ISSN 2664-4150 (Print) & ISSN 2664-794X (Online)
South Asian Research Journal of Engineering and Technology
Abbreviated Key Title: South Asian Res J Eng Tech

| Volume-6 | Issue-6 | Nov-Dec- 2024 | **DOI: https://doi.org/10.36346/sarjet.2024.v06i06.002**

Original Research Article

The Comparison some 1-Parameter Models such as Lindley (LL), Sujatha (SJ), Amarendra (AR), Devya (DV), and Shambhu (SB) Distribution for Duration Diabetes Data (YEARS)

Muhammad Iqbal Lubis[1*], Rado Yendra[1], Ari Pani Desvina[1], Muhammad Marizal[1]

[1]Universitas Islam Negeri Sultan Syarif Kasim Riau, Panam, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau 28293, Indonesia

***Corresponding Author:** Muhammad Ikbal Lubis
Universitas Islam Negeri Sultan Syarif Kasim Riau, Panam, Jl. HR. Soebrantas No.Km. 15, RW.15, Simpang Baru, Kota Pekanbaru, Riau 28293, Indonesia

Article History
Received: 17.10.2024
Accepted: 22.11.2024
Published: 26.11.2024

Abstract: This study examined the probability of death using survival data of individuals with diabetes. The main objective of this study was to identify the most appropriate distribution to represent the survival time of patients with diabetes from 50 individuals in the Bengkalis area using 1-parameter probability modeling, such as the Lindley (LL), Sujatha (SJ), Amarendra (AR), Devya (DV), and Shambhu (SB) distributions. The maximum likelihood method was used to estimate the parameter values of the distributions used in this study. In addition, graphical assessment (density-density plots) and numerical criteria (Akaike's Information Criterion (AIC), -2 log likelihood (-2*l)) will be used to determine the most appropriate model. In most cases, the results obtained from the graphical assessment were consistent but differed from the numerical criteria. The model with the lowest AIC value (-2*l) is selected as most appropriate. Overall, the Amarendra and Devya distributions were identified as the most appropriate models.

Keywords: Lindley, Sujatha, Amarendra, Devya, Shambhu, Diabetes.

INTRODUCTION

The Indonesian Minister of Health reported that diabetes was most prevalent in Riau Province, with an increase of 358.3% [1], whereas in Bengkalis, diabetes affected 10.57% of the population in 2019 [2]. Understanding the survival time of patients with diabetes is essential for estimating the risk of death from diabetes. Survival time studies can be conducted using statistical techniques to develop models that accurately represent survival patterns. Several studies have aimed to determine the optimal probability model for the use of diabetes survival time data. Most of the probability models used have 2 or more parameters, including: Alka and Gurpit [3] used the Weibull distribution to estimate the time to onset of nephropathy in patients with type 2 diabetes. Gurpit *et al*., discussed the estimation of survival functions in patients with diabetic nephropathy using distributions such as exponential, gamma, Weibull, log-normal, inverse Gaussian, and Rayleigh. The authors concluded that the gamma distribution is the most effective predictor of survival in patients with diabetic nephropathy. Previous studies have determined the best probability model for the survival time data of patients with diabetes. Ummu *et al*., [4] estimated the duration of diabetes survival using the Weibull, Gamma, and Log-Normal distributions. The results showed that the Weibull model best approximated the given observational data. This was also supported by numerical models, such as AIC and BIC, which gave the smallest values for both numerical methods compared to other probability models. Furthermore, Manda Lisa Usvita *et al*., [5] compared three distributions, Exponential (E), Weibull (W), and Rayleigh-Lomax (RL), which were applied to the survival time of patients with diabetes. Method of Moments is used to obtain the estimated parameters. Based on the smallest Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) values and graphical examination (probability density function (pdf)) of diabetes patient survival time, this study showed that RL is the most appropriate distribution for modeling diabetes patient survival

CITATION: Muhammad Iqbal Lubis, Rado Yendra, Ari Pani Desvina, Muhammad Marizal (2024). The Comparison some 1-Parameter Models such as Lindley (LL), Sujatha (SJ), Amarendra (AR), Devya (DV), and Shambhu (SB) Distribution for Duration Diabetes Data (YEARS). *South Asian Res J Eng Tech*, *6*(6): 168-173.

time at Mandau Regional Hospital, Bengkalis Regency, Riau Province. Sutriana *et al*., [6] used the Lindley distribution (LIN) and three modified Lindley distributions, namely, weighted Lindley exponential (WLE), Power Modified Lindley (PML), Lindley half-Cauchy (LHC), and Rayleight–Lomax distributions (RL). The most appropriate results were obtained for distributions with the lowest AIC, BIC, and -l values. In general, the Rayleight–Lomax (RL) distribution was selected as the best model. Marvasti *et al*., [7] compared the Cox and parametric models to analyze the effective time factor of type 2 nephropathy events using the log-normal distribution. The results of this study indicate that a log-normal distribution is appropriate for this case. Based on this description, the distributions commonly used in the survival analysis of diabetes patient data with parametric models are the Weibull, exponential, gamma, Rayleigh, and log-normal distributions that have 2 or more parameters. This study focused on the use of several probability models that have 1 parameter in the survival data of patients with diabetes. A total of 5 probability models with 1 parameter were used in this study, such as, Lindley (LL) [8], Sujatha (SJ) [9], Amarendra (AR) [10], Devya (DV) [11], and Shambhu (SB) [12] Distributions. Therefore, this study identified the most appropriate survival time distribution for patients with diabetes based on various goodness-of-fit criteria. A preliminary study was conducted on the survival time of patients with diabetes in 50 patients from the Bengkalis region. The proposed distribution was compared with existing distribution functions to assess its suitability for describing diabetes characteristics. Unknown parameter estimates were calculated using the Maximum Likelihood Method. Graphical methods, such as PDF (density) and CDF (cummulative) plots, and numerical criteria, such as AIC and -2*l, were used to determine the distribution that best fitted the diabetes survival time data.

MATERIALS

Duration 50 patients with diabetes (years) was collected from Mandau Regional General Hospital (RSUD), Bengkalis Regency, Riau Province, are presented in Table 1 and histogram of the diabetes survival times are illustrated in Figure 1.

Table 1: Duration 50 patients with diabetes (years)

3.6	0.7	2.4	5.8	4.6	4.4	7.4	3.2	0.7	6.3
6.1	6.5	1.2	2.3	2.1	1.3	1.5	1.7	3.0	4.3
5.2	6.3	1.8	4.7	4.3	1.8	2.6	7.1	3.4	3.3
0.8	0.3	4.0	3.3	5.8	4.2	5.6	6.0	9.0	1.1
3.0	9.3	2.8	7.3	3.3	2.8	3.1	5.9	4.7	4.9

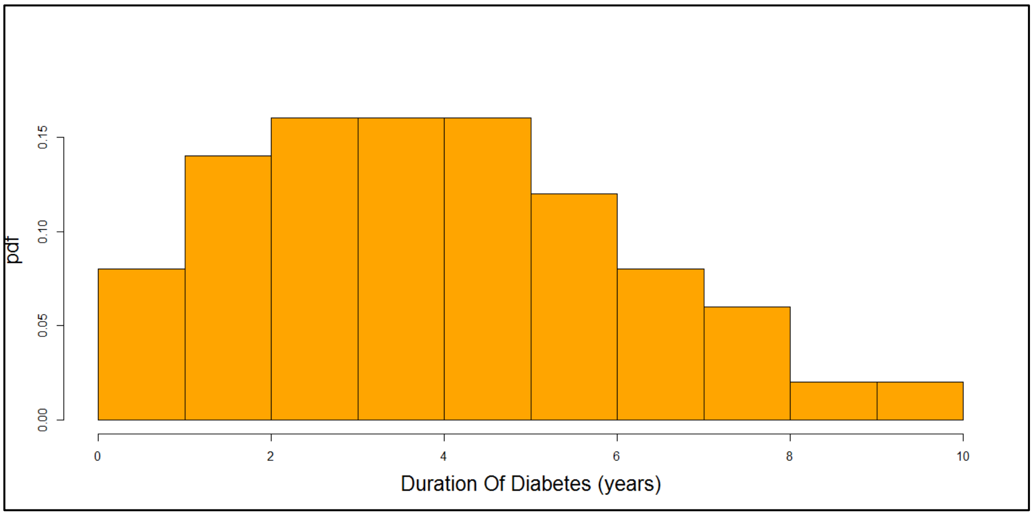

Figure 1: The Histogram Duration of 50 Patients with Diabetes

METHODS

Probability Density Function (pdf) and cumulative Distribution Function (CDF)

In this study, five probability density function (pdf) associated with modeling the duration of 50 patients with diabetes (years), LL, SJ, AR, DV, and SB are considered in this paper. The equations defining the probability density functions (pdf) and cumulative distribution functions (cdf) for various candidate distributions of interest are given below, for each distribution that we consider are given in Tables 2 and 3, respectively, where *x*(duration of diabetes) denote the observed values of the random variable representing the event of interest. In order to fit a particular theoretical distribution to the observed distribution of diabetes duration, parameters were estimated. The parameter estimation of the distribution function was performed using the maximum likelihood method. The parameter estimation of the distribution function was performed using the maximum likelihood method (MLE). The function of maximum likelihood in this model is implicit and complicated and we will not discuss detail in this paper. The nonlinear equation generated by the maximum log likelihood function (*l(Ω), Ω are parameters*) requires a numerical method, namely the Newton rapson method, to obtain the solution of the equation. However, this method was employed in the iteration system to find the solution. Initial values were tested for this procedure. If the initial value covers the same value, then the selected estimated parameter is considered. The goodness-of-fit test procedure for model selection, both numerically and graphically, is discussed.

Table 2: Probability Density Function (Pdf) of various distributions

	Pdf (x)
LL	$f(x;\theta) = \frac{\theta^2}{\theta+1}(1+x)e^{-\theta x}, x > 0, \theta > 0$
SJ	$f(x;\theta) = \frac{\theta^3}{\theta^2+\theta+2}(1+x+x^2)e^{-\theta x}, x > 0, \theta > 0$
AR	$f(x;\theta) = \frac{\theta^4}{\theta^3+\theta^2+2\theta+6}(1+x+x^2+x^3)e^{-\theta x}, x > 0, \theta > 0$
DV	$f(x;\theta) = \frac{\theta^5}{\theta^4+\theta^3+2\theta^2+6\theta+24}(1+x+x^2+x^3+x^4)e^{-\theta x}, x > 0, \theta > 0$
SB	$f(x;\theta) = \frac{\theta^6}{\theta^5+\theta^4+2\theta^3+6\theta^2+24\theta+120}(1+x+x^2+x^3+x^4+x^5)e^{-\theta x}, x > 0, \theta > 0$

Table 3: Distribution Function (Cdf) of various distributions

	Cdf (x)
LL	$F(x;\theta) = 1 - \left(1 + \frac{\theta x}{\theta+1}\right) e^{-\theta x}$
SJ	$F(x;\theta) = 1 - \left(1 + \frac{\theta x(\theta x+\theta+2)}{\theta^2+\theta+2}\right) e^{-\theta x}$
AR	$F(x;\theta) = 1 - \left(1 + \frac{\theta^3 x^3+\theta^2(\theta+3)x^2+\theta(\theta^2+2\theta+6)x}{\theta^3+\theta^2+2\theta+6}\right) e^{-\theta x}$
DV	$F(x;\theta) = 1 - \left(1 + \frac{\theta^4(x^4+x^3+x^2+x)+\theta^3(4x^3+3x^2+2x)+6\theta^2(2x^2+x)+24x\theta}{\theta^4+\theta^3+2\theta^2+\theta+24}\right) e^{-\theta x}$
SB	$F(x;\theta) = 1 - \left(1 + \frac{\theta^5(x^5+x^4+x^3+x^2+x)+\theta^4(5x^4+4x^3+3x^2+2x)+2\theta^3(10x^3+6x^2+3x)+120\theta^2(5x^2+2x)+120x\theta}{\theta^4+\theta^3+2\theta^2+\theta+24}\right) e^{-\theta x}$

Maximum Likelihood Estimate (MLE) and Goodness of Fit Tests (GOF)

Let $(x_1, x_2, \cdots, x_n)$ be random samples from the LL, SJ, AR, DV, and SB distributions. The log likelihood **(*l(Ω)*)** are presented in Table 4. The MLE $\widehat{\Omega}\ of\ \Omega$ is the solution of the equation $\frac{\partial l(\Omega)}{\partial \Omega} = 0$ and thus it is the solution of the following nonlinear equation. The most appropriate distribution was identified using results from several GOFs. The GOF tests considered were based on graphical inspection (pdf plot) and (cdf plot). Numerical criteria like Akaike's information criterion (AIC) and -2*Log Like Lihood (-2*l) was applied to determine the GOF criteria of the distributions. In most cases, graphical inspection gave the same result, but their AIC results differed. The best fitting result was selected as the distribution with the lowest AIC value. The formula of numerical methods such as AIC is exhibited in the following Table 5

Table 4: The Log Likelihood of Various Distributions

	l(x) (Log-Likelihood function)
LL	$l(x;\theta) = 2n\, log(\theta) - n\, log(\theta+1) + \sum_{i=1}^{n} log(1+x_i) - \theta \sum_{i=1}^{n} x_i$
SJ	$l(x;\theta) = 3n\, log(\theta) - n\, log(\theta^2+\theta+2) + \sum_{i=1}^{n} log(1+x_i+{x_i}^2) - \theta \sum_{i=1}^{n} x_i$
AR	$l(x;\theta) = 4n\, log(\theta) - n\, log(\theta^3+\theta^2+2\theta+6) + \sum_{i=1}^{n} log(1+x_i+{x_i}^2+{x_i}^3) - \theta \sum_{i=1}^{n} x_i$
DV	$l(x;\theta) = 5n\, log(\theta) - n\, log(\theta^4+\theta^3+2\theta^2+6\theta+24) + \sum_{i=1}^{n} log(1+x_i+{x_i}^2+{x_i}^3+{x_i}^4) - \theta \sum_{i=1}^{n} x_i$
SB	$l(x;\theta) = 6n\, log(\theta) - n\, log(\theta^5+\theta^4+2\theta^3+6\theta^2+24\theta+120) + \sum_{i=1}^{n} log(1+x_i+{x_i}^2+{x_i}^3+{x_i}^4+{x_i}^5) - \theta \sum_{i=1}^{n} x_i$

Table 5: The formulas of numerical criteria for model evaluation

Numerical Criteria	Formula
AIC	$-2l + 2p$
$-2l$	- 2*log likelihood

l = log likelihood, p = Number of parameters

RESULT

In this section, we present a dataset of diabetes survival times to demonstrate the performance of LL, SJ, AR, DV, and SB distributions in practice. The fitting of these distributions was assessed using the data. The computed parameter, AIC and -2*l for various probability density functions are presented in Table 6.

Table 6: Computed parameter, AIC and -2* Log Likelihood (-2*l)

	θ	AIC	-2*l
Lindley	0.431542	227.1103	225.1103
Sujatha	0.6405193	220.7192	218.7192
Amarendra	0.8709197	218.1556	216.1556
Devya	1.115399	218.7033	216.7033
Shambhu	1.368729	221.8296	219.8296

On the graphical presentation of the modeling of the duration of diabetes survival time data (i.e., on the duration of diabetes survival time histogram, the pdf plot and cdf plot for the LL, SJ, AR, DV, and SB distribution models are shown in Figures 2, 3, 4, 5 and 6respectively. When the pdf and cdf plot were examined, it was determined that some distributions yielded similar results. Based on these figures, the AR and DV distribution models could provide good results during diabetes survival time data. However, instead of graphical evaluation, Table 6 provides a more meaningful comparison using the AIC and -2*l values. Table 6 shows the AIC and -2*l values test statistics for the goodness of fit test for the fitness of the duration of diabetes survival time data based on Maximum Likelihood Estimators for LL, SJ, AR, DV, and SB distributions. According to these results, although similar results were obtained for all six distributions, the lowest AIC and -2*l values were obtained for the AR and DV distributions. In conclusion, the AR and DV distributions provide better modeling in terms of the numerical criteria.

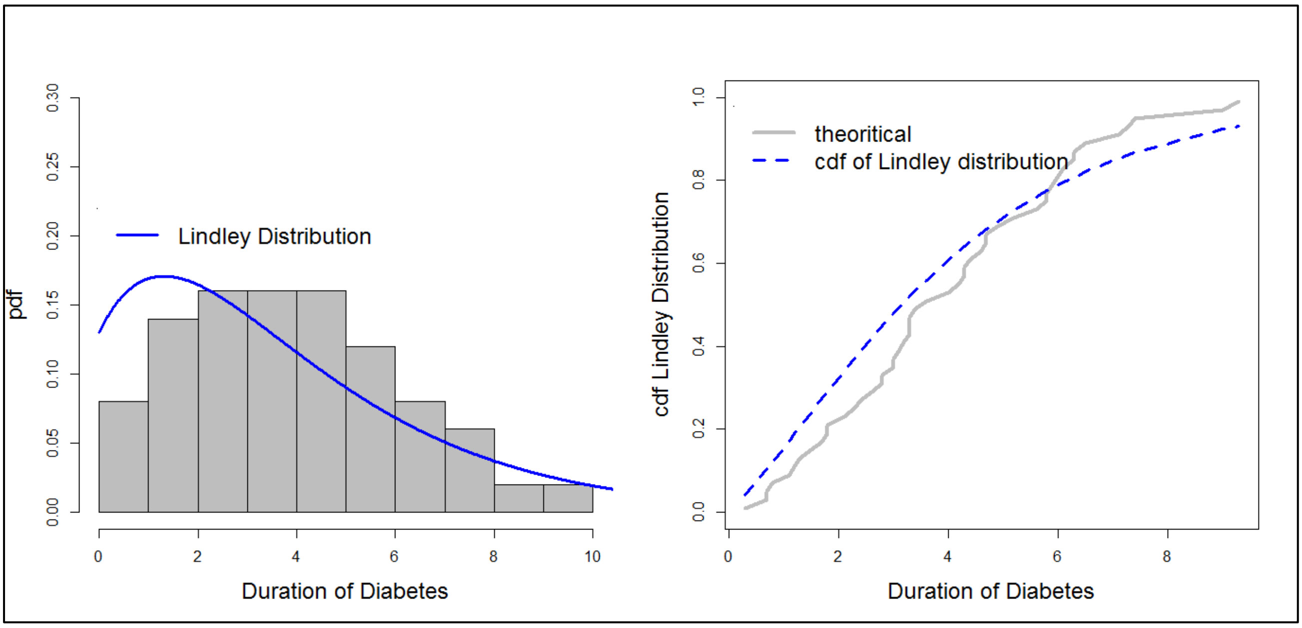

Figure 2: Fitted pdf and cdf of LL distributions, respectively

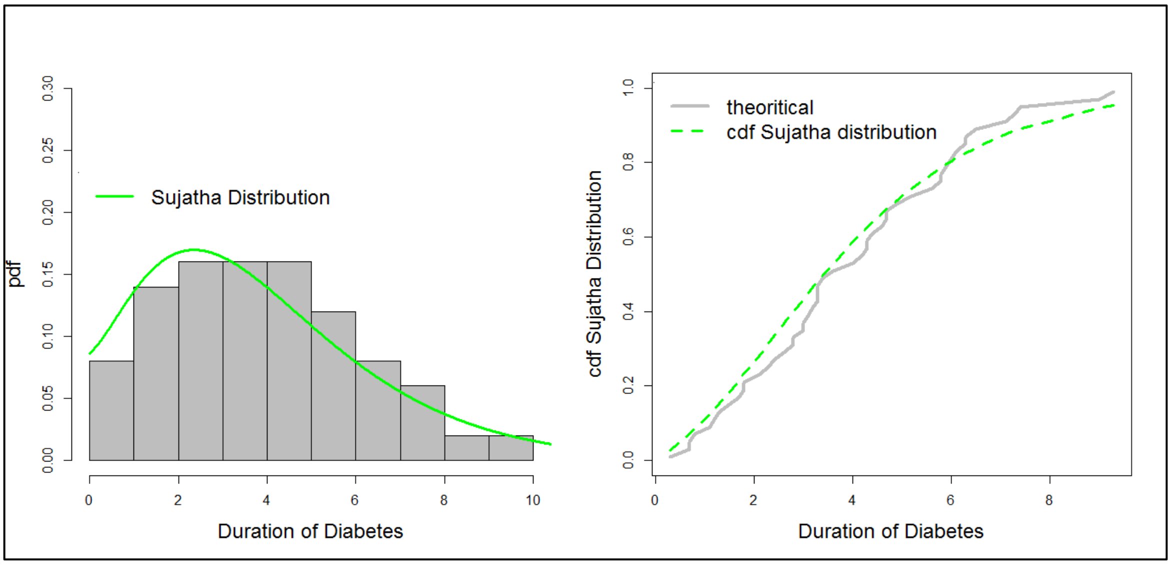

Figure 3: Fitted pdf and cdf of SJ distributions, respectively

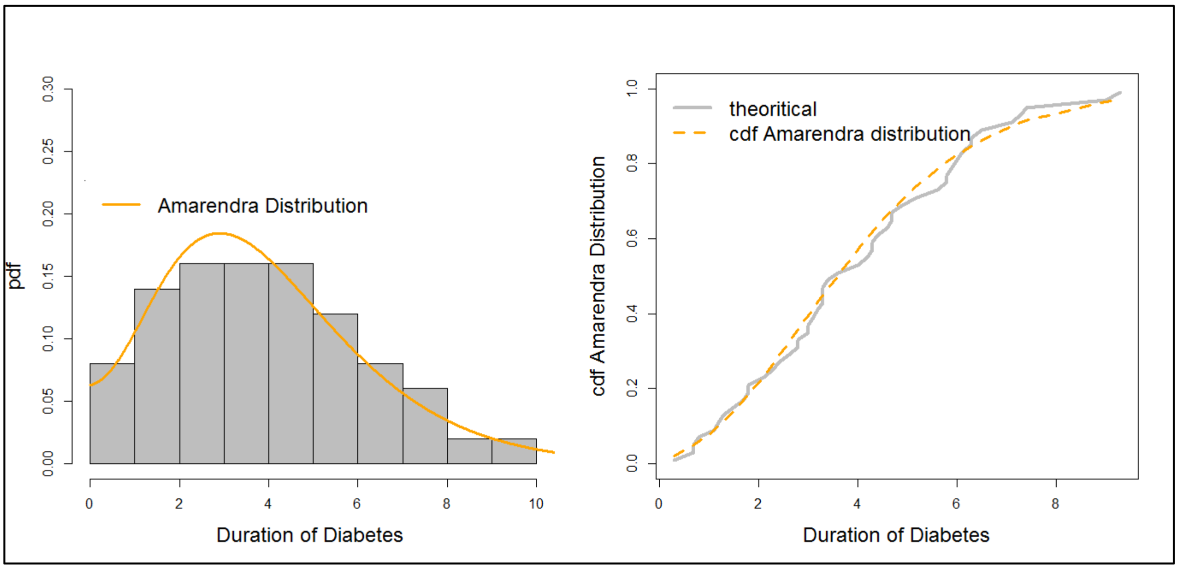

Figure 4: Fitted pdf and cdf of AR distributions, respectively

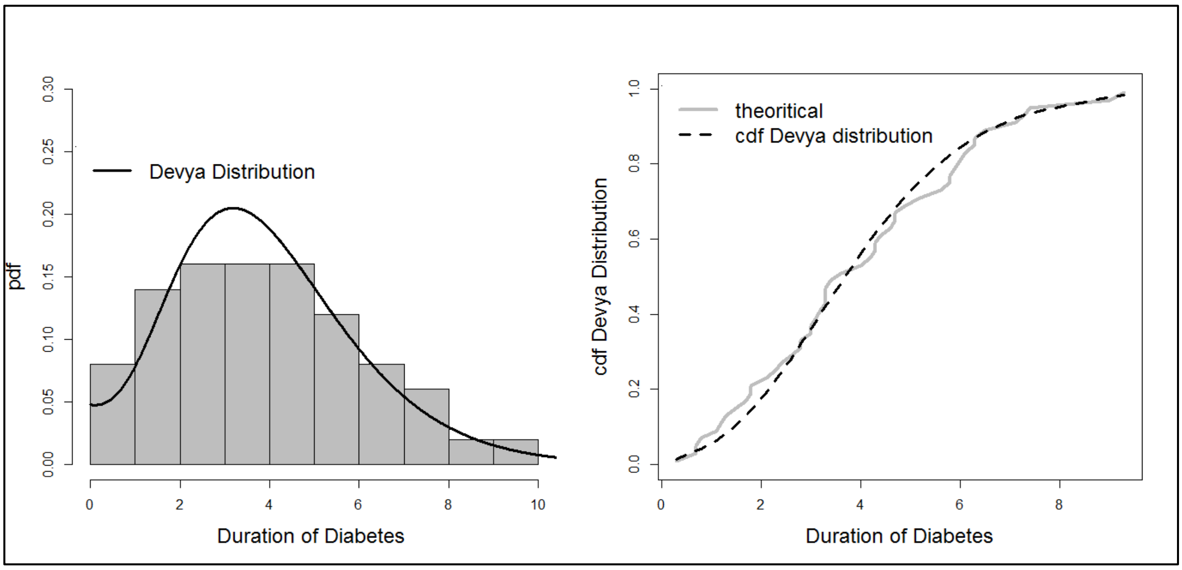

Figure 5: Fitted pdf and cdf of DV distributions, respectively

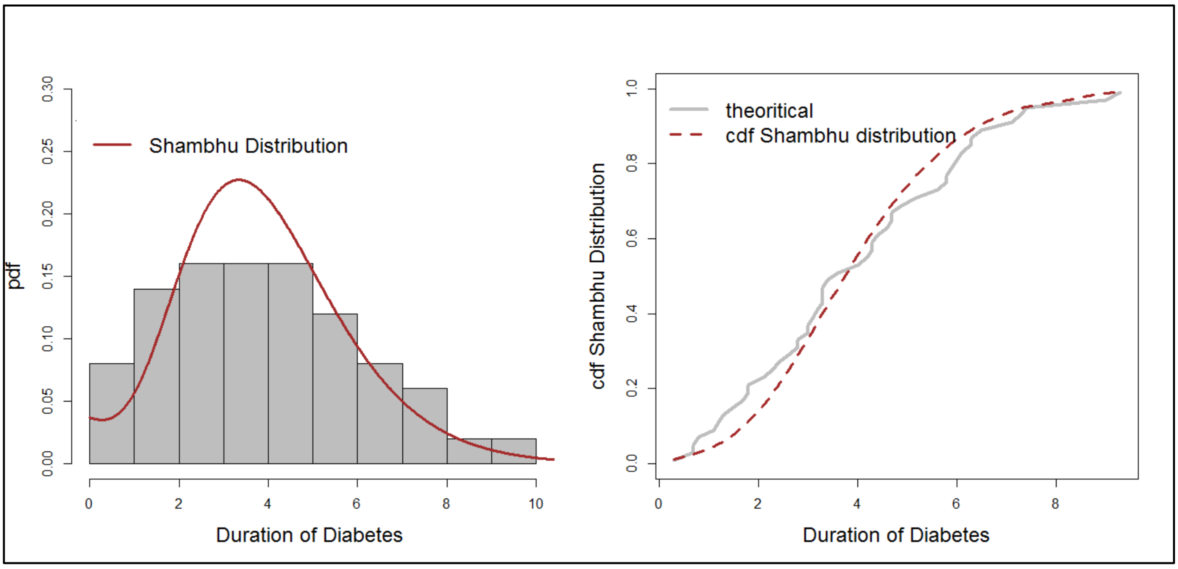

Figure 6: Fitted pdf and cdf of SB distributions, respectively

Furthermore, based on Figure 7, it can be concluded that in this study, the mixture of 4 component (k = 4) Gamma Distribution or Amarendra Distribution will result in a better model.

CONCLUSION

In this study, the focus was on determining the best statistical model for predicting the duration of diabetes survival time data. This study considers 5 distributions with 1-Parameter such as Lindley, Sujatha, Amarendra, Devya and Shambhu Distributions. It has been shown that probability density functions (Pdf), such as LL, are inadequate; hence, extended functions are used to more effectively model observed duration survival distributions more effectively. Results clearly show that the proposed extended Pdf, AR and DV, provide a viable alternative to other PDFs for describing the duration of diabetes survival time.

REFERENCES

1. Lusiana, V. (2019). Menkes sebut diabetes paling banyak ”serang” warga Riau ini sebabnya, 25 Maret 2019, https://riau.antaranews.com/berita/112025, accessed 8 juni 2019.
2. Juli, Kencing Manis dan Hipertensi Masuk 10 Pola Penyakit Utama di Bengkalis, 3 November 2019. http://infopublik.id/ kategori/ nusantara/ 384135, accessed 11 januari 2020.
3. Grover, G., & Sabharwal, A. (2012). A parametric approach to estimate survival time of diabetic nephropathy with left truncated and right censored data. *International Journal of Statistics and Probability*, *1*(1), 128-137.
4. Athifah, U., Yendra, R., Marizal, M., & Rahmadeni. (2021). Survival time data modeling for diabetics during the Covid 19 pandemic using several sensitive distributions (Weibull, gamma, and normal logs). Applied Mathematical Sciences,15(15), 717-724. doi: 10.12988/ams.2021.916579.
5. Usvita, M. L., Adnan, A., & Yendra, R. (2021). The Modelling Survival Times for Diabetes Patient Using Exponential, Weibull and Rayleigh-Lomax Distribution. *International Journal of Mathematics Trends and Technology-IJMTT*, *67*, 109-112.
6. Sutriana, Yendra, R., Rahmadeni, & Marizal, M. (2023). The Comparison Duration Diabetes Survival Times Modelling Using Lindley (LIN), Weighted Lindley Exponential (WLE), Power Modified Lindley (PML), Lindley Half-Cauchy (LHC) and Rayleight Lomax (RL) Distributions. *International Journal of Mathematics and Computer Research*, *11*(12), 3894-3898.
7. Kargarian-Marvasti, S., Rimaz, S., Abolghasemi, J., & Heydari, I. (2017). Comparing of Cox model and parametric models in analysis of effective factors on event time of neuropathy in patients with type 2 diabetes. *Journal of Research in Medical sciences*, *22*(1), 115.
8. Lindley, D. V. (1958). Fiducial distributions and Bayes’ theorem. *Journal of the Royal Statistical Society, Series B*, *20*(1), 102-107.
9. Shanker, R. (2016b) Sujatha distribution and Its Applications. To appear in “Statistics in Transition new Series”, *17*(3).
10. Shanker, R. (2016c). Amarendra Distribution and Its Applications. *American Journal of Mathematics and Statistics*, *6*(1), 44-56.
11. Shanker, R. (2016). Devya Distribution and Its Applications. *International Journal of Statistics and Applications*, *6*(4), 189-202. Doi: 10.5923/j.statistics.20160604.01
12. Shanker, R. (2016e). Shambhu distribution and its Applications, *International Journal of Probability and Statistics*, *5*(2), 48-63.