Jurnal
Teknologi

Full Paper

HYPOTHESIS TESTING FOR THE PARAMETERS OF LOG-LOGISTIC REGRESSION MODEL WITH LEFT-TRUNCATED AND RIGHT-CENSORED SURVIVAL DATA

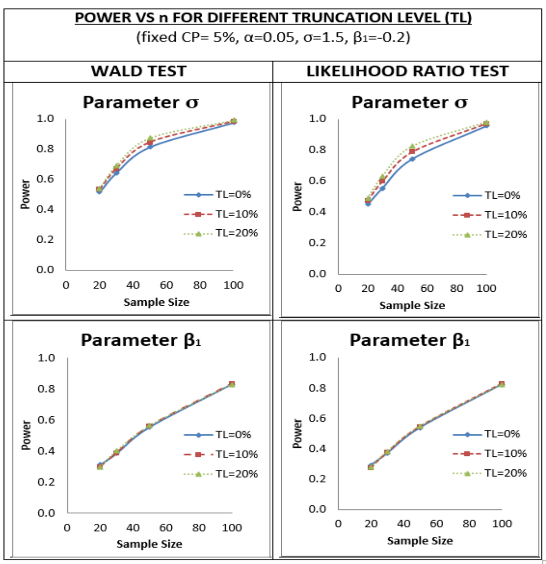
Wan Nur Atikah Wan Mohd Adnan*, Jayanthi Arasan

Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

Article history
Received
13 August 2017
Received in revised form
26 November 2017
Accepted
15 January 2018
Published online
1 April 2018

*Corresponding author
wanatikah@upm.edu.my

Graphical abstract



Abstract

Left-truncation and right-censoring (LTRC) arise naturally in lifetime data. Data may be left-truncated due to a limitation in the study design. Failure of a unit is observed only if it fails after a certain period. Usually, the units under study may not be followed until all of them have failed but the study has to be stopped at a certain time. This introduces the right censoring into the survival data. Log-logistic model is extended to accommodate the left-truncated and right-censored survival data. The bias, standard error (SE), and root mean square error (RMSE) of the parameter estimates are computed to evaluate the performance of the model at different sample sizes, censoring proportion (CP), and truncation level (TL). The results show that the SE of the parameter estimates increase as the truncation level (TL) and censoring proportion (CP) increase. Having low and high TL (5% and 15%) in the data, the graphs clearly show that the empirical power of both tests increases with the increase of TL for parameter σ and β_1. The SE and RMSE also decrease as the sample size increases. Following that, power analysis is conducted via simulation to compare the performance of hypothesis tests based on the Wald and Likelihood Ratio (LR) for the parameters. The results clearly indicate that the Wald performs slightly better than the LR when dealing with the proposed model.

Keywords: Log-logistic Regression Model, Left-Truncated and Right-Censored, Wald, Likelihood Ratio, Empirical Power

Abstrak

Pemangkasan kiri dan penapisan kanan (LTRC) timbul secara semula jadi dalam data seumur hidup. Data mungkin dipisahkan oleh sebab had dalam reka bentuk sesuatu kajian. Dalam kes ini, kegagalan unit yang diperhatikan adalah jika ia gagal selepas tempoh masa tertentu. Biasanya, unit di bawah kajian tidak boleh diikuti sehingga gagal kerana kajian itu perlu dihentikan pada masa yang tertentu. Oleh itu, data yang mengandungi tapisan kanan diperkenalkan. Model Log-logistik diperluaskan untuk menampung data mandirian yang dipangkas kiri dan ditapis kanan. Ralat, ralat piawai (RP), dan punca min ralat persegi (PMRP) daripada anggaran parameter model dikira untuk menilai prestasi model pada saiz sampel yang berbeza, bahagian tapisan kanan yang berbeza dan tahap pemangkasan kiri yang berbeza. Keputusan menunjukkan bahawa ralat piawai meningkat apabila kadar tangkasan kiri dan kadar tapisan kanan meningkat. Mempunyai TL rendah (5% dan 15%) dalam data, graf jelas menunjukkan bahawa kuasa empirikal kedua-dua ujian meningkat dengan peningkatan TL untuk parameter σ dan β_1 dan SE dan RMSE juga berkurangan apabila saiz sampel bertambah. Maka, analisis kuasa telah dijalankan melalui kaedah simulasi untuk membandingkan prestasi ujian hipotesis berdasarkan Wald dan Nisbah Kemungkinan (LR) bagi parameter-parameter dalam model ini. Keputusan jelas menunjukkan bahawa Wald adalah lebih baik daripada LR apabila berurusan dengan model regresi yang dicadangkan.

Kata kunci: Model Regresi Log-logistik, Pemangkasan Kiri dan Tapisan Kanan, Nisbah Kemungkinan, Kuasa Empirik

1.0 INTRODUCTION

Log-logistic is one of the distributions that is widely used to model lifetime data especially in medical field. It is often used in survival analysis as a parametric model because it has a non-monotonic hazard rate; the hazard rate that increases to a maximum and later decreases. Bennet (1983) dealt with Log-logistic distribution (LLD) to analyze survival data due to its non-monotonic hazard function, and concluded that it can be a good replacement for the other distribution that has similar hazard function such as the Weibull or Log-normal distribution. Other authors who have done significant work using this model are Cox and Lewis (1966), Cox *et al.* (1982). The model can be easily extended to accommodate the covariates, truncated data and all types of censored observations such as left, right and interval. Arasan and Adam (2014) agreed that the log logistic model was chosen as its popularity in most cancer studies and its ability to fit both fixed and time dependent covariates easily. Lawless (1982) has more discussion on the truncated data in the survival field. The parameters of this model had unique maximum likelihood estimates and this has been proved by Gupta *et al*. (1999).

Generally, truncation and censoring reduce the information about the survival data. Censored data arise when an individual's life length is known to occur only within a certain period of time. Lifetime data are sometimes truncated due to the study design. Ignoring left truncation may lead to inadequacy of statistical procedures as described by Luo and Tsai (2009). Emura and Wang (2012) agreed that the truncated data have some information about the lifetime therefore it is important to include it in the analysis of survival. The observations may enter the study at random time or delayed entry as the starting time might be random for each individuals (Shen (2009)).

Empirical likelihood ratio method was first used by Thomas and Grunkemeier (1975). They proposed a nonparametric likelihood ratio method for interval estimation of survival probabilities for randomly censored data. Their method based on confidence intervals has better performance as compared to normal approximation method. However, the computation of the empirical likelihood ratios with censored or truncated data is non-trivial. Turnbull (1976) used a modified self-consistency/ EM algorithm to compute a class of arbitrarily censored or truncated empirical likelihood ratios. Li (1995) discussed more on empirical likelihood of truncated data.

Arasan and Manoharan (2013) focused on assessing the performance of the log-normal distribution with left truncated and right censored (LTRC) survival data. Some of the observations on the left-tailed of the log-normal distribution will be disregarded consequently with the increase of the existing skewness of the distribution when fitted to LTRC data as demonstrated by Cain *et al.*, 2011.

The LR method has been shown to work with all types of censored survival data. This has been highlighted by authors Pan and Zhou (1999) and Murphy and Van der Vaart (1997). Pan and Zhou showed for right-censored data, the empirical LR has a chi-square limit (Wilks Theorem). Doganaksoy and Schmee (1993) showed that the likelihood ratio method performed better with parameters of the log-normal distribution compared to the Wald method. For skewed data, the assumption of normality often fails as it can't fully capture the sampling distribution of the sample statistics being studied, subsequently resulting in the poor performance of the Wald and LR methods in constructing the confidence intervals for the parameter estimates specifically when higher proportion of truncation and censoring is present in the data (Manoharan, Arasan, Midi & Adam (2015)).

None of the researchers had done the performance analysis for the LLD with LTRC data. On this basis, the simulation study is conducted to access the performance of proposed model with LTRC data using hypothesis testing based on the Wald and LR tests.

2.0 METHODOLOGY

2.1 Log-logistics Regression Model with Right-Censored and Left-Truncated Survival Data

Let $T \geq 0$ be the random variable representing survival time or lifetime with density function

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}. \quad (1)$$

The probability density function (PDF) of LLD is given by

$$f(t) = \frac{\lambda\sigma(\lambda t)^{\sigma-1}}{(1+(\lambda t)^{\sigma})^2}, t > 0,\ \lambda > 0,\ \sigma > 0. \quad (2)$$

LLD has 2 parameters which are λ and σ represent scale and shape respectively.

In this paper, the LLD is extended to accommodate the covariate effects by allowing $\lambda = e^{\beta' X}$ where $X' = (x_0, x_1, \ldots, x_p)$ is the vector of covariate values, $x_0 = 1$ and $\beta' = (\beta_0, \beta_1, \ldots, \beta_p)$ are unknown parameters where p denotes the number of covariates in the model.

Suppose we have n independent random variables $i = 1, 2, \ldots, n$. If $\lambda = e^{-\beta_0 - \beta_1 x_i}$, where x_i is a single fixed covariate for i^{th} subjects, then the density and survival function of LLD are given in (1) and (2) respectively as follows,

$$f(y_i, x_i, \beta, \sigma) = \frac{\sigma e^{(-\beta_0 - \beta_1 x_i)}\left(e^{(-\beta_0 - \beta_1 x_i)} e^{y_i}\right)^{\sigma-1}}{\left(1 + \left(e^{(-\beta_0 - \beta_1 x_i)} e^{y_i}\right)^{\sigma}\right)^2}, \quad (3)$$

$$S(y_i, x_i, \beta, \sigma) = \frac{1}{1 + \left(e^{(-\beta_0 - \beta_1 x_i)} e^{y_i}\right)^{\sigma}}. \quad (4)$$

If the censoring indicator, c_i is defined as

$$c_i = \begin{cases} 1 \text{ if the observation is uncensored} \\ 0 \text{ if the observation is right censored} \end{cases}$$

and t_i is the observed survival time for the i^{th} subjects, then the likelihood function consisting right-censored and left-truncated is defined as follows:

$$L(\hat{\theta}) = \prod_{i=1}^{n} \left(\frac{f(t_i)}{S(u_i)}\right)^{c_i} \left(\frac{S(t_i)}{S(U_i)}\right)^{1-c_i}$$

$$= \prod_{i=1}^{n} \left(\frac{\sigma e^{z_i \sigma}\left(1 + e^{z_{ui}\sigma}\right)}{1 + e^{z_i \sigma}}\right)^{c_i} \left(\frac{1 + e^{z_{ui}\sigma}}{1 + e^{z_i \varsigma}}\right)^{1-c_i} \quad (5)$$

where $z_i = y_i - (\beta_0 + \beta_1 x_i)$, y_i is the failure times and $z_{ui} = y_{ui} - (\beta_0 + \beta_1 x_i)$, y_{ui} is the left-truncation times. x_i is the covariates. Following that, the log-likelihood function of LLD for right-censored and left-truncated is obtained as follows:

$$\ell(\hat{\theta}) = \sum_{i=1}^{n} \{c_i \ln\sigma + c_i z_i \sigma - c_i \ln\left(1 + e^{z_i \sigma}\right) - \ln\left(1 + e^{z_i \sigma}\right) + \ln\left(1 + e^{z_{ui}\sigma}\right)\}. \quad (6)$$

2.2 Likelihood Equations and Estimation

The first and second order partial derivatives of the log-likelihood function with respect to parameters σ, β_0 and β_1 would be as follows:

$$\frac{\partial \ell}{\partial \sigma} = \sum_{i=1}^{n} \left(\frac{c_i}{\sigma} + c_i z_i - \frac{c_i z_i e^{z_i \sigma}}{1 + e^{z_i \sigma}} - \frac{z_i e^{z_i \sigma}}{1 + e^{z_i \sigma}} + \frac{z_{ui} e^{z_{ui}\sigma}}{1 + e^{z_{ui}\sigma}}\right) \quad (7)$$

$$\frac{\partial \ell}{\partial \beta_0} = \sum_{i=1}^{n} \left(-c_i \sigma + \frac{c_i \sigma e^{z_i \sigma}}{1 + e^{z_i \sigma}} - \frac{\sigma e^{z_i \sigma}}{1 + e^{z_i \sigma}} + \frac{\sigma e^{z_{ui}\sigma}}{1 + e^{z_{ui}\sigma}}\right) \quad (8)$$

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{n} \left(-c_i \sigma x_i + \frac{c_i \sigma x_i e^{z_i \sigma}}{1 + e^{z_i \sigma}} - \frac{\sigma x_i e^{z_i \sigma}}{1 + e^{z_i \sigma}} + \frac{\sigma x_i e^{z_{ui}\sigma}}{1 + e^{z_{ui}\sigma}}\right) \quad (9)$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \sum_{i=1}^{n} \left(\begin{aligned} &\frac{-c_i}{\sigma^2} - \frac{c_i z_i^2 e^{z_i \sigma}}{1 + e^{z_i \sigma}} + \frac{c_i z_i^2 \left(e^{z_i \sigma}\right)^2}{\left(1 + e^{z_i \sigma}\right)^2} - \frac{z_i^2 e^{z_i \sigma}}{1 + e^{z_i \sigma}} \\ &+ \frac{z_i^2 \left(e^{z_i \sigma}\right)^2}{\left(1 + e^{z_i \sigma}\right)^2} + \frac{z_{ui}^2 e^{z_{ui}\sigma}}{1 + e^{z_{ui}\sigma}} - \frac{z_{ui}^2 \left(e^{z_{ui}\sigma}\right)^2}{\left(1 + e^{z_{ui}\sigma}\right)^2} \end{aligned}\right) \quad (10)$$

$$\frac{\partial^2 \ell}{\partial \beta_0^2} = \sum_{i=1}^{n} \left(\begin{aligned} &\frac{-c_i \sigma^2 e^{z_i \sigma}}{1 + e^{z_i \sigma}} + \frac{c_i \sigma^2 e^{z_i \sigma}}{1 + e^{z_i \sigma}} - \frac{\sigma^2 e^{z_i \sigma}}{1 + e^{z_i \sigma}} \\ &+ \frac{\sigma^2 \left(e^{z_i \sigma}\right)^2}{\left(1 + e^{z_i \sigma}\right)^2} + \frac{\sigma^2 e^{z_{ui}\sigma}}{1 + e^{z_{ui}\sigma}} - \frac{\sigma^2 \left(e^{z_{ui}\sigma}\right)^2}{\left(1 + e^{z_{ui}\sigma}\right)^2} \end{aligned}\right) \quad (11)$$

$$\frac{\partial^2 \ell}{\partial \beta_1^2} = \sum_{i=1}^{n} \left(\begin{aligned} &\frac{-c_i x_i^2 \sigma^2 e^{z_i \sigma}}{1 + e^{z_i \sigma}} + \frac{c_i x_i^2 \sigma^2 \left(e^{z_i \sigma}\right)^2}{\left(1 + e^{z_i \sigma}\right)^2} - \frac{x_i^2 \sigma^2 e^{z_i \sigma}}{1 + e^{z_i \sigma}} \\ &+ \frac{x_i^2 \sigma^2 \left(e^{z_i \sigma}\right)^2}{\left(1 + e^{z_i \sigma}\right)^2} + \frac{x_i^2 \sigma^2 e^{z_{ui}\sigma}}{1 + e^{z_{ui}\sigma}} - \frac{x_i^2 \sigma^2 \left(e^{z_{ui}\sigma}\right)^2}{\left(1 + e^{z_{ui}\sigma}\right)^2} \end{aligned}\right) \quad (12)$$

$$\frac{\partial^2 \ell}{\partial\sigma\partial\beta_0}=\sum_{i=1}^{n}\left(\begin{array}{l} -c_i+\dfrac{c_i e^{z_i\sigma}}{1+e^{z_i\sigma}}+\dfrac{c_i z_i\sigma e^{z_i\sigma}}{1+e^{z_i\sigma}}-\dfrac{c_i z_i\sigma\left(e^{z_i\sigma}\right)^2}{\left(1+e^{z_i\sigma}\right)^2} \\ +\dfrac{e^{z_i\sigma}}{1+e^{z_i\sigma}}+\dfrac{z_i\sigma e^{z_i\sigma}}{1+e^{z_i\sigma}}-\dfrac{z_i\sigma\left(e^{z_i\sigma}\right)^2}{\left(1+e^{z_i\sigma}\right)^2}-\dfrac{e^{z_{Ui}\sigma}}{1+e^{z_{Ui}\sigma}} \\ -\dfrac{z_{Ui}\sigma e^{z_{Ui}\sigma}}{1+e^{z_{Ui}\sigma}}+\dfrac{z_{Ui}\sigma\left(e^{z_{Ui}\sigma}\right)^2}{\left(1+e^{z_{Ui}\sigma}\right)^2} \end{array}\right) \quad (13)$$

$$\frac{\partial^2 \ell}{\partial\sigma\partial\beta_1}=\sum_{i=1}^{n}\left(\begin{array}{l} -c_i x_i+\dfrac{c_i x_i e^{z_i\sigma}}{1+e^{z_i\sigma}}+\dfrac{c_i x_i z_i\sigma e^{z_i\sigma}}{1+e^{z_i\sigma}}-\dfrac{c_i x_i z_i\sigma\left(e^{z_i\sigma}\right)^2}{\left(1+e^{z_i\sigma}\right)^2} \\ +\dfrac{x_i e^{z_i\sigma}}{1+e^{z_i\sigma}}+\dfrac{x_i z_i\sigma e^{z_i\sigma}}{1+e^{z_i\sigma}}-\dfrac{x_i z_i\sigma e^{z_i\sigma}}{1+e^{z_i\sigma}}-\dfrac{x_i z_i\sigma\left(e^{z_i\sigma}\right)^2}{\left(1+e^{z_i\sigma}\right)^2} \\ -\dfrac{x_i e^{z_{Ui}\sigma}}{1+e^{z_{Ui}\sigma}}-\dfrac{x_i z_{Ui}\sigma e^{z_{Ui}\sigma}}{1+e^{z_{Ui}\sigma}}+\dfrac{x_i z_{Ui}\sigma\left(e^{z_{Ui}\sigma}\right)^2}{\left(1+e^{z_{Ui}\sigma}\right)^2} \end{array}\right) \quad (14)$$

$$\frac{\partial^2 \ell}{\partial\beta_0\partial\beta_1}=\sum_{i=1}^{n}\left(\begin{array}{l} -\dfrac{c_i x_i\sigma^2 e^{z_i\sigma}}{1+e^{z_i\sigma}}+\dfrac{c_i x_i\sigma^2\left(e^{z_i\sigma}\right)^2}{\left(1+e^{z_i\sigma}\right)^2}-\dfrac{x_i\sigma^2 e^{z_i\sigma}}{1+e^{z_i\sigma}} \\ +\dfrac{x_i\sigma^2\left(e^{z_i\sigma}\right)^2}{\left(1+e^{z_i\sigma}\right)^2}+\dfrac{x_i\sigma^2 e^{z_{Ui}\sigma}}{1+e^{z_{Ui}\sigma}}-\dfrac{x_i\sigma^2\left(e^{z_{Ui}\sigma}\right)^2}{\left(1+e^{z_{Ui}\sigma}\right)^2} \end{array}\right) \quad (15)$$

The inverse of the observed information matrix, which can be derived from the second partial derivatives of the log-likelihood function evaluated at $\hat{\sigma},\hat{\beta}_0,\hat{\beta}_1$, provides us with the estimates for the variance and covariance of $\hat{\theta}$.

$$\hat{\mathrm{var}}\left(\hat{\theta}\right)=\left[i\left(\hat{\sigma},\hat{\beta}_0,\hat{\beta}_1\right)\right]^{-1} \quad (16)$$

The Maximum Likelihood Estimation (MLE) of the parameters can be obtained by using the Newton-Raphson iterative procedure.

2.3 Inverse Transform Method

Inverse transform method is one of the methods used to simulate the lifetimes, t_i of LLD. By setting the survival function $S(t_i)=U_i$, where $U_i \sim U(0,1)$, the random numbers generated are converted from uniform distribution to random variables which are log-logically distributed.

$$S(t_i)=\frac{1}{1+(t_i e^{-(\beta_0+\beta_1 x_i)})^{\sigma}}$$

$$\frac{1}{U_i}-1=(t_i e^{-(\beta_0+\beta_1 x_i)})^{\sigma}$$

$$t_i=\frac{\left(\dfrac{1}{U_i}-1\right)^{\frac{1}{\sigma}}}{e^{-(\beta_0+\beta_1 x_i)}} \quad (17)$$

2.4 Hypothesis Testing Procedures

2.4.1 *Wald Test*

Let $\hat{\theta}$ be the maximum likelihood estimator for parameter and the log likelihood function of θ. Under mild regularity conditions, $\hat{\theta}$ is asymptotically normally distributed with mean θ and covariance matrix $i^{-1}(\theta)$, where $i(\theta)$ is the Fisher information matrix, evaluated at the true value of the parameter θ. The matrix $i(\theta)$ is not always available and can be replaced by the observed information matrix, $i\left(\hat{\theta}\right)$ which can be obtained from the log likelihood function evaluated at $\hat{\theta}$ as given below:

$$i\left(\hat{\theta}\right)=-\frac{\partial^2\ell\left(\hat{\theta}\right)}{\partial\hat{\theta}^2} \quad (18)$$

The Wald test statistics for testing the null hypothesis, $H_0:\theta=\theta_0$ versus the alternative hypothesis, $H_1:\theta\neq\theta_0$ is given as:

$$z=\frac{\hat{\theta}-\theta_0}{\sqrt{i^{-1}\left(\hat{\theta}\right)}} \quad (19)$$

where $i^{-1}\left(\hat{\theta}\right)$ is the $\hat{s.e}\left(\hat{\theta}\right)$ which can be obtain from the diagonal elements of the inverse of the observed Fisher information matrix or known as negative Hessian. Following that, the H_0 will be rejected if the test statistics, $|z|>z_{\alpha/2}$ where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution.

2.4.2 *Likelihood Ratio Test*

The LR compares the maximized likelihood of two nested models, the full model and the reduced model. The reduced model is restricted by certain conditions given in H_0. The LR testing the null hypothesis, $H_0:\theta=\theta_0$ versus alternative hypothesis, $H_1:\theta\neq\theta_0$ is given as follows:

$$\psi=-2\left(\ell\left(\theta_0,\tilde{\eta}\right)-\ell\left(\hat{\theta},\tilde{\eta}\right)\right) \quad (20)$$

where ℓ is the log likelihood function, η is the vector of nuisance parameters, $\left(\hat{\theta},\hat{\eta}\right)$ is the maximum likelihood estimator of (θ,η) and $\tilde{\eta}$ is the restricted maximum likelihood estimator of η under H_0. In particular, the LR test under the H_0 has asymptotically a chi square distribution with k degree of freedom (dof). In the decision rule, the H_0 will be rejected if the test statistics, $\psi>\chi^2_{\alpha,k}$ where k is the total number of parameters in full model subtract the total number of parameters in restricted model.

3.0 RESULTS AND DISCUSSIONS

A simulation study on left-truncated and right-censored survival data (LTRC) survival data proposed by Mitra (2013) is adopted. The estimates obtained from lung cancer data by Tai et.al (2003) are used as the true parameters values for the simulation study which are $(\sigma,\beta_0,\beta_1)=(2.65252, 3.39643, -0.00683)$. The week of truncation of the study, namely y, is fixed throughout the simulation. The beginning time point of the study or the truncation time is 70 weeks.

A set of random number of weeks which basically represents the patient's weeks of lung cancer diagnosis is simulated with equal probability with replacement; for before, y_{bk} and after, y_{aj} the week of truncation, where $k=1,2,\ldots,n_1$ and $j=1,2,\ldots,n_2$. In the simulation study, the total observation is denoted by $N=n_1+n_2$ and the truncation level (TL) is fixed at 10% and 20%.

The lifetimes, t_i are generated by:

$$t_i=e^{\beta_0+\beta_1 x_i}\left(U_i^{-1}-1\right)^{1/\sigma} \quad (21)$$

The covariate values are simulated from the standard normal distribution, $x_i\sim N(0,1)$. The lifetime, t_i are added to y_{bk} and y_{aj} and if the resulting failure time less than y which is the truncation week, these observations are removed and replaced by a new simulated set of y_{bk}, y_{aj}, t_i, U_i and x_i. The truncation times y_{ui} are obtained as $y_{ui}=y-y_{bk}$. The censoring times, c_i are simulated from the exponential distribution with parameter $\lambda, c_i\sim\exp(\lambda)$, where the value of λ could be adjusted in order to yield the desired approximate censoring proportion, CP in the data. In this study, the two levels of approximate CP, 0.05 and 0.15 are chosen to represent low and high levels of CP, respectively.

The simulation study is conducted to assess the accuracy and efficiency of the parameter estimates using N =1000 samples of size n=20,30,50,100. The true parameters, (σ,β_0,β_1)= (2.65252, 3.39643, -0.00683) are used to obtain the Bias, Standard Error (SE) and Root Mean Square Error (RMSE) of the parameter estimates.

Power of a statistical test is the total number of possibilities that the test will reject the null hypothesis, H_0 when a specific alternative, H_1 is true. The hypothesis tests the parameter σ at $H_0:\sigma=1$ versus $H_1:\sigma>1$ and tests the parameter β_1 at $H_0:\beta_1=0$ versus $H_1:\beta_1\neq 0$. Data are simulated from Log-logistics regression model under H_0 and H_1 for several effect sizes to estimate the empirical power, $1-\hat{\beta}$. Effect size refers to the difference between the true value for parameters σ and β_1 which are 2.65252 and -0.00683 and the value specified in H_1. Therefore, three effect sizes value are used for parameters σ and β_1 which are (1.5,2.0,2.5) and (-0.1,-0.2,-0.5) respectively.

The simulated data are adjusted to meet the desired truncation levels, TL = 0%, 10% and 20% and censoring proportions, CP = 5% and 15%. Then, the Wald and LR tested for parameters σ and β_1 are run for several effect sizes, n, TL and CP . The empirical power for the Wald and LR are estimated by the proportion of rejections for a test statistic at two values of significance levels which are $\alpha=0.10$ and $\alpha=0.05$. The proportions of rejection are calculated by counting the number of times the hypothesis tests reject the null hypothesis, H_0 and the number of rejections are divided with N =5000.

Tables 1, 2 and 3 show the results of the bias, SE and RMSE for the parameter estimates of the model, $\hat{\sigma},\hat{\beta}_0,\hat{\beta}_1$ at several n, TL and CP. Both bias and SE contribute to the average error size of an estimator, thus the $RMSE=\sqrt{Bias^2+SE^2}$ is used to measure the average overall error of the parameter estimates.

The results in Tables 1, 2 and 3 show that the values of Bias, SE, and Root Mean Square Error (RMSE) are relatively low with $\hat{\sigma}$ having higher values than $\hat{\beta}_0$ and $\hat{\beta}_1$. The trend for the bias is inconsistent and the values for the bias are relatively low for all the parameter estimates. This is due to the shape parameter which tends to produce high values of bias, SE and RMSE. As n increases, the SE and RMSE values clearly decrease. From the tables, it can be concluded that the SE and RMSE values are increasing as the CP increases but the values show no clear trend when the TL increases for all the parameter estimates. In general, we can conclude that the estimation procedure is working well with the proposed regression model.

The summary of simulation results, comparing the performance of the Wald and LR, is given in Tables 4 through 11. These tables display the empirical power, generated by each of these methods at several effect sizes, TL, CP, n and α for the Wald and LR for parameters σ and β_1 respectively.

Figures 1 through 4 give a graphical view of the empirical power for parameters σ and β_1 for the Wald and LR hypothesis tests. All the graphs show that the empirical power for both parameters increases with the increase in n for both tests. Besides that, the power of both tests also increase as the effect size which is the distance between the parameter values under H_0 & H_1 becomes larger for parameter σ and β_1. Apart from that, the power of both tests increases with the increase in the CP for both parameters involved. Having low and high TL (5% and 15%) in the data, the graphs clearly show that the empirical power of both tests increases with the increase of TL for parameter σ. However, the empirical power shows no clear changes for parameter β_1 for both tests. From Figure 4, the results show the empirical power of both tests is high at higher values of α. By comparing these two tests, we can conclude that the empirical power of the Wald is slightly better as compared with LR.

Table 1 Bias, SE and RMSE for $\hat{\sigma}$

Truncation (%)		0		10		20	
Censoring (%)		5	15	5	15	5	15
$n=20$	Bias	0.2681	0.2993	0.2796	0.3024	0.2761	0.3025
	SE	0.6826	0.7416	0.6630	0.7167	0.6613	0.7223
	RMSE	0.7334	0.7997	0.7196	0.7779	0.7167	0.7831
$n=30$	Bias	0.1647	0.1758	0.1727	0.1848	0.1652	0.1775
	SE	0.5121	0.5447	0.4905	0.5206	0.5067	0.5394
	RMSE	0.5380	0.5724	0.5201	0.5524	0.5330	0.5678
$n=50$	Bias	0.0908	0.0954	0.1013	0.1094	0.0921	0.0937
	SE	0.3640	0.3897	0.3738	0.3994	0.3701	0.3858
	RMSE	0.3752	0.4012	0.3872	0.4141	0.3813	0.3969
$n=100$	Bias	0.0601	0.0609	0.0581	0.0619	0.0547	0.0566
	SE	0.2621	0.2716	0.2630	0.2769	0.2408	0.2557
	RMSE	0.2689	0.2784	0.2694	0.2837	0.2470	0.2619

Table 2 Bias, SE and RMSE for $\hat{\beta}_0$

Truncation (%)		0		10		20	
Censoring (%)		5	15	5	15	5	15
$n=20$	Bias	-0.0106	-0.0123	-0.0078	-0.0084	-0.0016	-0.0024
	SE	0.1588	0.1644	0.1600	0.1674	0.1624	0.1707
	RMSE	0.1591	0.1649	0.1602	0.1677	0.1625	0.1707
$n=30$	Bias	0.0019	0.0014	0.0050	0.0036	0.0002	-0.0009
	SE	0.1289	0.1340	0.1251	0.1293	0.1215	0.1260
	RMSE	0.1289	0.1340	0.1252	0.1293	0.1215	0.1260
$n=50$	Bias	0.0015	0.0001	0.0024	0.0017	-0.0005	-0.0012
	SE	0.0980	0.1004	0.0966	0.0989	0.0960	0.1001
	RMSE	0.0979	0.1004	0.0966	0.0990	0.0960	0.1001
$n=100$	Bias	0.0020	0.0014	0.0011	0.0014	0.0016	0.0009
	SE	0.0648	0.0667	0.0666	0.0688	0.0670	0.0685
	RMSE	0.0648	0.0667	0.0666	0.0689	0.0671	0.0685

Table 3 Bias, SE and RMSE for $\hat{\beta}_1$

Truncation (%)		0		10		20	
Censoring (%)		5	15	5	15	5	15
$n=20$	Bias	-0.0030	-0.0057	-0.0013	-0.0051	0.0048	0.0053
	SE	0.1776	0.1863	0.1795	0.1859	0.1708	0.1767
	RMSE	0.1776	0.1864	0.1795	0.1859	0.1708	0.1767
$n=30$	Bias	-0.0037	-0.0043	-0.0042	-0.0046	-0.0048	-0.0057
	SE	0.1315	0.1394	0.1347	0.1392	0.1317	0.1356
	RMSE	0.1316	0.1395	0.1348	0.1393	0.1317	0.1358
$n=50$	Bias	-0.0016	-0.0027	0.0026	0.0032	0.0018	0.0008
	SE	0.1041	0.1071	0.0969	0.1014	0.0991	0.1025
	RMSE	0.1041	0.1071	0.0970	0.1015	0.0991	0.1025
$n=100$	Bias	0.0023	0.0027	-0.0007	-0.0001	-0.0003	-0.0002
	SE	0.0706	0.0728	0.0681	0.0694	0.0653	0.0680
	RMSE	0.0706	0.0729	0.0681	0.0694	0.0653	0.0680

Table 4 Empirical Power of Wald Test for Parameter σ at $\alpha=0.10$

Test			WALD					
Truncation (%)			0		10		20	
Censoring (%)			5	15	5	15	5	15
σ	$n=20$	1.5	0.6914	0.2170	0.7178	0.6714	0.7260	0.6796
		2.0	0.9566	0.9300	0.9688	0.9470	0.9756	0.9574
		2.5	0.9968	0.9922	0.9990	0.9926	0.9986	0.9964
	$n=30$	1.5	0.7922	0.7380	0.8148	0.7692	0.8378	0.7954
		2.0	0.9888	0.9732	0.9930	0.9858	0.9966	0.9904
		2.5	1.0000	0.9992	1.0000	0.9998	1.0000	1.0000
	$n=50$	1.5	0.9126	0.8690	0.9262	0.8924	0.9490	0.9146
		2.0	0.9994	0.9988	0.9996	1.0000	1.0000	1.0000
		2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$n=100$	1.5	0.9920	0.9816	0.9948	0.9880	0.9958	0.9934
		2.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 5 Empirical Power of Likelihood Ratio Test for Parameter σ at $\alpha=0.10$

Test			LIKELIHOOD RATIO					
Truncation (%)			0		10		20	
Censoring (%)			5	15	5	15	5	15
σ	$n=20$	1.5	0.5812	0.1764	0.6058	0.5608	0.6228	0.5820
		2.0	0.9224	0.8806	0.9396	0.9064	0.9524	0.9264
		2.5	0.9918	0.9828	0.9952	0.9846	0.9972	0.9910
	$n=30$	1.5	0.6914	0.6326	0.7240	0.6730	0.7504	0.7120
		2.0	0.9748	0.9508	0.9852	0.9698	0.9896	0.9774
		2.5	0.9999	0.9972	1.0000	0.9986	0.9998	0.9998
	$n=50$	1.5	0.8352	0.7912	0.8706	0.8194	0.8976	0.8566
		2.0	0.9992	0.9950	0.9992	0.9992	1.0000	0.9992
		2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$n=100$	1.5	0.9802	0.9570	0.9870	0.9730	0.9922	0.9828
		2.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6 Empirical Power of Wald Test for Parameter β_1 at $\alpha = 0.10$

Test			WALD					
Truncation (%)			0		10		20	
Censoring (%)			5	15	5	15	5	15
β_1	$n=20$	-0.1	0.2142	0.2098	0.2138	0.2074	0.2094	0.2082
		-0.2	0.4130	0.3996	0.4024	0.3856	0.4046	0.3888
		-0.5	0.9036	0.8846	0.9048	0.8924	0.9056	0.8914
	$n=30$	-0.1	0.2288	0.2206	0.2264	0.2226	0.2318	0.2274
		-0.2	0.5006	0.4810	0.5050	0.4848	0.5130	0.6494
		-0.5	0.9692	0.9588	0.9754	0.9664	0.9736	0.9686
	$n=50$	-0.1	0.2998	0.2844	0.2834	0.2786	0.3006	0.2850
		-0.2	0.6686	0.6472	0.6720	0.6466	0.6716	0.6494
		-0.5	0.9990	0.9982	0.9988	0.9978	0.9986	0.9984
	$n=100$	-0.1	0.4334	0.4244	0.4338	0.4234	0.4418	0.4278
		-0.2	0.8974	0.8842	0.9000	0.8864	0.8930	0.8796
		-0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 7 Empirical Power of Likelihood Ratio Test for Parameter β_1 at $\alpha = 0.10$

Test			LIKELIHOOD RATIO					
Truncation (%)			0		10		20	
Censoring (%)			5	15	5	15	5	15
β_1	$n=20$	-0.1	0.2020	0.2028	0.2022	0.1984	0.1970	0.1980
		-0.2	0.3980	0.3854	0.3842	0.3734	0.3872	0.3740
		-0.5	0.8978	0.8806	0.8988	0.8854	0.9002	0.8866
	$n=30$	-0.1	0.2222	0.2132	0.2212	0.2168	0.2228	0.2214
		-0.2	0.4934	0.4710	0.4998	0.4750	0.5048	0.6444
		-0.5	0.9682	0.9578	0.9744	0.9658	0.9728	0.9672
	$n=50$	-0.1	0.2970	0.2810	0.2788	0.2754	0.2970	0.1870
		-0.2	0.6622	0.6428	0.6668	0.6428	0.6688	0.6444
		-0.5	0.9990	0.9982	0.9982	0.9980	0.9988	0.9982
	$n=100$	-0.1	0.4324	0.4222	0.4332	0.4226	0.4402	0.3092
		-0.2	0.8966	0.8830	0.8994	0.8862	0.8920	0.8776
		-0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 8 Empirical Power of Wald Test for Parameter σ at $\alpha = 0.05$

Test			WALD					
Truncation (%)			0		10		20	
Censoring (%)			5	15	5	15	5	15
σ	$n=20$	1.5	0.5170	0.1556	0.5362	0.4862	0.5420	0.4950
		2.0	0.8930	0.8432	0.9114	0.8682	0.9238	0.8874
		2.5	0.9862	0.9728	0.9902	0.9764	0.9938	0.9844
	$n=30$	1.5	0.6444	0.5880	0.6730	0.6162	0.6926	0.6472
		2.0	0.9658	0.9374	0.9764	0.9592	0.9826	0.9664
		2.5	0.9998	0.9956	0.9960	0.9982	0.9998	0.9998
	$n=50$	1.5	0.8146	0.7622	0.8468	0.7924	0.8740	0.8214
		2.0	0.9986	0.9936	0.9984	0.9962	0.9998	0.9988
		2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$n=100$	1.5	0.9756	0.9504	0.9840	0.9690	0.9916	0.9784
		2.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 9 Empirical Power of Likelihood Ratio Test for Parameter σ at $\alpha = 0.05$

Test			LIKELIHOOD RATIO					
Truncation (%)			0		10		20	
Censoring (%)			5	15	5	15	5	15
σ	$n=20$	1.5	0.4502	0.1328	0.4748	0.4298	0.4896	0.4404
		2.0	0.8512	0.7946	0.8794	0.8272	0.9002	0.8606
		2.5	0.9764	0.9568	0.9826	0.9674	0.9914	0.9782
	$n=30$	1.5	0.5566	0.4990	0.5988	0.5436	0.6324	0.5730
		2.0	0.9422	0.9072	0.9642	0.9404	0.9730	0.9530
		2.5	0.9976	0.9934	0.9940	0.9964	0.9998	0.9988
	$n=50$	1.5	0.8352	0.7912	0.8706	0.8194	0.8976	0.8566
		2.0	0.9992	0.9950	0.9992	0.9992	1.0000	0.9992
		2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$n=100$	1.5	0.9564	0.9152	0.9698	0.9482	0.9800	0.9638
		2.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		2.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 10 Empirical Power of Wald Test for Parameter β_1 at $\alpha = 0.05$

Test			WALD					
Truncation (%)			0		10		20	
Censoring (%)			5	15	5	15	5	15
β_1	$n=20$	-0.1	0.1430	0.1414	0.1402	0.1390	0.1400	0.1356
		-0.2	0.3114	0.3006	0.3008	0.2914	0.3018	0.2900
		-0.5	0.8444	0.8192	0.8504	0.8290	0.8584	0.8334
	$n=30$	-0.1	0.1548	0.1470	0.1514	0.1486	0.1498	0.1510
		-0.2	0.3846	0.3682	0.3904	0.3712	0.4018	0.5380
		-0.5	0.9446	0.9298	0.9546	0.9390	0.9552	0.9430
	$n=50$	-0.1	0.2040	0.1960	0.1970	0.1916	0.2044	0.2818
		-0.2	0.5546	0.5212	0.5622	0.5356	0.5652	0.5380
		-0.5	0.9970	0.9952	0.9964	0.9946	0.9976	0.9960
	$n=100$	-0.1	0.3184	0.3032	0.3156	0.3002	0.3256	0.4248
		-0.2	0.8314	0.8138	0.8348	0.8128	0.8310	0.8120
		-0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 11 Empirical Power of Likelihood Ratio Test for Parameter β_1 at $\alpha = 0.05$

Test			LIKELIHOOD RATIO					
Truncation (%)			0		10		20	
Censoring (%)			5	15	5	15	5	15
β_1	$n=20$	-0.1	0.1274	0.1236	0.1250	0.1212	0.1240	0.1212
		-0.2	0.2894	0.2806	0.2752	0.2698	0.2804	0.2674
		-0.5	0.8320	0.8118	0.8370	0.8172	0.8452	0.8214
	$n=30$	-0.1	0.1436	0.1362	0.1386	0.1364	0.1376	0.1376
		-0.2	0.3688	0.3524	0.3752	0.3532	0.3806	0.5268
		-0.5	0.9402	0.9248	0.9512	0.9352	0.9504	0.9376
	$n=50$	-0.1	0.1956	0.1866	0.1880	0.1798	0.1982	0.1984
		-0.2	0.5396	0.5114	0.5466	0.5238	0.5496	0.5268
		-0.5	0.9968	0.9942	0.9958	0.9938	0.9970	0.9962
	$n=100$	-0.1	0.3140	0.2974	0.3096	0.2956	0.3206	0.3128
		-0.2	0.8272	0.8090	0.8298	0.8080	0.8258	0.8060
		-0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

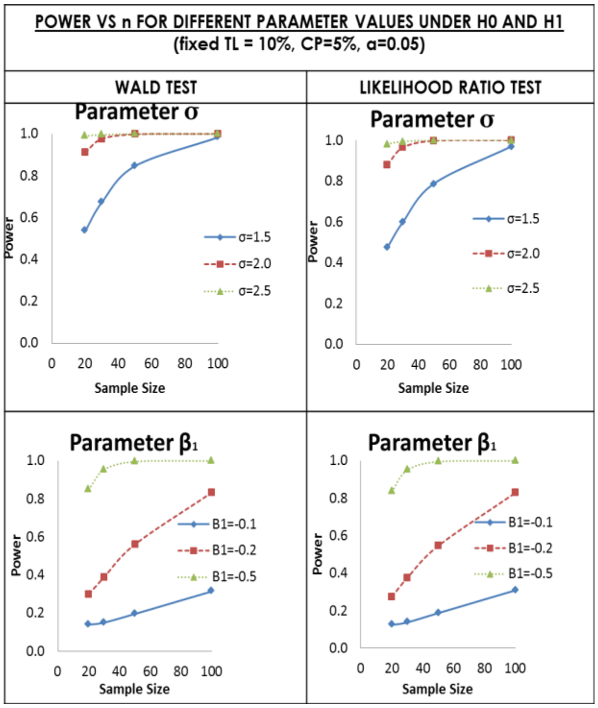


Figure 1 Power vs n for different parameter values under H_0 and H_1 for parameters σ and β_1

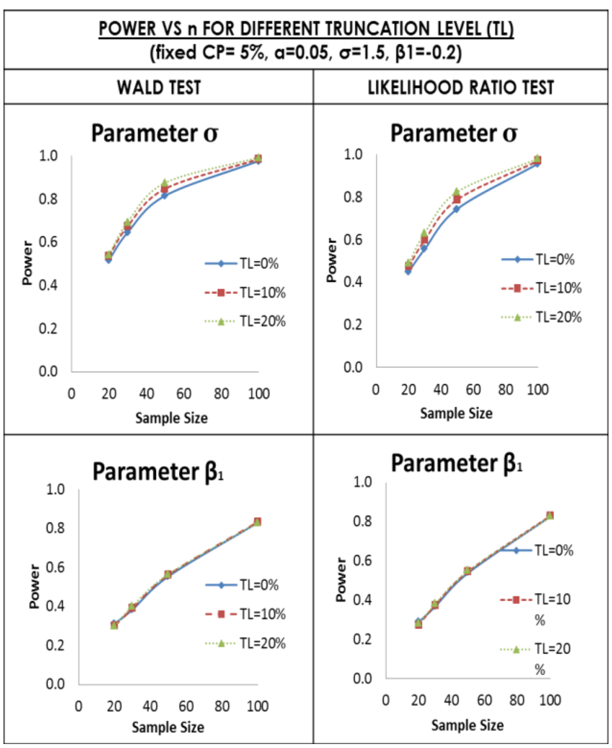


Figure 2 Power vs n for different TL for parameters σ and β_1

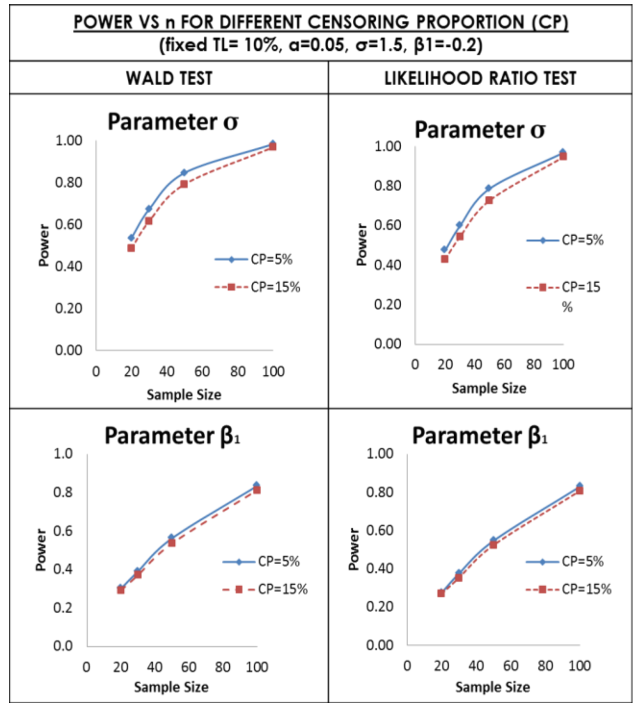


Figure 3 Power vs n for different CP for parameters σ and β_1

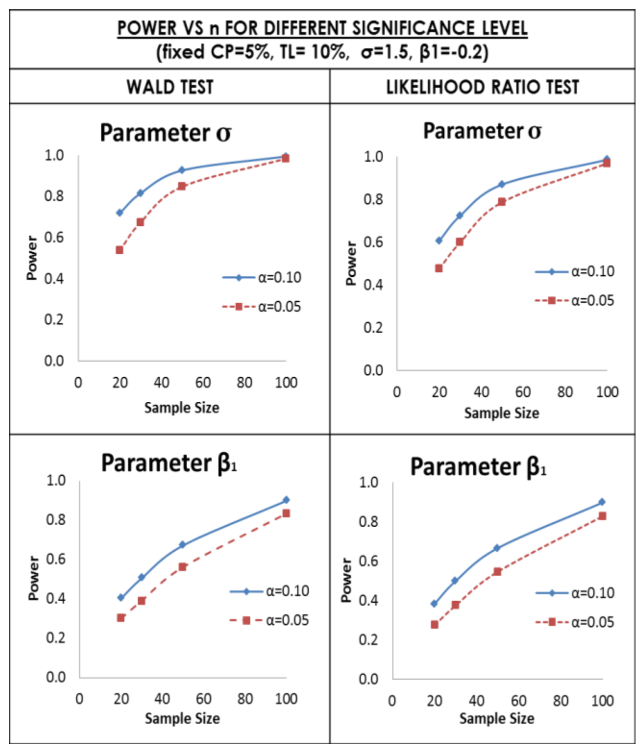


Figure 4 Power vs sample size, n for different significant level, α for parameters σ and β_1

4.0 CONCLUSION

Overall, the Wald clearly generates empirical power which is slightly higher than the LR. The overall results of the Wald are better than the LR for small and large sample size. It should be the preferred method when dealing with censored and truncated data. In general, we can conclude that the existence of the truncation and censoring in the survival data seem to affect the Wald more than the LR.

In future, other hypothesis tests can be analyzed especially those based on computer intensive techniques such as the bootstrap or jackknife. The Score test could be evaluated as an additional asymptotic test.

Acknowledgement

We would like to extend our gratitude to University Putra Malaysia and Dr. Patricia Tai, University of Saskatchewan, Saskatchtoon, Canada.

References

[1] Cain, K. C., Harlow, S. D., Little, R. J., Nan, B., Yosef, M., Taffe, J. R., and Elliott, M. R. 2011. Bias due to Left Truncation and Left Censoring in Longitudinal Studies of Developmental and Disease Processes. *American Journal of Epidemiology*. 173: 1078-84.

[2] Cox, D. R., Lewis, P. A. 1966. *Statistical Analysis of Series of Events*. London: Methuen.

[3] Doganaksoy, N., and Schmee J. 1993. Comparison of Approximate Confidence Interval Distributions used in Life-data Analysis. *Technometrics.* 2: 175-84.

[4] Emura, T. & Wang, W. 2012. Nonparametric Maximum Likelihood for Dependent Truncation Data based on Copulas. *Journal of Multivariate Analysis.* 110: 171-188.

[5] Gupta, R. C., Akman, O., & Lvin, S. 1999. A Study of Log-logistic Model in Survival Analysis. *Biometrical Journal*. 41(4): 431-443.

[6] Jayanthi Arasan & Mohd B. Adam. 2014. Double Bootstrap Confidence Interval Estimates with Censored and Truncated Data. *Journal of Modern Applied Statistical Methods.* 13(2): Article 22.

[7] Lawless, J. F. 1982. *Statistical Model and Methods for Lifetime Data*. Wiley, New York.

[8] Li, G. 1995. Nonparametric Likelihood Ratio Estimation of Probabilities for Truncated Data. *JASA*. 90: 997-1003.

[9] Mitra, Debanjan. 2013. Likelihood Inference for Left-Truncated & Right Censored Lifetime Data. McMaster University Library. Paper 7599:22-26.

[10] O'Quigley, J. & Struthers, L. 1982. Survival Models Based upon the Logistic & Log-Logistics Distribution. *Computer Programme in Biomedicine*. 15(1): 3-11.

[11] Pan, X, R. & Zhou, M. 1999. Using One Parameter Sub-family of Distributions in Empirical Likelihood with Censored Data. *J.Statist, Planning and Infer*. 75(2): 379-392.

[12] S. A. Murphy & A. W. Van Der Vaart. 1997. Semi-parametric Likelihood Ratio Inference. *The Annals of Statistics*. 25(4): 1471-1509.

[13] Shen, P. 2009. Semiparametric Analysis of Survival Data with Left Truncation and Right Censoring. *Computational Statistics and Data Analysis.* 53: 4417-32.

[14] Steve Bennet. 1983. Log-Logistic Regression Models for Survival Data. *Journal of the Royal Statistical Society*. 32(2): 165-17.

[15] Tai, P., Tonita, J., Yu, E., & Skarsgard, D. 2003. Twenty-year Follow-up Study of Long Term Survival of limited-stage Small-Cell Lung Cancer and Overview of Prognostic and Treatment Factors. *Int J. Radiation Oncology Bio/Phys*. 56: 626-633.

[16] Thirunanthini Manoharan, Jayanthi Arasan, Habshah Midi & Mohd Bakri Adam. 2015. A Coverage Probability on the Parameters of the Log-Normal Distribution in the presence of Left-Truncated and Right Censored Survival Data. *Malaysian Journal of Mathematical Sciences*. 9(1): 125-143

[17] Thirunanthini Manoharan & Jayanthi Arasan. 2013. Evaluating the Performance of the Log-Normal Survival Model with Left-Truncated & Right Censored Survival Data. Departments of Mathematics, Faculty of Science, University Putra Malaysia.

[18] Thomas, D. R. & G. L. Grunkemeier. 1975. Confidence Interval Estimation of Survival Probabilities for Censored Data, *J. Amer. Statist. Assoc*. 70: 865-871.

[19] Turnbull, B. 1976. The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *JRSS B*. 290-295.

[20] X. Luo, W. Y. Tsai. 2009. Nonparametric Estimation for Right-Censored Length Biased Data: A Pseudo-partial Likelihood Approach. *Biometrika*. 96: 873-886.