949)	4.2621(0.8973)	0.0590(0.0334)
	4	-2	4	0.0713	4.2966(1.0288)	-2.0601(0.4959)	4.1749(0.7851)	0.0763(0.0386)
	3	-3	4	0.1943	3.2544(0.7914)	-3.1580(0.6739)	4.2855(0.9535)	0.1997(0.0571)
	4	-3	4	0.2503	4.4009(1.1187)	-3.1040(0.6478)	4.2725(0.9196)	0.2495(0.0675)
	3	-2	2	0.3535	3.3679(0.9516)	-2.1216(0.4431)	2.2177(0.5881)	0.3350(0.0677)
	4	-2	2	0.4410	4.4740(1.2817)	-2.0994(0.4623)	2.1843(0.6008)	0.4428(0.0752)
	3	-3	2	0.6007	3.4642(1.1448)	-3.1878(0.8006)	2.2986(0.8364)	0.5984(0.0781)
	4	-3	2	0.7059	4.5851(1.5401)	-3.1684(0.8418)	2.2661(0.8095)	0.6996(0.0818)
200	3	-2	4	0.0540	3.1437(0.7782)	-2.0989(0.5949)	4.2621(0.8973)	0.0590(0.0334)
	4	-2	4	0.0713	4.1785(0.7655)	-2.0280(0.3406)	4.0952(0.5829)	0.0731(0.0260)
	3	-3	4	0.1943	3.2002(0.5995)	-3.0878(0.4635)	4.1907(0.6856)	0.1966(0.0396)
	4	-3	4	0.2503	4.2232(0.8512)	-3.0459(0.4200)	4.1433(0.6962)	0.2492(0.0415)
	3	-2	2	0.3535	3.2225(0.6950)	-2.0758(0.3459)	2.1373(0.4588)	0.3539(0.0475)
	4	-2	2	0.4410	4.2788(0.9378)	-2.0677(0.3067)	2.1208(0.4397)	0.4426(0.0497)
	3	-3	2	0.6007	3.3427(0.8827)	-3.1380(0.5365)	2.2301(0.6175)	0.5989(0.0480)
	4	-3	2	0.7059	4.4268(1.1217)	-3.1086(0.5280)	2.1867(0.5737)	0.7060(0.0459)
500	3	-2	4	0.0540	3.0664(0.3449)	-2.0207(0.2432)	4.0635(0.3897)	0.0546(0.0146)
	4	-2	4	0.0713	4.0768(0.4653)	-2.0217(0.2150)	4.0474(0.3714)	0.0728(0.0165)
	3	-3	4	0.1943	3.0722(0.3713)	-3.0253(0.2937)	4.0686(0.4496)	0.1942(0.0243)
	4	-3	4	0.2503	4.1164(0.5388)	-3.0421(0.2680)	4.0883(0.4557)	0.2515(0.0251)
	3	-2	2	0.3535	3.1040(0.4193)	-2.0317(0.2006)	2.0629(0.2815)	0.3534(0.0290)
	4	-2	2	0.4410	4.1027(0.5520)	-2.0250(0.1739)	2.0431(0.2547)	0.4420(0.0277)
	3	-3	2	0.6007	3.1615(0.5217)	-3.0813(0.3473)	2.1189(0.4067)	0.6015(0.0282)
	4	-3	2	0.7059	4.2403(0.7790)	-3.0621(0.3303)	2.1172(0.4103)	0.7046(0.0268)
1000	3	-2	4	0.0540	3.0306(0.2321)	-2.0080(0.1564)	4.0214(0.2655)	0.0545(0.0097)
	4	-2	4	0.0713	4.0287(0.3294)	-2.0102(0.1378)	4.0206(0.2499)	0.0728(0.0109)
	3	-3	4	0.1943	3.0312(0.2471)	-3.0183(0.1984)	4.0429(0.3100)	0.1939(0.0166)
	4	-3	4	0.2503	4.0526(0.3773)	-3.0113(0.1835)	4.0347(0.3206)	0.2501(0.0173)
	3	-2	2	0.3535	3.0307(0.2784)	-2.0095(0.1305)	2.0219(0.1886)	0.3526(0.0188)
	4	-2	2	0.4410	4.0472(0.4130)	-2.0069(0.1224)	2.0161(0.1884)	0.4412(0.0207)
	3	-3	2	0.6007	3.0730(0.3608)	-3.0362(0.2365)	2.0592(0.2812)	0.5995(0.0189)
	4	-3	2	0.7059	4.1542(0.5777)	-3.0415(0.2490)	2.0759(0.3094)	0.7055(0.0179)

Table 2: Simulation of the MLE for mean and standard deviation of the defective Topp-Leone inverse Gaussian distribution.

n	α	a	b	p	$\hat{\alpha}$(SD)	$\hat{a}$(SD)	$\hat{b}$(SD)	$\hat{p}$(SD)
100	2	-2	5	0.5145	1.9634(1.4274)	-2.0007(1.4319)	4.9664(1.4117)	0.4850(0.3265)
	4	-2	5	0.7643	3.8456(1.3395)	-2.0079(1.3561)	5.0027(1.3464)	0.6491(0.3120)
	2	-2	7	0.3430	1.9589(0.9739)	-2.0390(0.9784)	7.0021(0.9697)	0.3422(0.2260)
	4	-2	7	0.5684	3.9076(0.8989)	-2.0179(0.9407)	6.9934(0.9031)	0.5297(0.2496)
	2	-1.5	5	0.3657	1.9828(1.0028)	-1.5481(0.9688)	4.9949(0.9812)	0.3730(0.2695)
	4	-1.5	5	0.5977	3.9582(0.9194)	-1.5213(0.9755)	5.1160(0.9365)	0.5478(0.2899)
	2	-1.5	7	0.2282	2.0018(0.9609)	-1.4802(0.9679)	6.9942(1.0055)	0.2469(0.2060)
	4	-1.5	7	0.4044	3.9100(0.8925)	-1.5343(0.9043)	6.9769(0.8918)	0.4047(0.2434)
200	2	-2	5	0.5145	2.0207(1.4051)	-1.9479(1.4128)	5.0257(1.4321)	0.4782(0.3295)
	4	-2	5	0.7643	3.9728(1.4282)	-2.0206(1.3754)	4.9448(1.4046)	0.6563(0.3127)
	2	-2	7	0.3430	1.9639(1.0157)	-2.0007(1.0027)	6.9814(0.9997)	0.3365(0.2408)
	4	-2	7	0.5684	3.9588(0.9825)	-2.0326(0.9985)	7.0424(0.9895)	0.5315(0.2594)
	2	-1.5	5	0.3657	1.9900(0.9840)	-1.4556(1.0239)	4.9857(1.0043)	0.3606(0.2716)
	4	-1.5	5	0.5977	3.9513(1.0141)	-1.4859(1.0049)	5.0366(0.9999)	0.5348(0.3043)
	2	-1.5	7	0.2282	2.0018(1.0060)	-1.5284(0.9983)	6.9533(0.9952)	0.2613(0.2155)
	4	-1.5	7	0.4044	4.0195(0.9852)	-1.4677(0.9586)	6.9129(0.9855)	0.4003(0.2629)
500	2	-2	5	0.5145	2.0224(1.4167)	-1.9486(1.4153)	4.9419(1.4370)	0.4895(0.3308)
	4	-2	5	0.7643	3.9966(1.3472)	-1.9326(1.4285)	5.0010(1.4655)	0.6400(0.3270)
	2	-2	7	0.3430	1.9946(0.9618)	-2.0360(1.0169)	7.0113(1.0402)	0.3438(0.2313)
	4	-2	7	0.5684	4.0172(0.9460)	-1.9990(1.0286)	6.9555(1.0296)	0.5270(0.2719)
	2	-1.5	5	0.3657	1.9570(1.0142)	-1.5082(0.9975)	4.9915(1.0148)	0.3681(0.2808)
	4	-1.5	5	0.5977	4.0182(1.0283)	-1.5282(1.0285)	4.9681(1.0009)	0.5617(0.3031)
	2	-1.5	7	0.2282	2.0097(0.9903)	-1.5426(1.0068)	7.0346(1.0103)	0.2567(0.2173)
	4	-1.5	7	0.4044	4.0277(1.0320)	-1.5293(0.9649)	7.0478(0.9800)	0.4141(0.2656)
1000	2	-2	5	0.5145	1.9959(1.4123)	-2.0640(1.4047)	4.9724(1.3855)	0.4925(0.3312)
	4	-2	5	0.7643	4.0945(1.3663)	-1.9590(1.4237)	5.0724(1.3937)	0.6332(0.3314)
	2	-2	7	0.3430	2.0126(1.0187)	-1.9786(1.0048)	7.0185(1.0004)	0.3354(0.2279)
	4	-2	7	0.5684	4.0168(1.0067)	-2.0018(0.9927)	6.9511(0.9855)	0.5362(0.2682)
	2	-1.5	5	0.3657	1.9853(1.0191)	-1.5087(0.9803)	4.9838(0.9860)	0.3686(0.2729)
	4	-1.5	5	0.5977	43.9386(1.0012)	-1.4991(0.9828)	4.9854(1.0177)	0.5464(0.3001)
	2	-1.5	7	0.2282	1.9992(0.9629)	-1.4496(1.0034)	7.0300(1.0098)	0.2467(0.2111)
	4	-1.5	7	0.4044	4.0153(1.0153)	-1.4707(0.9978)	6.9994(1.0242)	0.3988(0.2716)

For the TL Gompertz distribution, the SD decreases as n increases; but, for the TL inverse Gaussian distribution, the SD is not fixed as n increases.

6 Application

In this section, we use three data sets to study the feasibility of the new defective distribution based on the TL generated family of distributions. The *optim* and *maxLik* functions in R are used to maximize the log-likelihood function, and the algorithm *SANN* was chosen for maximization. For comparing with other existing defective distributions, such as, Gompertz distribution, MO Gompertz distribution, inverse Gaussian distribution, and MO inverse Gaussian distributions, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are considered. For further comparison, we also obtain Kaplan-Meier curves for three data sets. The Kaplan-Meier curve reaches a stable level between 0 and 1 and is evidence that a cured individual exists. If the curve of the survival function fitted by the distribution is closer to the Kaplan-Meier curve, the model is better.

6.1 Colon cancer data

The first data set was the successful trial of adjuvant chemotherapy for colon cancer, which can be seen in the survival package of R. A detailed description of this data is given in [13]. The event of interest here is the recurrence or death of the patient under the recommended treatment. This data set contains 16 variables, having 1858 observations, of which 938 were censored, and the censored rate is 0.5048. The Kaplan-Meier curve of this data stabilizes at 0.465.

The results are summarized in Table 3. Based on the results obtained, we come to the conclusion that the estimated cure rates of Gompertz, TL Gompertz distributions were smaller than the end of the Kaplan-Meier curve, while the cure rates of Gompertz and TL inverse Gaussian distributions are closer to the end of kaplan-meier curve. The smallest values of AIC and BIC are the TL Gompertz distribution, the second smallest is the Gompertz distribution, the third smallest is the TL inverse Gaussian distribution and largest one is the MO Gompertz distribution.

According to the value of the parameters that are estimated, we can draw the corresponding survival curves. Figure 7 shows that the fitted curves of the survival function by the TL Gompertz and Gompertz distributions are closer to the Kaplan-Meier curves, but the curves fitted by Gompertz distributions are closer to the Kaplan-Meier curves within the range of (0.6, 1). The fitted curves of the survival function by the TL inverse Gaussian distribution is closer to the Kaplan-Meier curves in the Figure 8.

Table 3: MLEs for the fitted distributions for the colon data set.

Model	$\hat{\alpha}$	$\hat{r}$	$\hat{a}$	$\hat{b}$	$\hat{p}$	Max	AIC	BIC
Gompertz	-	-	-1.8191	1.7065	0.3444	-762.1549	1528.3097	1539.3642
TL Gompertz	2.3335	-	-5.2531	3.5790	0.1196	-742.4400	1490.8801	1499.9346
MO Gompertz	-	5.7666	-5.1659	8.2514	0.8689	-1133.8974	2273.7949	2282.8494
inverse Gaussian	-	-	-2.2849	4.4609	0.6410	-1042.4832	2088.9663	2100.0208
TL inverse Gaussian	0.9395	-	-4.2302	6.5587	0.5252	-797.3239	1600.6478	1609.7023
MO inverse Gaussian	-	1.5100	-3.1009	7.2738	0.6702	-978.1472	1962.2944	1971.3489

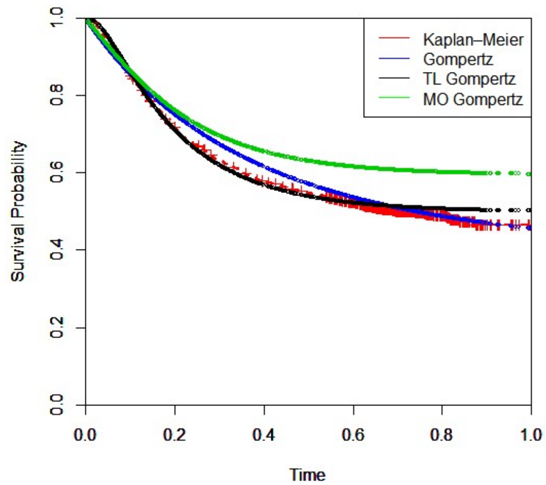

Figure 7: Survival curves for the fitted defective Gompertz, defective MO Gompertz and defective TL Gompertz distributions for the colon cancer data set.

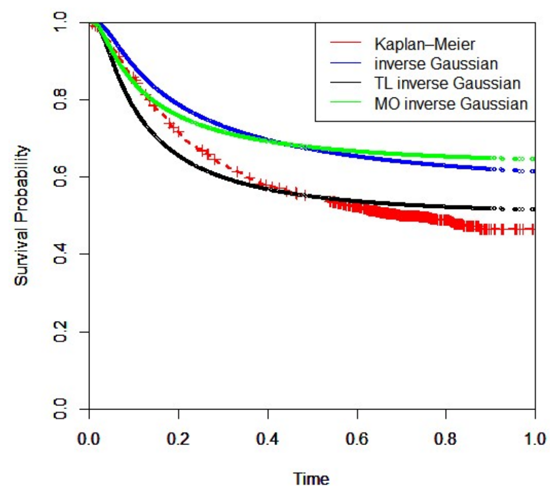

Figure 8: Survival curves for the fitted defective inverse Gaussian, defective MO inverse Gaussian and defective TL inverse Gaussian distributions for the colon cancer data set.

6.2 Melanoma data

The melanoma data come from the Eastern Cooperative Oncology Group (ECOG) phase III clinical trial e1684 which is used for modeling semicure PH mixture cure model (Kirkwood, et al., 1996) and is available in the smcure package in R, see [11]. Melanoma is a cancer that develops from melanocytes, which are the main cause of skin pigmentation. This is a potentially serious malignancy that can occur in the skin, mucous membranes, eyes, and central nervous system, with a high risk of metastasis and a high mortality rate. There are 285 observed values, of which 88 were censored, with a censored rate 0.3088. The event of concern here is the recurrence and cure of the patient after removal of the malignant melanoma. The Kaplan-Meier curve of this data is stabilizes at 0.281.

The proposed distribution is applied to this data set, and the final result is shown in Table 4. Based on the results obtained, we come to the conclusion that the smallest values of AIC and BIC are the TL Gompertz and TL inverse Gaussian distributions, the second smallest is the Gompertz distribution and the largest one is the MO Gompertz distribution.

According to the value of the parameters that are estimated, we can draw the corresponding survival curve. Figure 9 shows that the fitted curves of the survival function by the TL Gompertz and Gompertz distributions is closer to the Kaplan-Meier curves. Figure 10 shows that the fitted curves of the survival function by the TL inverse Gaussian distribution are closer to the Kaplan-Meier curves.

Table 4: MLEs for the fitted distributions for the melanoma data set.

Model	$\hat{\alpha}$	$\hat{r}$	$\hat{a}$	$\hat{b}$	$\hat{p}$	Max	AIC	BIC
Gompertz	-	-	-0.5881	0.7120	0.4378	-383.4901	770.9803	778.2852
TL Gompertz	1.2416	-	-0.7153	0.5075	0.0735	-382.3855	770.7711	776.0760
MO Gompertz	-	2.6697	-1.2071	3.6650	0.8725	-616.8638	1239.7277	1245.0327
inverse Gaussian	-	-	-1.5500	7.7741	0.3288	-434.9962	873.9923	881.2973
TL inverse Gaussian	3.5332	-	-1.384252	7.9424	0.0866	-376.9198	759.8397	765.1446
MO inverse Gaussian	-	3.0071	-0.8064	10.6182	0.3303	-399.4102	804.8204	810.1254

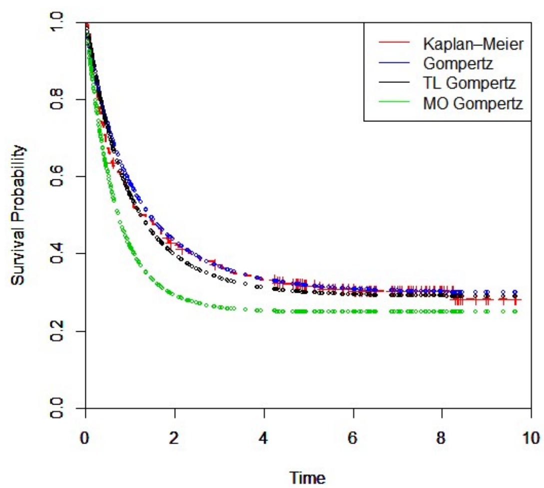

Figure 9: Survival curves for the fitted defective Gompertz, defective MO Gompertz and defective TL Gompertz distributions for the melanoma data set.

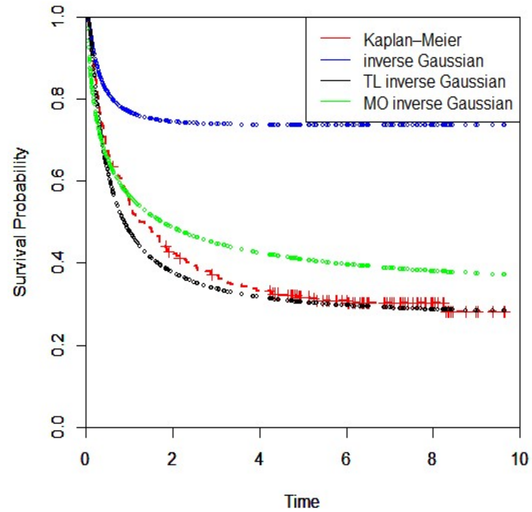

Figure 10: Survival curves for the fitted defective inverse Gaussian, defective MO inverse Gaussian and defective TL inverse Gaussian distributions for the melanoma data set.

6.3 Lung cancer data

The data describes the survival in patients with advanced lung cancer from the North Central Cancer Treatment Group, which is available in the *smcure* package in R; see [14]. The event of interest here is the censored or death of the patient under the recommended treatment during the study period. There are 228 observed values, of which 63 were censored, with a censored rate of 0.2763. The Kaplan-Meier curve of this data stabilizes at 0.0503.

The proposed distribution is applied to the data set, and the final result is shown in Table 5. Based on the results obtained, we come to the conclusion that the smallest values of AIC and BIC are the TL Gompertz distribution, the second smallest is the Gompertz distribution and the largest one is the MO Gompertz distribution.

According to the value of the parameters that are estimated, we can draw the corresponding survival curve. Figure 11 shows that the fitted curves of the survival function by the MO Gompertz distribution is closer to the Kaplan-Meier curves in the range of (0.2, 0.5). The fitted curves of the survival function by the TL Gompertz distribution is closer to the Kaplan-Meier curves within the range of (0, 0.1) and (0.5, 1). It shows that the fitted curves of the survival function by the TL inverse Gaussian distribution is closer to the Kaplan-Meier curves on the whole in the Figure 12.

Table 5: MLEs for the fitted distributions for the lung data set.

Model	$\hat{\alpha}$	$\hat{r}$	$\hat{a}$	$\hat{b}$	$\hat{p}$	Max	AIC	BIC
Gompertz	-	-	-2.2794	3.9158	0.5587	-52.9484	109.8968	116.7555
TL Gompertz	2.8225	-	-3.6645	4.9703	0.5198	-49.3093	104.6186	109.4772
MO Gompertz	-	1.7661	-2.9145	4.8015	0.6790	-62.5651	131.1302	135.9889
inverse Gaussian	-	-	-1.2849	8.4609	0.2619	-94.3641	192.7283	199.5870
TL inverse Gaussian	1.6931	-	-0.8203	9.5465	0.0249	-82.8689	171.7397	176.5984
MO inverse Gaussian	-	1.5100	-2.1009	11.2738	0.4055	-113.8730	233.7459	238.6046

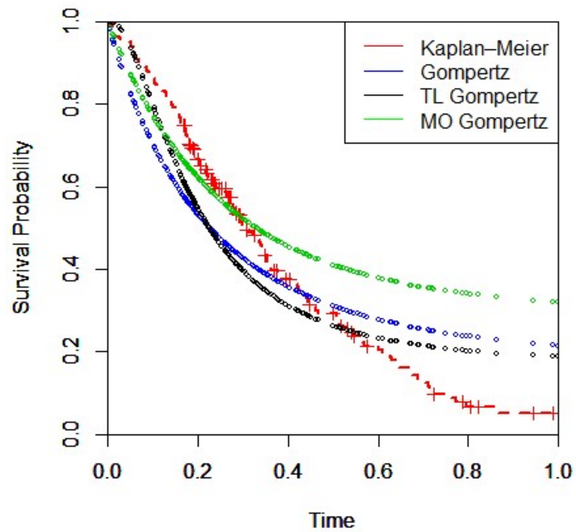

Figure 11: Survival curves for the fitted defective Gompertz, defective MO Gompertz and defective TL Gompertz distributions for the lung cancer data set.

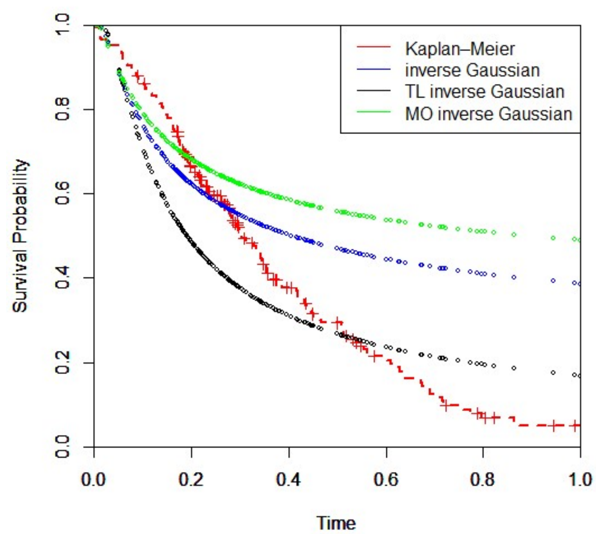

Figure 12: Survival curves for the fitted defective inverse Gaussian, defective MO inverse Gaussian and defective TL inverse Gaussian distributions for the lung cancer data set.

For all three data sets, the defective TL Gompertz distribution is superior to the defective Gompertz and defective MO Gompertz distributions, and the survival curve fitted by the defective TL Gompertz is closer to the Kaplan-Meier curve. After comparing the values of AIC and BIC, we get the defective TL inverse Gaussian distribution which is better than the defective inverse Gaussian and defective MO inverse Gaussian distributions for all three data sets, and all the fitted survival curves by the defective TL inverse Gaussian are closer to the Kaplan-Meier curve.

7 Conclusion

In this paper, we propose two new defective distributions: the defective TL Gompertz distribution and the defective TL inverse Gaussian distribution. To illustrate the useful of our proposed model, we compare it with other defective models. The results show that the defective models based on the TL distribution is a competitive model for modeling the proportion of cured. Further research, like the regression models, will be studied.

Acknowledgments

We would like to express our gratitude to a referee for providing insightful comments that contributed to improving this paper.

References

[1] Berkson, J., Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**(259), 501-515.
[2] Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B* (Methodological), **11**(1), 15-53.
[3] Balka, J., Desmond, A. F., McNicholas, P. D. (2009). Review and implementation of cure models based on first hitting times for Wiener processes. *Lifetime data analysis*, **15**(2), 147.
[4] Balka, J., Desmond, A. F., McNicholas, P. D. (2011). Bayesian and likelihood inference for cure rates based on defective inverse Gaussian regression models. *Journal of Applied Statistics*, **38**(1), 127-144.
[5] Cantor, A. B., Shuster, J. J. (1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statistics in Medicine*, **11**(7), 931-937.
[6] Chen, Z. (1997). Parameter estimation of the Gompertz population. *Biometrical Journal*, **39**(1), 117-124.
[7] Gen, A. I., Moments of order statistics of Topp−Leone distribution. *Statistical Papers*, *53*(1), 117-131, 2012.
[8] Genc, A. I., Estimation of P (X> Y) with Topp-Leone distribution. *Journal of Statistical Computation and Simulation*, ***83***(2), 326-339, 2013.
[9] Gieser, P. W., Chang, M. N., Rao, P. V., Shuster, J. J., Pullen, J. (1998). Modelling cure rates using the Gompertz model with covariate information. *Statistics in medicine*, **17**(8), 831-839.
[10] Ghitany, M. E., Kotz, S., Xie, M. (2005). On some reliability measures and their stochastic orderings for the Topp-Leone distribution. *Journal of Applied Statistics*, **32**(7), 715-722.
[11] Ibrahim, J. G., Chen, M. H., Sinha, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, **57**(2), 383-388.
[12] Kotz, S., Seier, E. (2007). Kurtosis of the Topp-Leone distributions. *Interstat*, **1**, 1-15.
[13] Laurie, J. A., Moertel, C., Fleming, T. R., Wieand, H. S., Leigh, J. E., Rubin, J., Malliard, J. (1989). Surgical adjuvant therapy of large-bowel carcinoma: an evaluation of levamisole and the combination of levamisole and fluorouracil. The North Central Cancer Treatment Group and the Mayo Clinic. *Journal of Clinical Oncology*, *7*(10), 1447-1456.
[14] Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., Klatt, N. E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. *Journal of Clinical Oncology*, **12**(3), 601-607.

[15] Makany, R. (1991). A theoretical basis for Gompertz's curve. *Biometrical Journal*, **33**(1), 121-128.

[16] Maller, R. A., Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, 1197-1205.

[17] Nadarajah, S., Kotz, S. (2003). Moments of some J-shaped distributions. *Journal of Applied Statistics*, **30**(3), 311-317.

[18] Rocha, R., Nadarajah, S., Tomazella, V., Louzada, F. (2016). Two new defective distributions based on the Marshall-Olkin extension. *Lifetime data analysis*, **22**(2), 216-240.

[19] Rocha, R., Nadarajah, S., Tomazella, V., Louzada, F. (2017). A new class of defective models based on the Marshall-Olkin family of distributions for cure rate modeling. *Computational Statistics Data Analysis*, **107**, 48-63.

[20] Rocha, R., Nadarajah, S., Tomazella, V., Louzada, F., Eudes, A. (2017). New defective models based on the Kumaraswamy family of distributions with application to cancer data sets. *Statistical methods in medical research*, **26**(4), 1737-1755.

[21] Rezaei, S., Sadr, B. B., Alizadeh, M., Nadarajah, S. (2017). Topp-Leone generated family of distributions: Properties and applications. *Communications in Statistics-Theory and Methods*, **46**(6), 2893-2909.

[22] Sindhu, T. N., Saleem, M., Aslam, M., Bayesian estimation for Topp Leone distribution under trimmed samples. *Journal of Basic and Applied Scientific Research*, **3**(10), 347-360, 2013.

[23] Topp, C. W., Leone, F. C. (1955). A family of J-shaped frequency functions. *Journal of the American Statistical Association*, **50**(269), 209-219.

[24] Wu, J. W., Hung, W. L., Tsai, C. H. (2004). Estimation of parameters of the Gompertz distribution using the least squares method. *Applied mathematics and computation*, **158**(1), 133-147.

[25] Zghoul, A. A., Record values from a family of J-shaped distributions. *Statistica,* **71**(3), 355-365, 2011.

[26] Zhou, M., Yang, D. W., Wang, Y., Nadarajah, S., Some J-shaped distributions: Sums, products and ratios. In RAMS'06. *Annual Reliability and Maintainability Symposium*, 175-181, 2006.