Parameter estimation and fitting of life time distributions in the presence of competing risks

Talawar A. S. *and* **Rangoli A. M.**
Departemnt of Statistics, Karnatak University, Dharwad. India
Email: talwarappu@gmail.com and ajmr1008@gmail.com

Abstract: In the present paper we consider the prostate cancer data to study the applications of life time parametric distributions such as exponential, Weibull and two forms of modified Weibull distributions in the presence of competing risks with three causes of failures. We found that all the four life time distributions give good fit to the prostate cancer data and in comparison, the modified Weibull-II distribution fits well. We have also calculated hazard for the causes and seen that exponential and modified Weibull-I distribution show constant hazard rate. Where as Weibull and modified Weibull-II distributions have increased hazard rate. We have compared the survival curves of Kaplan-Meier and all four distributions and seen that modified Weibull-II distribution survival curve and Kaplan-Meier survival curve coincides.

Key words: Modified Weibull distribution, Maximum likelihood estimator, competing risks, Kaplan-Meier estimator and information criterion.

1. Introduction:

Generally statistical analysis is different from the survival Analysis because of the presence of the censored observations. In the case of censoring, may be of right, left or interval, we cannot go with usual statistical analysis [9]. In survival analysis or medical studies it is quite common that more than one cause of failure may be directed to an object at the same time. In many situations we see that, subjects or patients can experience more than one cause of failure. Thus the concept of competing risks arises. Here the analysis of competing risks can be done through parametric and non-parametric approaches. Many authors like David and Moeschberger [4], Iyer et al. [5] and Alwasel [1] gave the explanation regarding the parametric life time distributions such as exponential, Weibull and modified forms of Weibull Distributions. The analysis of non-parametric version of this competing risks model can be done by Kaplan and Meier [6]. Generally Weibull distribution is preferred because it has increasing, decreasing and constant hazard rate. We can see in Alwasel [1] where they have considered exponential, Weibull and modified Weibull-II distribution and found that modified Weibull-II distribution fits well for the electronic appliances dataset. Sarhan and Zaindin have considered exponential, Rayleigh, linear failure rate, Weibull and modified Weibull distribution-I [10] for Aarset dataset and found that out of these distributions, modified Weibull-I distribution fits well. In present paper we are considering four life time parametric distributions for mortality analysis such as exponential, Weibull and modified forms of Weibull distribution ([10] and [1]). The selection of the distribution is based on the dataset. For this we are using the prostate cancer data **[2]** with three causes of failure such as death due to cancer, CVD and other causes, with right censored observations.

2. Research Methodology:

Let $T_1 T_2 \ldots . T_n$ be the independently distributed failure time of n patients out of which n_1 items failed from k different causes and rest $n_2 = n - n_1$ were right censored.

Let $\delta_{ij} = \begin{cases} 1 & i^{th} subject\ fail\ due\ to\ j^{th}\ cause,\ j = 1,2, \ldots \ldots, k \\ 0 & i^{th} subject\ does\ not\ fail\ due\ to\ j^{th} cause\ (Censored) \end{cases}$

Let $f(t)\ and F(t)$ be density and distribution functions of t, where $F(t) = \int_0^t f(u)du$

The basic quantity employed to describe time-to-event phenomena is the survival function, the probability of an individual surviving beyond time t (experiencing the event after time t) [6], is defined as

$S(t) = p(T > t)$ or $S(t) = 1 - F(t)$

2.1 Kaplan-Meier (K-M) Estimator:

The Kaplan-Meier estimator known as the product limit estimator is a non-parametric statistic used to estimate the survival function from lifetime data [7]. An important benefit of the Kaplan–Meier curve is that, the method can take into account some types of censored data, particularly right-censoring, which occurs if a patient withdraws from a study, or is lost to follow-up, or is alive without event incidence at last follow-up. The Kaplan-Meier estimate is an easiest way of computing survival over time. The Kaplan Meier estimator of survival function is defined as

$$\hat{S}(t) = \prod_{i:t_i<t}(1 - \frac{d_i}{n_i}) \quad (1)$$

Where t_i is the failure time, d_i is the number of events that occurs at time t_i and n_i is the number individuals at risk of experiencing the event immediately prior to t_i.

2.2 Hazard Function:

It is a instantaneous rate at which failures occurs for items that are surviving at time t [9], which is denoted by

$$\lambda(t) = \lim_{h \to 0} \frac{p(t \le T < t + h/T \ge t)}{h}$$

or Hazard function in terms of density and survival, $\lambda(t) = \frac{f(t)}{S(t)}$

Let $\Lambda(t) = \int_0^t \lambda(u)du$ be the cumulative hazard function and the survival function in terms of cumulative hazard function

$$S(t) = e^{-\Lambda(t)} = e^{-\int_0^t \lambda(u)du} \quad (2)$$

Now the cause specific hazard function can also be written as

$$\lambda_j(t) = \lim_{h \to 0} \frac{p(t \le T < t + h, J = j/T \ge t)}{h}, J = 1, 2, \ldots . k$$

$$\lambda_j(t) = \lim_{h \to 0} \frac{p(t \le T < t + h, J = j)}{hp(T > t)}, J = 1, 2, \ldots . k$$

And thus, we have $\lambda_j(t) = \frac{f_j(t)}{S(t)}$ (3)

Where $f_j(t)$ is the sub-density function and $F_j(t)$ is the sub-distribution function

Also $\lambda(t) = \sum_{j=1}^k \lambda_j(t)$ and $S(t) = e^{-\int_0^t \sum_{j=1}^k \lambda_j(u)du}$

2.3 Maximum Likelihood Estimator (MLE):

The method of estimation of unknown parameters proceeds as follows,

The likelihood function for the censored data is defined as [9]

$$L = \prod_{i=1}^n \prod_{j=1}^k (f_j(t))^{\delta_{ij}} S(t)^{1-\delta_{ij}}$$

$$L = \prod_{i=1}^n \prod_{j=1}^k (\lambda_j(t))^{\delta_{ij}} S(t) \quad (5)$$

Now the log likelihood can be written as,

$$l = logL = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\delta_{ij} * \log\left(\lambda_j(t)\right)\right) + \log(S(t)) \quad (6)$$

Now we can consider life time parametric distributions such as Exponential, Weibull, modified Weibull-I [10] and modified Weibull-II [1] Distributions for the data with three causes of failures.

2.4 Life time Distribution:

Now we consider life time parametric distributions such as Exponential, Weibull, modified Weibull-I ([10]) and modified Weibull-II ([1]) distributions for the prostate cancer data with three causes of failures.

2.4.1 Exponential Distribution:

Let $T \sim exp(\alpha)$ where T be the failure time

$f(t) = \alpha\, e^{-\alpha t}, \ \ x \geq 0 \quad \alpha > 0 \quad \alpha \rightarrow Scale \ \ Parameter.$

Therefore we have $\lambda(t) = \alpha$ and $S(t) = e^{-\alpha t}$

$$l_E = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\delta_{ij} * \log(\alpha_j)\right) + \log(e^{-\alpha_j t_i}) \quad (7)$$

$where\ l_E$ is the log-likelihood function of exponential distribution.

2.4.2 Weibull Distribution :

Let $T \sim Weibull(\alpha, \beta)$ and the density is

$f(t) = \alpha\, \beta\, t^{\beta-1} e^{-\alpha t^{\beta}}, \ x \geq 0 \quad \alpha, \beta > 0$

Where α *is the scale parameter* and β *is the shape parameter*.

Thus

$$\lambda(t) = \alpha\, \beta\, t^{\beta-1}$$
$$S(t) = e^{-\alpha t^{\beta}}$$

Here if shape parameter tends to 1 then Weibull distribution tends to exponential distribution.

$$l_E = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\delta_{ij} * \log(\alpha_j \beta_j t_i^{\beta_j - 1})\right) + \log\left(e^{-\alpha_j t_i^{\beta_j}}\right) \quad (8)$$

$where\ l_M$ is the log-likelihood function of Weibull distribution.

2.4.3 Modified Weibull-I Distribution(due to Sarhan and Zaindin [10]):

Let $T \sim Modified\ Weibull$ with parameters $a, \beta\ and\ \gamma$ and the density is

$f(t) = (a + \beta\, \gamma\, t^{\gamma})e^{-at-\beta t^{\gamma}}, \ x \geq 0 \ a, \beta, \gamma > 0$

Where both a *and* β *are the scale parameters and* γ *is the shape parameter*

Therefore, we have

$$\lambda(t) = a + \beta\, \gamma\, t^{\gamma}$$
$$S(t) = e^{-at-\beta t^{\gamma}}$$

$$l_{MW-I} = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\delta_{ij} * \log(a_j + \beta_j \gamma_j t_i^{\gamma_j})\right) + \log\left(e^{-a_j t_i - \beta_j t_i^{\gamma_j}}\right) \quad (9)$$

$where\ l_{MW-I}$ is the log-likelihood function of Modified Weibull [9] distribution.

Here if scale parameter $\beta = 0$ and shape parameter $\gamma = 1$ then modified Weibull distribution converges to Exponential distribution

2.4.4 Modified Weibull-II Distribution (due to Alwasel [1]):

Let $T \sim Modified\ Weibull$ with parameters $\alpha, \beta\ and\ \lambda$ and the density is

$$f(t) = \alpha\, (\beta + \lambda\, t)t^{b-1}e^{\lambda\, t}e^{-\alpha\ t^{\beta}e^{\lambda t}} \ , t \geq 0 \ \ \alpha, \beta > 0 \ , \lambda \geq 0$$

Where β is the shape parameter and, α and λ is the scale parameters

$$\lambda(t) = \alpha\, (\beta + \lambda\, t)t^{b-1}e^{\lambda\, t}$$
$$S(t) = e^{-\alpha\ t^{\beta}e^{\lambda t}}$$

$$l_{MW-II} = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\delta_{ij} * \log\left(\alpha_j(\beta_j + \lambda_j t_i) t_i^{b_j - 1} e^{\lambda_j t_i}\right)\right) + \log\left(e^{-\alpha_j t_i^{\beta_j} e^{\lambda_j t_i}}\right) \quad (10)$$

where l_{MW-II} is the log-likelihood function of Modified Weibull [1] distribution.

Generally we don't get explicit form to estimate the unknown parameters, so we consider numerical analysis using Newton-Raphason method.

2.5 Information Criterion

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) methods are used to know which of the following distributions fits the data well [9]. The distribution fits the data well, whose AIC and BIC values are less.

$$AIC = 2K - 2\ lnL$$

$$BIC = K\ ln(n) - 2\ lnL$$

Where L is the likelihood function, n is the sample size and K is the number of parameters estimated

2.6 Goodness of Fit:

The testing of goodness of fit and the calculation of confidence intervals can be done by Liklihood ratio test. The liklihood ratio test can be done as [9],

$$\Lambda = \frac{L(\theta_{H_0})}{L(\theta_{H_1})}$$

Where θ is a vector of parameters

Under null hypothesis, $X_L = -2\ln\Lambda = -2\ (l_{H_0} - l_{H_1})$

Where

$X_L \sim$ Chisquare distribution with k degrees offreedom, k is number of parameters

2.6.1 Exponential with Weibull Distribution :

To test suitablility of the model, the null and alternative hypotheses can be stated as

$$H_0{:}\ \beta = 1 \quad v/s H_1{:}\ \beta \neq 1$$

or equivalently

$$H_0{:}\ Exponential\ distribution\ \ v/s\ \ H_1{:}\ Weibull\ distribution$$

Under null hypothesis $X_L = -2\ln\Lambda = -2\ (l_E - l_W)$

Where l_E →logliklihood of Exponential Distribution

l_W →logliklihood of Weibull distribution

2.6.2 Weibull with modified Weibull-II Distribution :

Similarly, to test Weibull versus modified Weibull-II, we state null hypothesis and alternative hypothesis as

$$H_0{:}\ \lambda = 0 \quad v/s H_1{:}\ \lambda \neq 0$$

$$H_0{:}\ Weibull\ distribution\ \ v/s\ \ H_1{:}\ Modified\ Weibull - II\ distribution$$

Thus under H_0, $X_L = -2\ln\Lambda = -2\ (l_W - l_{MW-II})$

Where l_W →logliklihood of Weibull distribution

l_{MW-II} →logliklihood of modified Weibull-II distribution

Here l_E, l_W and $l_{MW\text{-}II}$ are the corresponding log liklihood functions of Exponential, Weibull, modified Weibull-I and modified Weibull-II distributions after replacing the unknown parameters value by their respective MLE's.

2.7 Asymptotic Confidence bounds:

The MLE's do not have closed form, to know the distribution and to calculate confidence intervals, we use asymptotic distribution of the MLE of the parameters. It is known that the asymptotic distribution of the MLE $\hat{\theta}$ is given by

$$(\hat{\theta} - \theta) \rightarrow N_9(0, I^{-1}(\theta))$$

Where $I^{-1}(\theta) \rightarrow$ Fisher Information Matrix of the unknown Parameters
$\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \lambda_1, \lambda_2, \lambda_3)$ parameters for modified Weibull-II distribution.
The elements of the $9X9$ matrix I^{-1}, $I_{ij}(\theta)$, $i, j = 1,2, \dots 9$ can be approximated by $I_{ij}(\hat{\theta})$, where

$$I_{ij}(\hat{\theta}) = -\left.\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}\right|_{\theta = \hat{\theta}} \quad (11)$$

$\hat{\theta} = (\widehat{\alpha_1}, \widehat{\alpha_2}, \widehat{\alpha_3}, \widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}, \widehat{\lambda_1}, \widehat{\lambda_2}, \widehat{\lambda_3})$ estimated parameters.

Therefore, the approximation $100(1-\gamma)\%$ two sided confidence interval for θ is given by

$$\hat{\theta} \pm Z_{\gamma/2}\sqrt{I^{-1}(\hat{\theta})} \quad (12)$$

Here $Z_{\gamma/2}$is the upper $\gamma/2$ th percentile of a standard normal distribution.

3 Results and Discussion:

For the validity purpose, we have considered prostate cancer data set with 316 observations [2]. Data having three causes of failure i,e,. death due cancer, CVD and other causes with right censored observation. Table I gives the summary of the data. Table 2 gives the estimated values of parameters of the four distributions viz, Exponential, Weibull, modified Weibull-I and modified Weibull-II for three causes of failures. The parameter of these distributions are estimated using MLE. Here the modified Weibull-I distribution converges to exponential distribution as the estimated shape parameter is $\gamma \approx 1$ and scale parameter $\beta \approx 0$. From Table 3, we can see the AIC and BIC values of these distributions. We conclude that modified Weibull-II distribution has less AIC and BIC values as compared to others. Hence modified Weibull-II distribution fits well to this dataset.

On the basis of the likelihood ratio test, for testing exponential versus Weibull distribution, the calculated vaue of X_L =52, gives (χ^2Table value at 3 df is 7.81) Weibull distribution fits well as comapred to exponential distribution. In case of Weibull with modified Weibull-II, from the value X_L=699.618(χ^2Table value at 3 df is 7.81) , we conclude that modified Weibull-II distribution fits well to this dataset.

Table 4 contains standard error, lower and upper confidence limits of the parameters of the modified Weibull-II distribution for three causes of failures of the good fit model.

Figure 1 explains about histogram of the data and fitting of the four distributions, and we can see that modified Weibull-II distribution (red line) fits well to this dataset.

Figure 2 explains about the hazard for cause 1 in that we can see that Exponential and modified Weibull-I distribution shows constant hazard rate as modified Weibull-I distribution converges to Exponential distribution, where as Weibull and modified Weibull-II distributions have increasing hazard rate. Similarly Figure 3 and figure 4 explains about hazard curve for cause 2 and hazard curve for cause 3. Figure 5 explains about the survival curve of the patients having cause 1, here we have considered Kaplan Meier survival curve with all four distributions survival curves, and can be seen that Kaplan-Meier (blackline) and modified Weibull-II (redline) distribution have same survival curve. Similaly Figure 6 and Figure 7 explains about the survial curve for cause 2 and cause 3 respectively.

Table I : Description of the data.

	Cause I	Cause II	Cause III	Censored
No of observation	110	86	107	13
Mean	21.70909	22.05814	23.01869	-
SD	12.66748	14.63867	14.47573	-

Table2: The estimated parameters of the four distributions for three causes using MLE.

Distribution		α	β	γ	λ	a
Exponential	Cause 1	0.0148448	-	-	-	-
	Cause 2	0.01160594	-	-	-	-
	Cause 3	0.01443995	-	-	-	-
Weibull	Cause 1	0.002379926	1.532118704	-	-	-
	Cause 2	0.003616316	1.341140959	-	-	-
	Cause 3	0.003003756	1.457522202	-	-	-
Modified Weibull-I	Cause 1	-	$1.652093*10^{-10}$	0.998746	-	0.0148448
	Cause 2	-	$1.586036*10^{-8}$	0.999999	-	0.01160592
	Cause 3	-	$1.442505*10^{-8}$	0.99999999	-	0.01443993
Modified Weibull-II	Cause 1	0.002622533	1.482020794	-	0.002189726	-
	Cause 2	0.009103926	0.792041241	-	0.027896175	-
	Cause 3	0.009023049	0.834223757	-	0.030100299	-

Table 3: The values of the AIC and BIC for four distributions

Distribution	Logl	AIC	BIC
Exponential	-1602.788	3211.576	3222.843
Weibull	-1576.778	3165.556	3188.09
Modified Weibull-I	-1602.788	3223.576	3257.378
Modified Weibull-II	-1226.969	2471.938	2505.74

Table 4: Estimated parameters, Standard Error (S.E), Lower Confidence Interval (LCI) and Upper Confidence Interval (UCI) of the Parameters for three causes of failures for the modified Weibull-II distribution.

Parameter	S.E	LCI	UCI
$\alpha_1 = 0.002622533$	0.001655443	-0.0006055799	0.005850646
$\alpha_2 = 1.482020794$	0.266847865	0.9616674567	2.002374131
$\alpha_3 = 0.002189726$	0.010468486	-0.0182238218	0.022603274
$\beta_1 = 0.009103926$	0.004170795	0.0009708760	0.017236976
$\beta_2 = 0.792041241$	0.189084295	0.4233268653	1.160755617
$\beta_3 = 0.027896175$	0.009067516	0.0102145184	0.045577832
$\lambda_1 = 0.009023049$	0.004107799	0.0010128416	0.017033256
$\lambda_2 = 0.834223757$	0.189219226	0.4652462662	1.203201248
$\lambda_3 = 0.030100299$	0.008725024	0.0130865015	0.047114096

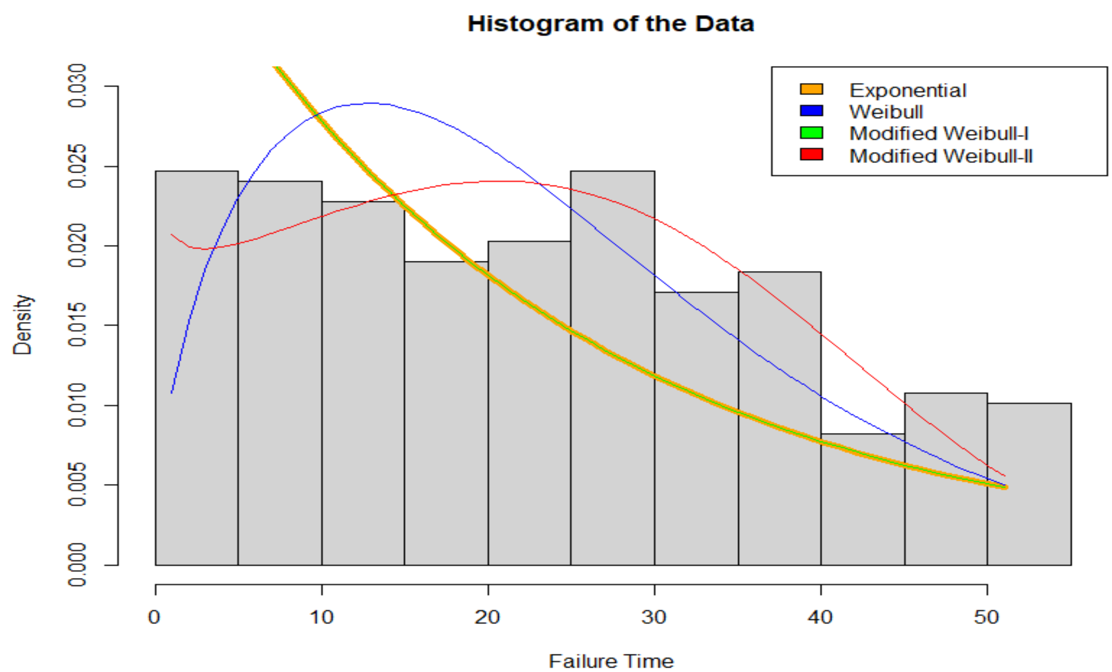

Figure 1: Histogram of Cancer data with fitted four Distributions.

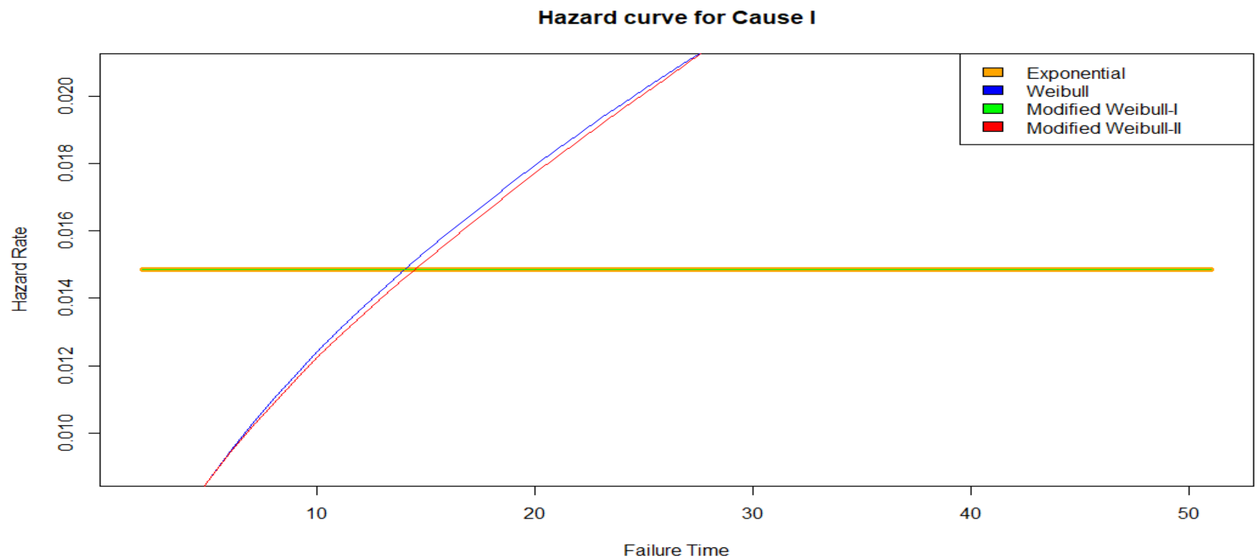

Figure 2: Hazard Curve for cause I with four distributions.

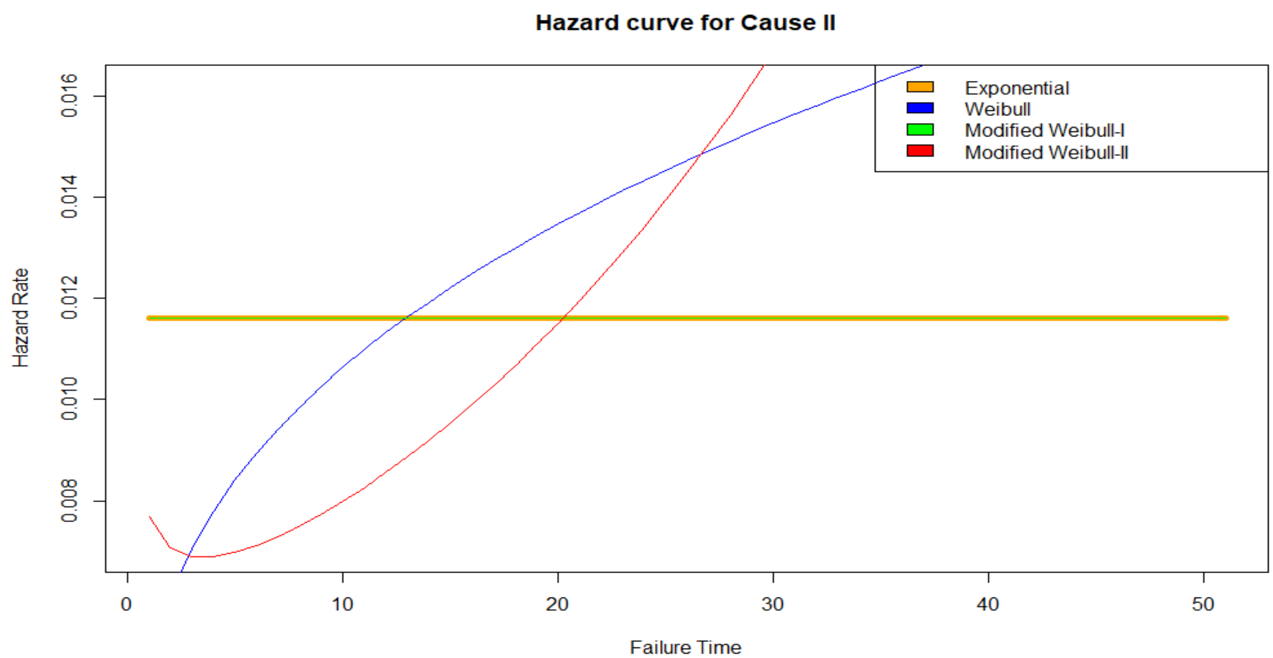

Figure 3: Hazard Curve for cause II with four distributions

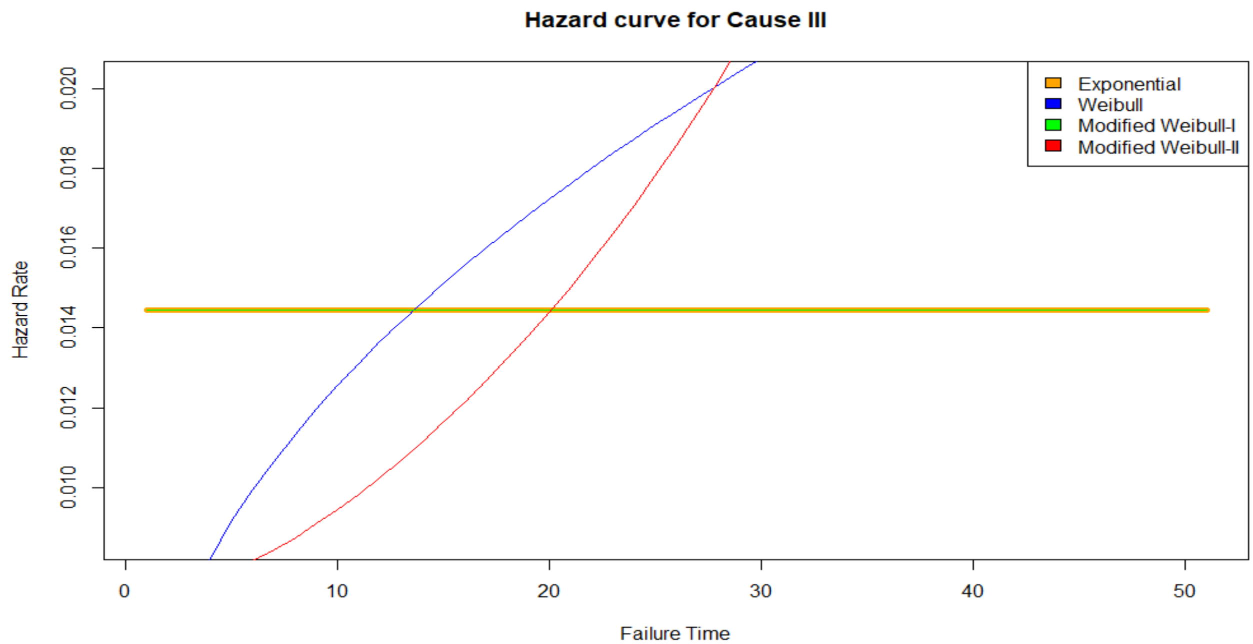

Figure 4: Hazard Curve for cause III with four distributions

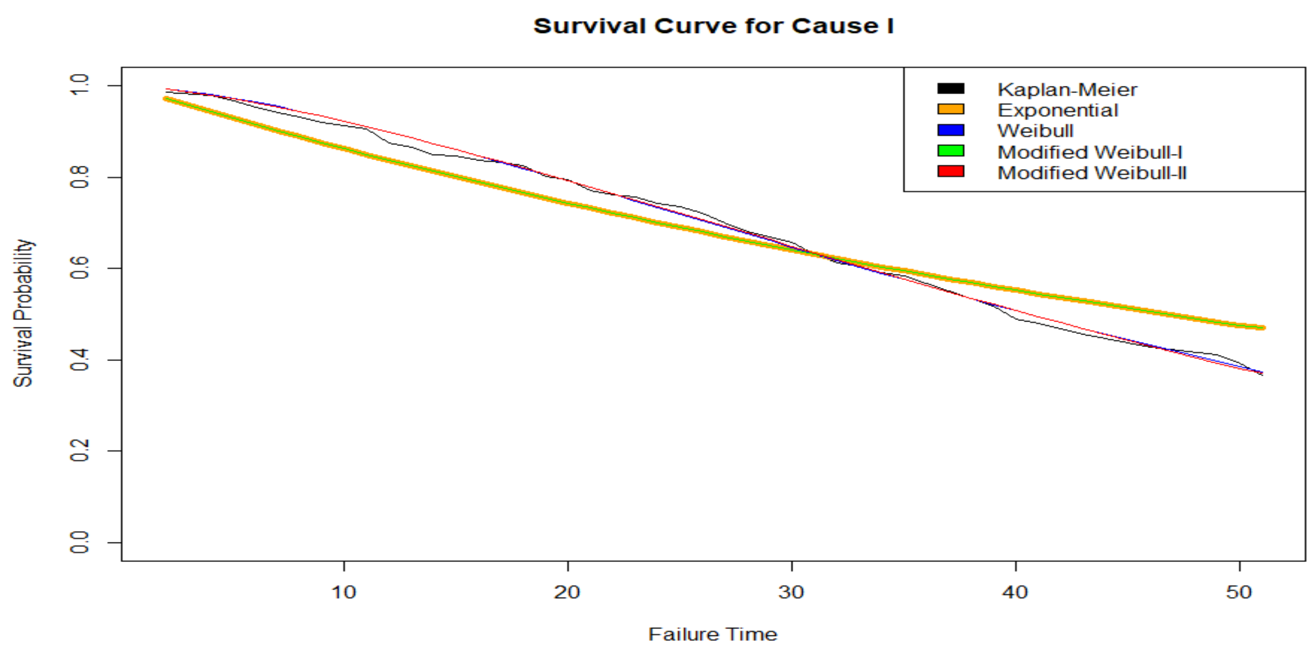

Figure 5: Survival Curve for cause I with Kaplan-Meier and four distributions

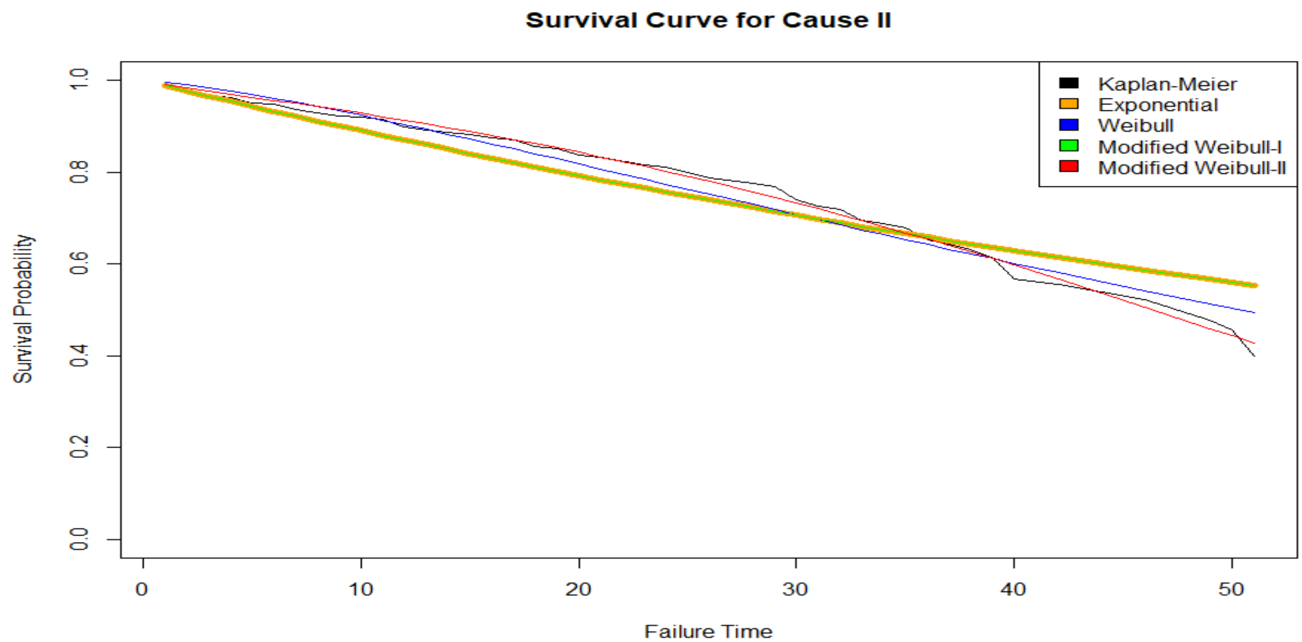

Figure 6: Survival Curve for cause II with Kaplan-Meier and four distributions

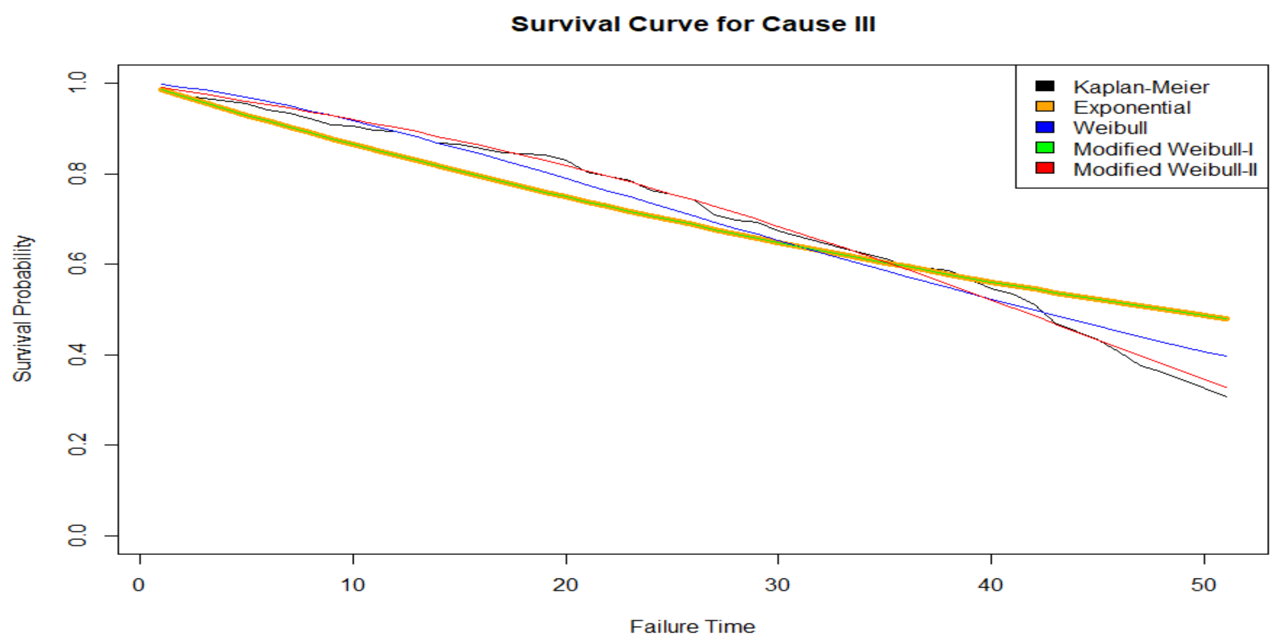

Figure 7: Survival Curve for cause III with Kaplan-Meier and four distributions

Conclusion: We conclude that, out of the four life time parametric distributions of mortality, the modified Weibull-II distribution fits well to this dataset. We can see that modified Weibull-I distribution converges to exponential distribution and hence have constant hazard rate for all three causes of failure. And as we have compared Kaplan-Meier survival curve with these four distributions survival curves and seen that modified Weibull-II distribution and Kaplan-Meier coincides.

Acknowledgement: The second author is thankful to Department of Science and Technology, innovation in science pursuit for inspired research (DST-INSPIRE) for financial support.

References:

1. Alwasel, I. A.(2009). Statistical Infernce of a competing Risks Model with Modified Weibull Distribution. *International Journal of Math. Analysis*, Vol. 3, no, 905-918.
2. Andrews, D. F. and Herzberg, A. M. (2012). *Data: a collection of problems from many fields for the student and research worker*. Springer Science & Business Media.
3. Crowder, M. J. (2001).*Classical Competing Risks.* Chapman and Hall/CRC, P.N: 09.
4. David, H. A., and Moeschberger, M. L.(1978). *The Theory of competing Risks*. Griffin, London.
5. Iyer, S.K., Jammalamadaka, S. R., and Kundu, D. (2008). Analysis of middle-censored data with exponential lifetime distributions. *Journal of Statistical Planning and Inference*, 138 3550 – 3560.
6. Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.
7. Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observation. *Journal of American Statistical Association*, 53, 457-481.
8. Lai, C. D., Xie, M., and Murthy, D.N.(2003). A Modified Weibull Distribution. *IEEE Trans. Rel.*, Vol. 52, No. 1, PP. 33-37.
9. Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York.
10. Sarhan, A.M., and Zaindin, M. (2009). Modified Weibull distribution .*Applied Sciences,* Vol.11, 2009, pp. 123-136.