# Application of Cure Rate Model Based on Zeghdoudi Distribution Using Real life data

**M. Ahajeeth[1], R. Kannan[2]**

[1]Research Scholar, Department of Statistics, Annamalai University, Chidambaram, Tamil Nadu-608 002, India
[2]Professor, Department of Statistics, Annamalai University, Chidambaram, Tamil Nadu-608 002, India

**Abstract: This paper focus on a new two parametric survival model for modelling the population of interest consists of long term survivors. This proposed model is said to be Zeghdoudi Cure Rate model. Various statistical properties like moments and maximum likelihood estimation is established. Finally, a model application to real data is shown, and the fit of many other well-known two-parameter distributions is compared.**

**Keywords: Zeghdoudi Distribution, Gamma distribution, Exponential distribution, Moments, Maximum Likelihood Estimation, Standard Cure Rate.**

## I. INTRODUCTION

A specialised area of mathematical statistics known as survival analysis, or failure time analysis, was established to investigate a specific type of random variable with positive values and censored observations, of which failure time or survival time events are the most frequent. Even though failure or survival are often not correlated to the events, "time to event" could be a better term. In analysing the data, survival analysis is carried out to take into consideration how often time will elapse before the event. The response is generally referred to as the event time, failure time, or survival time.

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. The time variable is usually referred to as survival time, because it gives the time that an individual has "survived" over some follow-up period.

Othus et al. (2012) discussed and revealed that cure models can be a useful tool to analyse and describe cancer survival data. Yin et al. (2014) compared Poisson cure models with Compound Poisson Cure models and revealed that Compound Poisson Cure model gives the better fit. Yin et al. (2009) proposed a cure rate models for modelling the right censored failure time data. Kim et al. (2007) dealt with the cure rate model with interval censored data where the diseases that often progress without symptoms. Guha et al. (2007) introduced a new mixture cure rate model from the perspective of competing risk and model the dependence between censoring and survival time by using the class of Archimedean coupla models. Tsodikov et al. (2015) appealing that bounded cumulative hazard model in cure estimation is better than the two component mixture models. Rodrigues et al. (2009) presented a new flexible cure rate model assuming that the causes of time to event follows Conway – Maxwell Poisson distribution. Chin – Shang Li et al. (2001) investigated the identifiability two forms of cure model namely standard curve model and Non – Mixture PH model. Peng and Dear (2000) discussed a non – parametric mixture model using the breast cancer data. Messaadia et al. (2018) proposed a new one parameter Zeghdoudi distribution with properties to model real life data. Molay Kumar Ruidas (2020) introduced a new two parameter Transmuted Zeghdoudi distribution which is an extension of one parameter Zeghdoudi distribution. Chouia et al. (2021) proposed a new one parameter size biased Zeghdoudi distribution to model count data which structurally excludes zero counts. Hamida et al. (2021) proposed a new model, the upper truncated Zeghdoudi distribution. Hussain et al. (2022) have deliberate a new version of one parameter Zeghdoudi distribution known as Weighted two parameters Zeghdoudi distribution. This new distribution is generated using weighting technique. Ahajeeth et al. (2022) presented a new two parameter survival model named Garima cure rate model for modelling lifetime data. Ahajeeth et al. (2022) developed a new two parameter survival model named Generalized Aradhana cure rate model for the population of interest that consists long term survivors or immunes.

Cure rate model have been widely used to analyse survival time of cancer patients. Many patients with cancer can be long term survivors of their disease. No patients can be "Cured" of death, so in these situations cure models can be used to model long – term survivors rather than cured patients. Cure models can be useful alternative to the Standard Cox proportional hazard models for data with survival trends. The Survival function for entire population of study is expressed as

(1)

where, is the mixing parameters.

- cured patients, - uncured patients, - survival function for uncured groups.

## II. Zeghdoudi Distribution

Zeghdoudi Distribution was developed by Hamouda Messaadia and Halim Zeghdoudi (2018), is a one – parametric lifetime model for modelling lifetime data. Zeghdoudi Distribution is based on mixtures of the ordinary exponential ( $\theta$ ) and Gamma (3, $\theta$ ) distribution.

The probability density function (pdf) of Zeghdoudi Distribution is given by

$$f(t,\theta)=\frac{\theta^3 x(1+x)e^{-\theta x}}{2+\theta} \quad (2)$$

The cumulative distribution function (cdf) of Zeghdoudi Distribution is given by

$$F_u(t,\theta)=1-\left(\frac{x^2\theta^2+\theta(\theta+2)x+\theta+2}{\theta+2}\right)e^{-x\theta} \quad (3)$$

The survival function of the Zeghdoudi Distribution is given by

$$S_u(t,\theta)=\left(\frac{x^2\theta^2+\theta(\theta+2)x+\theta+2}{\theta+2}\right)e^{-x\theta} \quad (4)$$

The hazard function of the Zeghdoudi distribution is specified by

$$h_u(t,\theta)=\left(\frac{\theta^3x(x+1)}{x^2\theta^2+x(\theta+2)+\theta+2}\right) \quad (5)$$

## III. Zeghdoudi Cure Rate Model

Using the standard cure rate mixture model and substituting equation (4) in equation (1) we obtain zeghdoudi Cure Rate Model and is given by

$$S(t)=\pi+(1-\pi)\left(\frac{x^2\theta^2+\theta(\theta+2)x+\theta+2}{\theta+2}\right)e^{-x\theta} \quad (6)$$

The cumulative distribution of cure rate model is defined by

$$F(t)=(1-\pi)F_u(t) \quad (7)$$

Using equation (3) in equation (7) we will obtain cumulative distribution function (cdf) of Zeghdoudi Cure Rate Model as

$$F(t)=(1-\pi)\left(\frac{\theta^3x(x+1)}{x^2\theta^2+x(\theta+2)+\theta+2}\right) \quad (8)$$

The probability density function (pdf) of Cure Rate Model is given by

$$f(t)=(1-\pi)\,f_u(t) \quad (9)$$

where, $f_u(t)$ is the pdf of uncured individuals.

Using equation (2) in equation (9) we will get the probability density function (pdf) of Zeghdoudi Cure Rate Model as,

$$f(t)=(1-\pi)\left(\frac{\theta^3x(1+x)e^{-\theta x}}{2+\theta}\right) \quad (10)$$

The hazard function Zeghdoudi Cure Rate Model is given by

$$h(t)=(1-\pi)h_u(t) \quad (11)$$

By inserting equation (5) in equation (11), the equation (11) takes the form as

$$h(t)=(1-\pi)\left(\frac{\theta^3x(x+1)}{x^2\theta^2+x(\theta+2)+\theta+2}\right) \quad (12)$$

Figures 1 and 2 shows the probability density function and cumulative distribution function of the Zeghdoudi cure rate model. The pdf of said model showing decreasing trend for different combination of parameters.

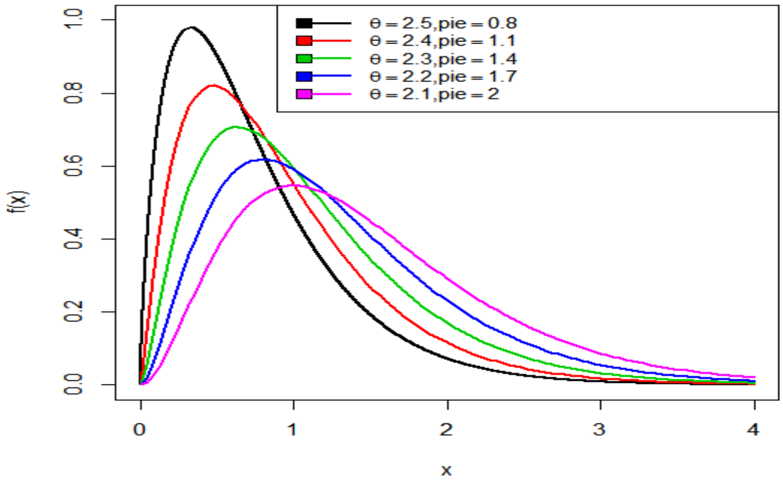


Fig.1:Pdf plot of Zeghdoudi cure rate model

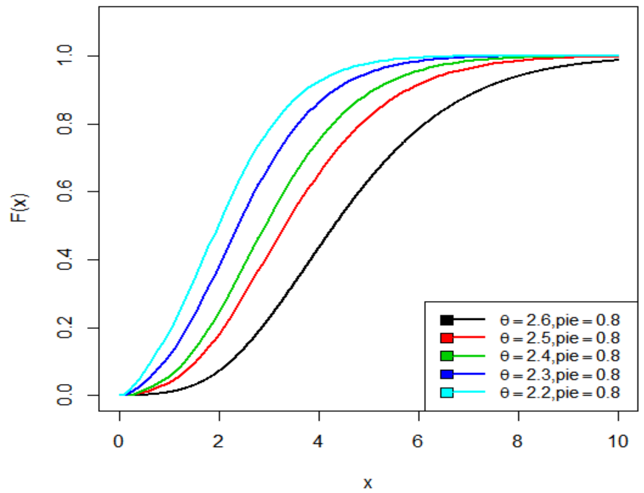


Fig.2 cdf plot of Zeghdoudi cure rate model

Figures 3 and 4 depict the survival plot and hazard function of the Zeghdoudi cure rate model. The hazard rate is increasing and then showing constant trend for different combination of parameters.

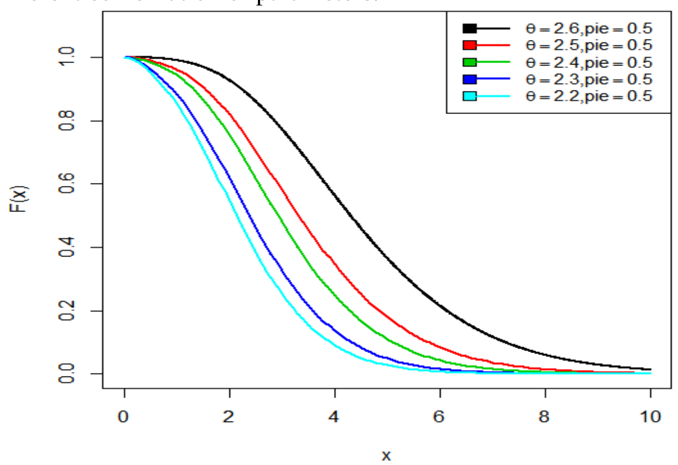


Fig.3 survival function of Zeghdoudi cure rate model

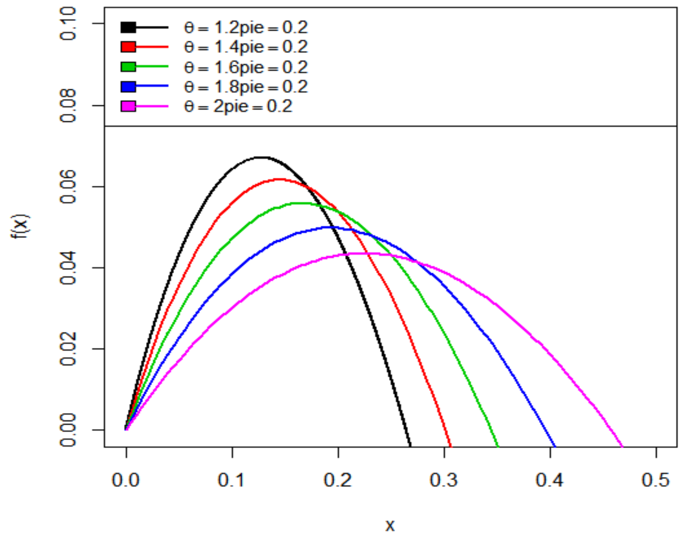


Fig.4:Hazard function of Zeghdoudi cure rate model

## IV. Statistical Properties

Some properties of the proposed model has been discussed in this section.

### IV.I Moments

Let $T$ be the random variable following the Zeghdoudi distribution then the $r^{th}$ moment can be obtained as

$$E\left(T^r\right)=\int_0^\infty t^r f\left(t\right) dt$$

$$=\int_0^\infty t^r \frac{\theta^3 t(1+t)e^{-\theta t}}{\theta+2} dt$$

$$=\left(\frac{\theta^3}{\theta+2}\right)\int_0^\infty t^{r+1}(1+t)e^{-\theta t}dt$$

$$=\left(\frac{\theta^3}{\theta+2}\right)\left[\int_0^\infty t^{r+1}e^{-\theta t}dt+\int_0^\infty t^{r+2}e^{-\theta t}dt\right]$$

$$=\left(\frac{\theta^3}{\theta+2}\right)\left[\int_0^\infty t^{r+2-1}e^{-\theta t}dt+\int_0^\infty t^{r+3-1}e^{-\theta t}dt\right]$$

$$=\left(\frac{\theta^3}{\theta+2}\right)\left[\frac{\Gamma r+2}{\theta^{r+2}}+\frac{\Gamma r+3}{\theta^{r+3}}\right]$$

$$=\frac{\theta^3}{\theta^r(\theta+2)}\left[\frac{\theta(r+1)!+(r+2)!}{\theta^3}\right]$$

$$\mu_r'=\frac{\theta(r+1)!+(r+2)!}{\theta^r(\theta+2)} \qquad (13)$$

$$\mu_1'=E(X)=\frac{2\theta+6}{\theta(\theta+2)}$$

$$\mu_2'=E(X^2)=\frac{6\theta+24}{\theta^2(\theta+2)}$$

$$\mu^2=Variance=\mu_2'-(\mu_1')^2$$

$$=\frac{6\theta+24}{\theta^2(\theta+2)}-\frac{(2\theta+6)^2}{\theta^2(\theta+2)^2}$$

$$=\frac{(6\theta+24)(\theta+2)-(4\theta^2+36+12\theta)}{\theta^2(\theta+2)^2}$$

$$=\frac{6\theta^2+12\theta+24\theta+48-4\theta^2-36-12\theta}{\theta^2(\theta+2)^2}$$

$$=\frac{2\theta^2+24\theta-12}{\theta^2(\theta+2)^2}$$

$$\mu_3'=E(X^3)=\frac{4!\theta+5!}{\theta^3(\theta+2)}$$

$$=\frac{24\theta+120}{\theta^3(\theta+2)}$$

$$\mu_4'=E(X^4)=\frac{5!\theta+6!}{\theta^4(\theta+2)}$$

$$=\frac{120\theta+720}{\theta^4(\theta+2)}$$

$$S.D.=\sqrt{\frac{2\theta^2+24\theta-12}{\theta^2(\theta+2)}}$$

$$Co-efficeient\ of\ variation=\frac{\sigma}{\bar{x}}$$

$$= \frac{\sqrt{2\theta^2 + 24\theta - 12}}{\theta(\theta+2)} \times \frac{\theta(\theta+2)}{2\theta+6}$$

$$= \sqrt{\frac{2\theta^2 + 24\theta - 12}{2\theta+6}}$$

**V. Parameter Estimation for Censored Data**

In this section, we will estimate the parameters of the proposed model by using maximum likelihood technique.

**V.i Maximum Likelihood Estimation**

Let $t_1, t_2 \ldots, t_n$ be the random sample of size n drawn from population with probability density function $f(t, \theta, \pi)$. Then the likelihood function of $n$ observations is given by

$$L = \prod_{i=1}^{n} \left(f(t_i)\right)^{\delta_i} \left(S(t_i)\right)^{1-\delta_i}$$

$$= \prod_{i=1}^{n} \left((1-\pi) f_u(t_i)\right)^{\delta_i} \left(\pi + (1-\pi) S_u(t)\right)^{1-\delta_i}$$

where, $\delta_i$ is censoring indicator variable, $\delta_i = 1$, for observed lifetime and $\delta_i = 0$ for censored lifetime.

$$L = \prod_{i=1}^{n} \left[\frac{(1-\pi)\theta^3 t_i (1+t_i) e^{-\theta t_i}}{\theta+2}\right]^{\delta_i} \left[\pi + (1-\pi)\left[t_i^2\theta^2 + \theta(\theta+2)t_i + (\theta+2)e^{-\theta t_i}\right]\right]^{1-\delta_i}$$

The log – likelihood function of Zeghdoudi Cure rate model is defined as

$$\log L = \sum_{i=1}^{n} \log\left[\frac{(1-\pi)\theta^3 t_i (1+t_i) e^{-\theta t_i}}{\theta+2}\right]^{\delta_i} + \sum_{i=1}^{n} \log\left[\pi + (1-\pi)\left[t_i^2\theta^2 + \theta(\theta+2)t_i + (\theta+2)\right] e^{-\theta t_i}\right]^{1-\delta_i}$$

$$\log L = \sum_{i=1}^{n} \delta_i \left[\log(1-\pi) + \log\theta^3 t_i + \log(1+t_i) - \theta t_i - \log(\theta+2)\right] +$$
$$\sum_{i=1}^{n} (1-\delta_i)\left[\log\pi + \log(1-\pi) + \log\left(t_i\theta^2\right)\right] + \log\left[t_i\left(\theta(\theta+2)\right) + \log(\theta+2) - \theta t_i\right] \quad (14)$$

Differentiating with respect to $\pi$ and $\theta$ and setting equal to zero.

$$\frac{\partial \log L}{\partial \pi} = \sum_{i=1}^{n} -\delta_i\left(\frac{1}{1-\pi}\right) + \sum_{i=1}^{n} (1-\delta_i)\left[\frac{1}{\pi} - \frac{1}{1-\pi}\right]$$

$$\frac{\partial \log L}{\partial \pi} = \left(\frac{1}{1-\pi}\right)\sum \delta_i + \frac{1}{\pi}\sum(1-\delta_i) - \left(\frac{1}{1-\pi}\right)\sum(1-\delta_i)$$

$$\frac{\partial \log L}{\partial \pi} = \left(\frac{1}{1-\pi}\right)\left(\sum \delta_i - \sum(1-\delta_i)\right) + \frac{1}{\pi}\sum(1-\delta_i) \quad (15)$$

$$\frac{\partial \log L}{\partial \theta} = \sum_{i=1}^{n} \delta_i\left[\frac{3}{\theta} - t_i - \frac{1}{\theta+2}\right] + \sum_{i=1}^{n} (1-\delta_i)\left[\frac{2}{\theta} + \frac{1}{\theta+2} - t_i + \frac{1}{\theta} + \frac{1}{\theta+2}\right]$$

$$\frac{\partial \log L}{\partial \theta} = \sum_{i=1}^{n} \delta_i\left[\frac{3(\theta+2) - \theta(\theta+2)t_i - \theta}{\theta(\theta+2)}\right] + \sum_{i=1}^{n} (1-\delta_i)\left[\frac{2}{\theta+2} + \frac{3}{\theta} - t_i\right]$$

$$\partial \log L = \sum_{i=1}^{n} \delta_i\left[\frac{2\theta - \theta(\theta+2)t_i + 6}{\theta(\theta+2)}\right] + \sum_{i=1}^{n} (1-\delta_i)\left[\frac{5\theta + 6 - \theta(\theta+2)t_i}{\theta(\theta+2)}\right]$$

$$\frac{\partial \log L}{\partial \theta} = \left(\frac{2\theta - \theta(\theta+2)t_i + 6}{\theta(\theta+2)}\right)\sum_{i=1}^{n} \delta_i + \left[\frac{5\theta - \theta(\theta+2)t_i + 6}{\theta(\theta+2)}\right]\sum_{i=1}^{n} (1-\delta_i) \quad (16)$$

The solution of equations (15) and (16) gives the MLE of the parameters for the Zeghdoudi cure rate model. However, the equations are complex so it cannot be solved analytically, thus we solved numerically using R programming with some data set.

**VI. Data Analysis**

This model incorporated real-life data, including individuals with a lung cancer, to compare the performance of existing models to a newly established model.

**Data set**

The data set consists of survival times (in months) of 64 patients suffering from lung cancer reported by Pena (2002). The data set are given below:
0.99, 1.28, 1.77, 1.97, 2.17, 2.63, 2.66, 2.76, 2.79, 2.86, 2.99, 3.06, 3.15, 3.45, 3.71, 3.75, 3.81, 4.11, 4.27, 4.34, 4.40, 4.63, 4.73, 4.93, 4.93, 5.03, 5.16, 5.17, 5.49, 5.68, 5.72, 5.85, 5.98, 8.15, 8.62, 8.48, 8.61, 9.46, 9.53, 10.05, 10.15, 10.94, 10.94, 11.24, 11.63, 12.26, 12.65, 12.78, 13.18, 13.47, 13.96, 14.88, 15.05, 15.31, 16.13, 16.46, 17.45, 17.61, 18.20, 18.37, 19.60, 20.70, 22.54, 23.36.
To evaluate the performance of the models using the goodness of fit criterion in order to compare them. The formulas are listed below.

$$AIC = 2P - 2\log L, \quad BIC = P\log n - 2\log L, \quad AICC = AIC + \frac{2P(P+1)}{n-P-1}$$

In Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Corrected Akaike Information Criterion (AICc), a term that measures how well the model fits the data is combined with a term that penalises the model proportional to the number of parameters.

**Table.1: Survival Times (in months) of 64 Lung Cancer Patients**

| Model | Parameter | S.E | -2log L | AIC | BIC | AICC |
|---|---|---|---|---|---|---|
| Zeghdoudi Cure Rate Model | $\hat{\theta}$= 0.38491<br>$\hat{\pi}$= 0.21054 | $\hat{\theta}$= 0.04570<br>$\hat{\pi}$= 0.34711 | 389.084 | 391.084 | 393.4017 | 391.2807 |
| Exponential Cure Rate Model | $\hat{\theta}$= 0.82641<br>$\hat{\pi}$= 0.53761 | $\hat{\theta}$= 0.07863<br>$\hat{\pi}$= 0.03879 | 392.0752 | 394.0752 | 396.234 | 394.1719 |
| Lindley Cure Rate Model | $\hat{\theta}$= 0.11481<br>$\hat{\pi}$= 0.06343 | $\hat{\theta}$= 0.01435<br>$\hat{\pi}$= 0.01360 | 405.0524 | 407.0524 | 409.2113 | 407.1491 |

As shown by the results in the above table, the Zeghdoudi cure rate rate model has a better fit in terms of the AIC, BIC, and AICC because it contains the lowest value for the data set that has been selected for comparison between the existing cure rate model and the new model.

**7. Conclusion**

In this present study, a new two parameter survival model has been introduced and the model is known as Zeghdoudi cure rate model. The Zeghdoudi cure rate model is a mixture model having two parameters obtained from the standard cure rate model provided by Boag (1949). The model's properties have been established. The parameters of the proposed model are generated using the maximum likelihood technique. Utilizing real life data, the model's accuracy is evaluated. The study's findings indicate that the Zeghdoudi cure rate approach represents a superior fit for the data under consideration.

**References**

1. H. Messadia and H. Zeghdoudi, (2018), Zeghdoudi distribution and its applications, International Journal of Computing Science and Mathematics, 9(1), 58 – 65.
2. T. Hamida and A. Hiba, (2021), On Truncated Zeghdoudi Distribution: Posterior Analysis Under Different Loss Functions for Type II Censored Data, Pakistan Journal of Statistics and Operation Research, 17(2), 497 – 508.
3. Molay Kumar Ruidas, (2020), Transmuted Zeghdoudi Distribution: Theory and Applications, International Journal of Scientific Research in Mathematical and Statistiacal Sciences, 7(2), 95 – 101.
4. B.A. Hussain and R.S. Mohammed, (2022), Weighted Zeghdoudi Distribution: Properties and Applications, ResearchJet Journal of Analysis and Inventions, 3(4), 50 – 63.
5. S. Chouia, H. Zeghdoudi, R. Vinoth and A. Beghriche, (2021), A New Size Biased Distribution with Application, Journal of Applied Probability and Statistics, 16(1), 111 – 125.
6. Chin – Shang Li, Jeremy M.G. Taylor and P. Judy (2001), Identifiability of Cure Models, Statistics and Probability Letters, 54, pp 389 – 395.
7. David G. Kleinbaum and Mitchel Klein, (2005), A self-learning text, Survival Analysis, Springer, pp. 17.
8. Josemar Rodrigues, Mariode castro, Vincente G.Cancho and N. Balakrisnan, (2009), COM – Poisson Cure Rate Survival Models and An Application to a Cutaneous Melanoma Data, Journal of Statistical Planning and Inference, 139, pp. 3605 – 3611.
9. M. Rahimzadeh, B. Kavehie and M. Reza Zali, (2014), Cure Models in Analysing Long term Survivors, Transl Gastrointest Cancer, 3(4), 149 – 154.
10. Yingwei Peng and Keith B. G. Dear, (2000), A Non – Parametric Mixture Model for Cure Rate Estimation, Biometrics 56, pp. 237 – 243.
11. Y. Li, R. C. Tiwari and S. Guha, (2007), Mixture Cure Survival Models with Dependent Censoring, J. R. Statist. Soc.B, 69, pp. 285 – 306.
12. Guosheng Yin and Joseph G. Ibrahim (2005), Cure Rate Models: a Unified Approach, The Canadian Journal of Statistics, vol. 33, No. 4, pp. 559 – 570.
13. M. Ahajeeth and R. Kannan, (2022), Utilization of a Mixture Cure Rate Model based on the Garima Distribution Using Real Life Data, High Technology Letters, 28(8), 385 – 394.

14. M. Ahajeeth and R. kannan, (2022), Generalized Cure Rate Model Based on Aradhana Distribution, International Journal of Statistics and Applied Mathematics, 7(5), 135 – 141.
15. Megan Othus, Bart Barlogie, Michael L. Leblanc and John J. Crowley (2012), Cure Models as a Useful Statistical Tool for Analyzing Survival, NIH Public Access, 18(14)3731 – 3736.A.D. Tsodikov, J.G. Ibrahim and [16] A.Y. Yakovlev (2015), Estimating Cure Rates From Survival Data, Journal of American Statistical Association, DOI: 1198/016221450300000001007.
16. A.E. Pena (2022), Goodness of fit test with censored data, http://www.stat.sc.edu/pena/TalksPresented/TalkAtUSC Aug 2000.pdf.