Menemui Matematik (Discovering Mathematics)
Vol. 36, No. 2: 38-50 (2014)

Parametric Survival Model in the Presence of Left-truncation and Case-*k* Interval Censoring with Fixed Covariate

Thirunanthini Manoharan[1], Jayanthi Arasan[2], Habshah Midi[3] and Mohd Bakri Adam[4]

[1,3] *Department of Mathematics, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor*
[2,4] *Applied* & Computational Statistics Laboratory , *Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor,*
[1]*mthirunanthini@gmail.com,* [2]*jayanthi@upm.edu.my*

ABSTRACT

The purpose of this study is to evaluate the performance of parametric survival model in the presence of left-truncation and case-*k* interval censoring where individuals are monitored periodically with fixed *k*-inspection times and the event of interest occurs between any two following inspection times. The log-normal distribution is extended to incorporate fixed covariate and the properties of bias, standard error (SE) and root mean square error (RMSE) were compared in the presence of low and high percentage of truncation with fixed width of inspection times. Also, the properties of bias, SE and RMSE were equally compared when the midpoint imputation technique were implemented. The simulation study indicates that the bias, SE and RMSE of the parameter estimates increases as the percentage of truncation and the width of the inspection times increases. Following that, a coverage probability study were implemented to study the performance of the Wald confidence interval method for the parameters of the log-normal distribution. The results from this study is equally applicable to the parameters of the log-logistic distribution which shares similar hazard rate with the log-normal distribution.

Keywords: left-truncation, case-*k* interval censoring, fixed covariate, log-normal distribution, midpoint imputation, Wald interval

INTRODUCTION

Left-truncation usually occurs in clinical studies when it is not feasible to observe an individual from time of contraction of certain disease but some time point later which may be due to the study design, cost or time constraint. In other words individuals are not observed from the beginning time point of the study, but some time point later *u*. It can be equally said that these individuals have been diagnosed with some initial or transitional events e.g. diagnosed with cancer or diabetes for a time period *u* before being recruited into the study. However, those who have experienced the event of interest prior e.g. cancer metastasis or death are excluded from the study or remains unobserved by the researcher. The selected individuals are equally said to be left-truncated at *u.*

Mathematically, this can be expressed as t_i being the lifetime of the i^{th} individual in the study for $i = 1,2,...,n$, and u_i is the left-truncation time. As individuals have to survive long enough before being recruited into the study, their lifetime, $t_i \geq u_i$. Consequently, individuals with $t_i < u_i$ are disregarded or excluded from the research. The selected individuals are followed prospectively with fixed *k* inspection times where the exact event time is unknown except that it falls within an interval of $(t_{l_i}, t_{r_i}]$ where $t_i \in (t_{l_i}, t_{r_i}]$ with $\Pr(t_{l_i} \leq t_{r_i}) = 1$. This type of data is known as left-truncated and case-*k* interval censored (LTIC) survival data, where left-truncated observations are existing

cases usually sampled from registry record. As date of diagnosis may differ from one individual to the other, individuals may enter the study at random age or time points, however in the presence of left-truncation only those who are free from failure are observed by the researcher (Guo, 1976). Further, other factors that affect the lifetime, known as covariates, x are only observed from the time of entry of individuals into the study, see Guo (1976), Lawless (1982) and Klein and Moeschberger (2003). Also, the lifetime after selection forms the response variable, hence the term prospective, Lawless (1982). In other words, the truncation time u contains no information on lifetime t, or equally t is independent of u.

Two types of covariates that are usually measured in a survival study are fixed and time-dependent covariates. Fixed covariates are covariates that is measured at the beginning time-point of the study and stays constant throughout the study. Example of such covariate includes the gender and ethnicity of individuals in the study. In contrary, time-dependent covariates vary over time. This study focuses on fixed covariates.

RESEARCH BACKGROUND

Sun (2007) indicated that research work involving left-truncation and interval censoring is limited although truncation is equally observed in the presence of censoring. Furthermore, statistical packages such as *R* and *S-Plus* does not accommodate the analysis of LTIC survival data, thus requires statisticians to develop functions in accordance to a specified model which fits the data in hand.

Following that, parametric survival models often remain a useful tool as they are fitted much faster and offers more efficient estimates under conditions such as dependency of lifetime of individual on covariates or when parameter values are far from zero, see Klein and Moeschberger (2003), Nardi and Schemper (2003), Cox and Oakes (1984). Parametric models involving left-truncated survival data has been discussed by Lawless (1982), Guo (1992), Klein and Moeschberger (2003), Cain et.al (2011) and Balakrishnan and Mitra (2011,2012,2014) among others, where all the researchers unanimously agreed to the conditional likelihood approach in estimating parameters of a model fitted with left-truncated survival data.

Midpoint imputation is another procedure adopted by researchers when data is interval censored. By utilizing this method, the effect of interval censoring is practically ignored by treating the interval censored failure times as exact failures by taking the midpoint of intervals where the event has occurred. However, Lindsey (1998) reported that this approximation is not constantly consistent. Additionally, Shen (2011) reported that using midpoint imputation method to estimate the parameters of Cox's semi-parametric model involving LTIC survival data resulted in larger bias, standard error (SE) and root mean square error (RMSE) compared to when the event is considered as interval censored in the estimation procedure. Also, Stovring and Kristiansen (2011) compared midpoint and multiple imputation procedure to identify suitable parametric model to estimate mean survival times of patients for left-truncated and grouped survival data. he further indicated that estimation procedure involving midpoint imputation resulted in satisfactory bias, SE and coverage probability provided that the inspection width is no more than 6months.

Nevertheless, many of the existing research on left-truncation involving semi-parametric and parametric models does not accommodate the covariates effects on the lifetimes, although this is a significant reason on employing these models which allows survival to be measured with reference to several covariates.

In this research the log-normal distribution is considered as it is often a popular choice to model cancer survival data based on the ability to accommodate non-

monotonic hazard rate, the hazard that increases, reaches a peak and later decreases. A detailed review on survival times and cancer sites has been discussed by Tai.et.al (2004). Also, the log-normal distribution belonging to the family of log-location scale model shares similar hazard function with the log-logistic distribution. Following that, the survival times of observations that satisfy a log-normal distribution has low mortality in the beginning, reaches a peak where the rate of mortality is the highest after which it slowly decreases with time Tai.et.al (2004).

RESEARCH OBJECTIVES

In this research, the log-normal survival model is extended to incorporate observations from prevalence (existing cases) and incidence (new cases) cohort encountered in a cancer survival study whom are monitored periodically with fixed *k* inspection times, where the exact event time is known to fall between two following inspection times. Also, the covariate factors which influence their lifetime are equally measured.

The performance of the proposed parametric model is assessed based on the bias, SE and RMSE of the parameter estimates. The robustness of this model is equally compared using midpoint imputation (mid.imp) method, where the exact event time is taken as the midpoint of the two following inspection times. Following that, the coverage probability study is conducted to study the Wald confidence interval method for the parameters of the log-normal distribution. In addition, the suitability of the parameterization of $\log(\sigma)$, based on the Wald method is equally analyzed for the scale parameter σ.

LOG-NORMAL MODEL WITH LEFT-TRUNCATION AND FIXED COVARIATE

In this study, we considered a single fixed covariate. The density and survival function of the log-normal distribution is given by (1) and (2) respectively,

$$f(t_i) = \frac{1}{t_i \sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left[\left(\frac{\log t_i - (\beta_0 + \beta_1 x_i)}{\sigma}\right)^2\right]}, \quad (1)$$

$$S(t_i) = 1 - \Phi\left(\frac{\log t_i - (\beta_0 + \beta_1 x_i)}{\sigma}\right), \quad (2)$$

with $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, σ the scale or the nuisance parameter, β_0 the intercept parameter, β_1 the covariate parameter and $i = 1, 2, ..., n$. Following that, the likelihood function for the observations from the prevalence cohort and incidence cohort with $\boldsymbol{\theta} = (\sigma, \beta_0, \beta_1)$ is given in (3) and (4).

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{\frac{f(t_i)}{S(u_i)}\right\}^{\delta E_k} \left\{\frac{S(t_{l_i})}{S(u_i)}\right\}^{\delta R_i} \left\{\frac{S(t_{l_i}) - S(t_{r_i})}{S(u_i)}\right\}^{\delta I_i}, \quad (3)$$

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ f(t_i) \right\}^{\delta E_i} \left\{ S(t_{l_i}) \right\}^{\delta R_i} \left\{ S(t_{l_i}) - S(t_{r_i}) \right\}^{\delta I_i} \left\{ 1 - S(t_{r_i}) \right\}^{\delta L_i}, \qquad (4)$$

with t_{l_i} and t_{r_i} are the left and right end points of an interval $(t_{l_i}, t_{r_i}]$ where $\Pr(t_{l_i} \le t_{r_i}) = 1$, t_i is the exact failure times and u_i is the left truncated times. Also the censoring indicators is defined in (5) as follows:

$$\delta E_i = \begin{cases} 1 \text{ if individual's failure time is observed exactly} \\ 0 \text{ otherwise} \end{cases}$$

$$\delta R_i = \begin{cases} 1 \text{ if individual's failure time is right censored} \\ 0 \text{ otherwise} \end{cases}$$

$$\delta I_i = \begin{cases} 1 \text{ if individual's failure time is interval censored} \\ 0 \text{ otherwise} \end{cases}$$

$$\delta L_i = \begin{cases} 1 \text{ if individual's failure time is left censored} \\ 0 \text{ otherwise.} \end{cases} \qquad (5)$$

Note that left-censored observations are only observed among the incidence cohort. Therefore, the log-likelihood function for the prevalence cohort and incidence cohort can be derived by combining the likelihood function as in (3), (4) and (5) by including a truncation indicator v_i. This is defined in (6) and (7).

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & -\sum_{i=1}^{n} \left[\delta E_i \log\{t_i \sigma\} \right] - \sum_{i=1}^{n} \left[\frac{\delta E_i}{2\sigma^2} \left\{ (\log t_i - \mu)^2 \right\} \right] + \sum_{i=1}^{n} \left[\delta R_i \log \left\{ 1 - \Phi\left(\frac{\log t_{l_i} - \mu}{\sigma} \right) \right\} \right] \\ & + \sum_{i=1}^{n} \left[\delta I_i \log \left\{ \Phi\left(\frac{\log t_{r_i} - \mu}{\sigma} \right) - \Phi\left(\frac{\log t_{l_i} - \mu}{\sigma} \right) \right\} \right] + \sum_{i=1}^{n} \left[\delta L_i \log \left\{ \Phi\left(\frac{\log t_{r_i} - \mu}{\sigma} \right) \right\} \right] \\ & - \sum_{i=1}^{n} \left[(1 - v_i) \log \left\{ 1 - \Phi\left(\frac{\log u_i - \mu}{\sigma} \right) \right\} \right], \end{aligned} \qquad (6)$$

with $\mu = \beta_0 + \beta_1 x_i$ and the truncation indicator v_i, is defined in (7) as follows:

$$v_i = \begin{cases} 0 \text{ if subject is left-truncated} \\ 1 \text{ otherwise.} \end{cases} \qquad (7)$$

In the case when midpoint imputation procedure is applied to the LTIC survival data, the likelihood in (6) reduces to observations with exact failure times (midpoint of two following inspection times) and right censored (if the event of interest is not observed even after the study ends). For instance, let $t_i \in (t_{l_i}, t_{r_i}]$, with t_{l_i} and t_{r_i} are the right and left end points of the pre-specified intervals. Thus using the midpoint imputation (mid.imp), the interval censored time is assigned as exact failure times with $\tilde{t}_i = (t_{l_i} + t_{r_i})/2$, which transforms the data to left-truncated and right-censored observations. This likelihood is defined in (8) as follows:

$$\ell(\boldsymbol{\theta}) = -\sum_{i=1}^{n}\left[\delta E_i \log\{\tilde{t}_i\sigma\}\right] - \sum_{i=1}^{n}\left[\frac{\delta E_i}{2\sigma^2}\left\{\left(\log\tilde{t}_i - \mu\right)^2\right\}\right] + \sum_{i=1}^{n}\left[\delta R_i \log\left\{1-\Phi\left(\frac{\log t_{l_i} - \mu}{\sigma}\right)\right\}\right] + \sum_{i=1}^{n}\left[(1-v_i)\log\left\{1-\Phi\left(\frac{\log u_i - \mu}{\sigma}\right)\right\}\right]. \quad (8)$$

CONFIDENCE INTERVAL ESTIMATE

Asymptotic based confidence interval (C.I) such as Wald is most commonly applied confidence interval method in any survival study as it easily computable and readily available in any statistical packages. The performance of the Wald and parameterization based C.I such as $\log(\sigma)$ for the scale parameter, σ is equally explored through a coverage probability study.

A coverage probability study is the probability of a C.I containing the true parameter value. In specific, we do not want a conservative (anticonservative) interval which the probability of an interval containing the true parameter value is wider (shorter) than it needs to be. Further, we do not want an asymmetrical interval where the probability that the true value of the parameter estimate failing on one end of an interval is higher/lower than the other end of the interval.

An optimal and reliable C.I method generates least number of anticonservative, conservative and asymmetrical intervals in addition that the total error probability closer to the nominal error being evaluated (Doganaksoy and Schmee, 1993). The construction of the Wald and parameterization of $\log(\sigma)$ addressed as PLS C.I method in this study is estimated using parameter σ as an example, which equally applies to the rest of the parameters in the model. Note that the PLS method is only applicable to the scale parameter σ.

Let $\hat{\sigma}$ be the mle of σ. The $100(1-\alpha)\%$ C.I for the parameter σ can be estimated as in (9) given by,

$$\hat{\sigma} - z_{1-\frac{\alpha}{2}}\sqrt{\text{var}(\hat{\sigma})} < \sigma < \hat{\sigma} + z_{1-\frac{\alpha}{2}}\sqrt{\text{var}(\hat{\sigma})}, \quad (9)$$

with $\text{var}(\hat{\sigma})$ the first diagonal element of the observed Fisher information matrix $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimate (mle) of vector of parameter $\boldsymbol{\theta}$. By utilizing the same principle, the PLS C.I for $\log(\sigma)$ is given as in (10),

$$\log(\hat{\sigma}) \pm z_{1-\frac{\alpha}{2}}\sqrt{\text{var}\left[\log(\hat{\sigma})\right]}, \quad (10)$$

where the variance of $\log(\sigma)$ can be estimated using the delta method and is given as $\text{var}\left[\log(\hat{\sigma})\right] \approx \frac{\text{var}(\hat{\sigma})}{\left[\exp\left[\log(\hat{\sigma})\right]\right]^2} \approx \frac{\text{var}(\hat{\sigma})}{\hat{\sigma}^2}$. Therefore the $100(1-\alpha)\%$ C.I for the parameter σ using the result from (10) can be obtained using the back transformation

method given by,

$$\hat{\sigma}\exp\left(-z_{1-\frac{\alpha}{2}}\frac{\sqrt{\text{var}(\hat{\sigma})}}{\hat{\sigma}}\right)<\sigma<\hat{\sigma}\exp\left(z_{1-\frac{\alpha}{2}}\frac{\sqrt{\text{var}(\hat{\sigma})}}{\hat{\sigma}}\right). \quad (11)$$

SIMULATION AND COVERAGE PROBABILITY STUDY

The simulation method mimics the small cell lung cancer survival data studied by Tai et.al (2007) which provides a satisfactory fit with the log-normal distribution. The year of truncation, namely y is fixed. A set of random number of years is simulated with unequal probabilities with replacement; before (y_{b_r}) and after (y_{a_s}) the year of truncation with r=1,2,..,n_1 *and* s=1,2,..,n_2 to represent the prevalence cohort (PC) and incidence cohort (DC) respectively.

The percentage of truncation is fixed at 20% (T2) and 60% (T6) for a sample size of $n = 80,100,150$ and 200. Note that the total sample size $n = n_1 + n_2$. The lifetimes, t_i are simulated from the log-normal distribution as $t_i = \exp(\sigma \times \Phi(1 - z_i) + \mu)$ for $i = 1,2,...,n$ with $z_i \sim unif(0,1)$, σ and μ are the scale and the mean parameter respectively. Additional parameters, β_0 and β_1 are modelled through μ with covariate $x_i \sim N(0,1)$ as $\mu = \beta_0 + \beta_1 \times x_i$. The true values of the parameters given as the vector of $\boldsymbol{\theta} = (\sigma, \beta_0, \beta_1) = (0.50, 2.87, 0.05)$.

The lifetimes, t_i are added to y_{b_r} and y_{a_s}; if the resulting failure times are less than y, these observations are removed and new set of y_{b_k}, y_{a_j}, t_i, z_i and x_i are simulated. Individuals are monitored periodically with fixed k inspection times and sequence of potential inspection times $m_{i1} \le m_{i2} \le ... \le m_{ik}$ are assumed to the same for all the individuals in the study. Further, the exact survival times for the PC and DC is unknown despite the event time t_i falls between any two following inspection times; e.g. $\left(t_{l_i}, t_{r_i}\right] = \left(m_{i(j-1)}, m_{ij}\right]$ with $1 \le j \le k$ or after the last inspection times; $\left(t_{l_i}, t_{r_i}\right] = \left(m_{ik}, \infty\right]$ which produces interval (IC) and right censored (RC) observations with t_{l_i} and t_{r_i} are the right and left end points of the intervals. Additionally, for the DC, the event of interest could have occurred at unknown time, m_{i1}; e.g. $\left(t_{l_i}, t_{r_i}\right] = \left(0, m_{i1}\right]$ and after the time origin producing left censored (LC) observations. Exact observation (EO) of lifetimes are available for PC and DC if the event of interest is observed within the observational window of $\left[t_{r_i} - \varepsilon, t_{r_i}\right] = \left[E_{1i}, E_{2i}\right]$ with $\varepsilon = 0.90$, where the event of interest occurs closer to the time of inspection. Additionally, by implementing the midpoint imputation procedure, all the event times that falls within the interval of $\left(t_{l_i}, t_{r_i}\right]$ with $t_{r_i} \ne \infty$ is imputed as EO of lifetimes.

The study period is assumed to be 60 months with the width of inspection intervals 2 months (W2) ($k = 30$) and 4 months (W4) ($k = 15$). The attributes of bias, SE and RMSE for parameter estimates $\hat{\sigma}, \hat{\beta}_0$ and $\hat{\beta}_1$ are compared under four different

settings, M1 (T2,W2), M2 (T2,W4), M3 (T6,W2) and M4 (T6, W4), equally using the midpoint imputation method. In order to aid the conduct of the coverage probability study, 2000 samples of size $n = 80,100,150$ and 200 were generated. The nominal probability error (npe) is set at $\alpha = 0.05$. The performance of the Wald and PLS C.I method are evaluated under similar settings indicated above. The error probabilities on the left (lep) and right (rep) for parameter σ were estimated as the number of times the C.I did not contain the true value of σ divided by the number of simulations; 2000 times. Therefore, the estimated total error probability (tep) for σ is simply the sum of lep and rep.

Following that outcome, a CI method is termed anticonservative (AC) if $\text{tep} > \alpha + 2.58 \times \text{se}(\hat{\alpha})$, conservative (C) if $\text{tep} < \alpha - 2.58 \times \text{se}(\hat{\alpha})$ with $se(\hat{\alpha}) = \sqrt{\alpha(1-\alpha)/N}$. Also, the estimated error probabilities are asymmetric (AS) when the larger error probabilities on one side of the interval is greater than 1.5 times the smaller one. A preferred confidence interval method produces least number of AS, C and AC intervals, the value of the lep and rep closer to 0.025 and the value of the tep closer to npe of 0.05, refer Doganaksoy and Schmee (1993).

In this study, it is assumed that t_i, u_i and censoring times are non-informative and independent of each other. Also, the exact month of diagnosis is known for all observations in this study and these observations were event free at the time of entry into the study. All the analysis is done with *R* statistical software and the parameter estimates are obtained using the Newton-Raphson iteration procedure.

RESULTS AND DISCUSSION

The results in Table 1 shows the average percentages of IC, RC, LC and EO failure times under the settings of M1 (T2,W2), M2 (T2,W4), M3 (T6,W2) and M4 (T6, W4).

By fixing the width of inspection times, e.g. compare (M1 and M3) or (M2 and M4), the percentage of IC and EO failure times are approximately closer to each other. However, the percentage of RC observations are slightly higher for M1 as opposed to M3. This may be due to the lower percentage of left-truncated observations in M1, which consequently increases the recruitment of new cases observed under M1 compared to M3. In other words, the possibility of patients not experiencing the event of interest (e.g. cancer metastasis) even after the last inspection times are higher among those whom have been recently diagnosed with a disease compared to those whom have been living with the disease for some period of time prior to entry into the study.

In contrary, by fixing the percentage of truncation, e.g. compare (M1 and M2) or (M3 and M4), the percentage of observations with EO lifetimes are higher when the width of inspection times are narrower. This is as expected, as shorter width of inspection times increases frequency of inspection times. For instance width of inspection intervals of 2 months (4 months) results in $k = 30$ ($k = 15$) for a total period of 60 months. This subsequently increases the possibility of observing event of interest occurring very close to the inspection times resulting in increased number of EO lifetimes. This equally reduces the percentage of IC observations comparatively when the width of inspection times are wider. Further, there are small percentage of LC lifetimes among the DC with wider width of inspection times, e.g. M2 and M4 as longer awaiting time may result observations experiencing event of interest at unknown times before the first inspection time.

Table 1: The average percentage of IC, RC, LC and EO for M1, M2, M3 and M4

%	M1	M2	M3	M4
IC	0.5475	0.7717	0.5496	0.7764
RC	0.0069	0.0068	0.0057	0.0057
LC	0.0000	0.0002	0.0000	0.0001
EO	0.4456	0.2213	0.4447	0.2178

Table 2 depicts the bias, SE and RMSE of the parameter estimates $\hat{\sigma}, \hat{\beta}_0$ and $\hat{\beta}_1$. It can be observed that the absolute bias for parameter estimate $\hat{\sigma}$ generally decreases with the increase in the sample size although the trend seems to be unclear for parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. However, none of these values seems to be a concern as these values are insignificant at $\alpha = 0.05$ or 0.10 level of significance. Further, the SE and RMSE decreased with increase in sample size under all settings.

However, by fixing the width of inspection times, it is evident that the SE and RMSE are higher when the percentage of left-truncation is higher. This is due to the increase number of observations that is excluded from the left-tail of the log-normal distribution in the presence of higher percentage of truncation consequently increasing sampling bias, SE and subsequently the RMSE of the parameter estimates.

On the other hand, the values of RMSE of the parameter estimates are lower at shorter width of inspection times. This is due to the fact that, shorter width of inspection times results in higher number of inspection which additionally resulted in higher percentage of EO of lifetimes, see Table 1. As more information is gained on the lifetimes of observations in the study, this results in lower values of SE and RMSE of the parameter estimates comparatively when larger width of inspection time is observed.

Similar results are observed on the bias, SE and RMSE of the parameter estimates $\hat{\sigma}, \hat{\beta}_0$ and $\hat{\beta}_1$ by adopting the mid.imp procedure, see Table 3. Further, the bias, SE and RMSE of the parameter estimates are approximately closer to the values observed in Table 2 specifically at lower percentage of truncation. Nevertheless, when higher proportion of truncation is present, e.g. refer M3 and M4, the estimation procedure fails to converge especially for sample size of 80, see Table 3.

Table 2: Bias , SE and RMSE of estimates $\hat{\sigma}, \hat{\beta}_0$ and $\hat{\beta}_1$ for M1, M2, M3 and M4

parameter		$\hat{\sigma}$			$\hat{\beta}_0$			$\hat{\beta}_1$		
setting	*n*	bias	SE	RMSE	bias	SE	RMSE	bias	SE	RMSE
M1	80	-0.0207	0.0391	0.0443	-0.0326	0.0554	0.0643	-0.0019	0.0570	0.0571
	100	-0.0211	0.0352	0.0410	-0.0284	0.0496	0.0571	-0.0010	0.0510	0.0510
	150	-0.0178	0.0291	0.0341	-0.0306	0.0405	0.0507	-0.0015	0.0401	0.0401
	200	0.0179	0.0254	0.0311	-0.0294	0.0352	0.0459	0.0018	0.0357	0.0358
M2	80	-0.0217	0.0402	0.0457	-0.0326	0.0558	0.0646	-0.0019	.0574	0.0575
	100	0.0222	0.0358	0.0422	-0.0285	0.0498	0.0574	-0.0009	.0513	0.0513
	150	-0.0191	0.0298	0.0354	-0.0310	0.0407	0.0512	-0.0015	.0404	0.0404
	200	-0.0192	0.0260	0.0323	-0.0297	0.0356	0.0464	-0.0017	.0359	0.0360
M3	80	-0.0215	0.0414	0.0466	-0.0348	0.0578	0.0675	-0.0002	0.0592	0.0592
	100	-0.0216	0.0362	0.0421	-0.0343	0.0521	0.0623	-0.0006	0.0510	0.0510
	150	-0.0203	0.0305	0.0366	-0.0340	0.0416	0.0537	-0.0014	0.0424	0.0424
	200	-0.0176	0.0263	0.0317	-0.0344	0.0371	0.0506	-0.0015	0.0371	0.0372
M4	80	0.0223	0.0478	0.0478	-0.0345	0.0584	0.0678	-0.0004	0.0597	0.0597
	100	0.0224	0.0372	0.0435	-0.0343	0.0524	0.0626	-0.0007	0.0513	0.0513

150	0.0211	).0310	0.0375	).0339	).0417	0.0538	-0.0014	0.0427	0.0427
200	0.0187	).0270	0.0328	).0345	).0373	0.0508	-0.0014	0.0374	0.0375

Table 3: Bias , SE and RMSE of estimates $\hat{\sigma}, \hat{\beta}_0$ and $\hat{\beta}_1$ for M1, M2, M3 and M4 using mid.imp

paramete		$\hat{\sigma}$			$\hat{\beta}_0$			$\hat{\beta}_1$		
settin	n	bias	SE	RMS	bias	SE	RMS	bias	SE	RMS
M1	8(	-0.021	0.038	0.043	0.029	0.056	0.063	-0.001	0.056	0.056
	10	-0.020	0.035	0.040	0.029	0.049	0.057	-0.001	0.049	0.049
	15	-0.017	0.028	0.033	0.028	0.039	0.048	0.000	0.041	0.041
	20	-0.016	0.024	0.029	0.031	0.035	0.046	-0.003	0.034	0.034
M2	8(	-0.018	0.039	0.043	0.027	0.058	0.064	-0.002	0.057	0.057
	10	-0.015	0.035	0.038	0.026	0.050	0.056	-0.001	0.050	0.050
	15	-0.014	0.029	0.032	0.028	0.040	0.050	-0.002	0.041	0.041
	20	-0.012	0.025	0.028	0.029	0.035	0.046	-0.001	0.035	0.035
M3	8(	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**
	10	-0.021	0.036	0.042	0.033	0.051	0.061	-0.003	0.052	0.052
	15	-0.018	0.029	0.035	0.034	0.041	0.053	0.000	0.040	0.040
	20	-0.016	0.026	0.031	0.034	0.037	0.051	-0.000	0.036	0.036
M4	8(	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**
	10	-0.016	0.038	0.041	0.032	0.052	0.062	0.002	0.052	0.052
	15	-0.015	0.030	0.033	0.033	0.043	0.054	-0.003	0.043	0.043
	20	-0.014	0.027	0.030	0.032	0.037	0.049	-0.000	0.037	0.037

The coverage probability study indicates the Wald method performed poorly with parameter σ and β_0 as more AC and AS C.I are produced regardless of large sample sizes under all settings, see Table 4. Also the parameterization of the Wald method, the PLS method did not improve the performance of the Wald interval for parameter σ as the number of AC and AS intervals remained the same. Additionally, the estimated tep are far from the npe of 0.05 and distance increased with the increase in sample size, see Table 5 and 6.

In contrary, the Wald method performed fairly well for parameter β_1 with least number of AC, C and AS intervals, observed in the presence of higher proportion of left-truncation, see Table 4. Additionally, the estimated tep are closer to 0.05 despite percentage of truncation and width of inspection times.

Similar results are observed by implementing the mid.imp procedure, see Table 7-9. However, in the presence of higher percentage of left-truncation, e.g. refer M3 and M4, the Wald and PLS C.I method failed to work for samples less than 80, see Table 8-9.

This results from the exclusion of observations from the left-tail of the log-normal distribution with lifetime $t < u$, consequently resulting in skewed data, see Cain et.al (2013) and Manoharan (2013). Further, as censoring equally causes the data to be incomplete, assumption of normality often fails as it can't fully capture the sampling distribution of the sample statistics being studied. Therefore, inferential techniques that is heavily dependable on normality assumptions such as Wald and likelihood ratio may perform poorly with the parameters from a specific distribution. This has been equally demonstrated by Manoharan (2015) where asymptotic intervals such as Wald and likelihood ratio generated many AC and AS intervals with parameter σ and β_0 of the

log-normal distribution when data is highly truncated and right-censored.

Table 4: **Number of AS, C and AC C.I for estimates $\hat{\sigma}, \hat{\beta}_0$ and $\hat{\beta}_1$ for M1, M2, M3 and M4**

C.I	parameter	M1			M2			M3			M4		
		AC	C	AS	AC	C	AS	AC	C	AS	AC	C	AS
Wald	$\hat{\sigma}$	4	0	4	4	0	4	4	0	4	4	0	4
	$\hat{\beta}_0$	4	0	4	4	0	4	4	0	4	4	0	4
	$\hat{\beta}_1$	0	0	0	0	0	0	0	0	1	0	0	1
PLS	$\hat{\sigma}$	4	0	4	4	0	4	4	0	4	4	0	4

Table 5: Estimated Wald error probabilities for parameters σ, β_0 and β_1 for M1, M2, M3 and M4

parameter		σ			β_0			β_1		
setting	n	lep	rep	tep	lep	rep	tep	lep	rep	tep
M1	80	0.004	0.108	0.112	0.086	0.005	0.091	0.028	0.031	0.059
	100	0.003	0.126	0.128	0.093	0.007	0.100	0.035	0.029	0.063
	150	0.003	0.118	0.121	0.110	0.003	0.113	0.025	0.022	0.047
	200	0.001	0.148	0.149	0.135	0.002	0.137	0.025	0.0315	0.056
M2	80	0.004	0.112	0.116	0.089	0.006	0.095	0.028	0.032	0.060
	100	0.003	0.125	0.128	0.096	0.007	0.103	0.032	0.029	0.061
	150	0.003	0.124	0.127	0.109	0.003	0.112	0.028	0.022	0.050
	200	0.001	0.148	0.149	0.136	0.002	0.138	0.024	0.033	0.057
M3	80	0.003	0.111	0.114	0.098	0.003	0.101	0.026	0.026	0.052
	100	0.003	0.118	0.121	0.109	0.003	0.112	0.022	0.023	0.045
	150	0.002	0.130	0.132	0.124	0.003	0.127	0.024	0.025	0.049
	200	0.003	0.130	0.133	0.175	0.001	0.176	0.024	0.037	0.061
M4	80	0.002	0.108	0.110	0.096	0.002	0.098	0.025	0.028	0.053
	100	0.001	0.119	0.120	0.109	0.003	0.112	0.019	0.022	0.041
	150	0.002	0.132	0.134	0.122	0.003	0.125	0.026	0.023	0.049
	200	0.003	0.137	0.140	0.176	0.001	0.177	0.024	0.037	0.061

Table 6: Estimated PLS error probabilities for parameters σ for M1, M2, M3 and M4

setting	M1			M2			M3			M4		
n	lep	rep	tep	lep	rep	tep	lep	rep	tep	lep	rep	tep
80	0.001	0.123	0.124	0.001	0.133	0.134	0.001	0.131	0.132	0.001	0.135	0.136
100	0.002	0.165	0.137	0.002	0.164	0.166	0.003	0.153	0.156	0.001	0.159	0.160
150	0.002	0.143	0.145	0.002	0.144	0.146	0.001	0.163	0.164	0.001	0.155	0.156
200	0.000	0.169	0.169	0.000	0.170	0.170	0.002	0.156	0.158	0.001	0.163	0.164

Table 7: **Number of AS, C and AC C.I for estimates $\hat{\sigma}, \hat{\beta}_0$ and $\hat{\beta}_1$ for M1, M2, M3 and M4 using mid.point**

C.I	parameter	M1			M2			M3			M4		
		AC	C	AS	AC	C	AS	AC	C	AS	AC	C	AS
Wald	$\hat{\sigma}$	4	0	4	4	0	4	3	0	3	3	0	3
	$\hat{\beta}_0$	4	0	4	4	0	4	3	0	3	3	0	3
	$\hat{\beta}_1$	0	0	1	0	0	0	0	0	0	0	0	0
PLS	$\hat{\sigma}$	4	0	4	4	0	4	3	0	3	3	0	3

Table 8: **Estimated Wald error probabilities for parameters σ, β_0 and β_1 for M1, M2, M3 and M4 using mid.point**

parameter		σ			β_0			β_1		
setting	n	lep	rep	tep	lep	rep	tep	lep	rep	tep
M1	80	0.002	0.105	0.107	0.085	0.007	0.092	0.025	0.032	0.057
	100	0.001	0.121	0.122	0.090	0.005	0.095	0.027	0.028	0.055
	150	0.002	0.121	0.123	0.112	0.002	0.114	0.027	0.029	0.056
	200	0.004	0.108	0.112	0.148	0.003	0.151	0.017	0.032	0.049
M2	80	0.004	0.104	0.108	0.075	0.013	0.088	0.027	0.036	0.063
	100	0.004	0.091	0.095	0.081	0.006	0.087	0.023	0.024	0.047
	150	0.005	0.099	0.104	0.117	0.003	0.120	0.022	0.032	0.054
	200	0.007	0.096	0.103	0.134	0.003	0.137	0.021	0.026	0.047
M3	80	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**
	100	0.000	0.122	0.122	0.105	0.003	0.108	0.027	0.029	0.056
	150	0.001	0.120	0.120	0.125	0.003	0.128	0.023	0.020	0.043
	200	0.003	0.125	0.128	0.168	0.003	0.171	0.030	0.027	0.057
M4	80	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**	**Fail**
	100	0.004	0.099	0.103	0.101	0.003	0.103	0.023	0.027	0.050
	150	0.003	0.102	0.105	0.129	0.003	0.132	0.025	0.036	0.061
	200	0.003	0.119	0.122	0.149	0.002	0.151	0.026	0.027	0.053

Table 9: **Estimated PLS error probabilities for parameters σ for M1, M2, M3 and M4 using mid.point**

setting	M1			M2			M3			M4		
n	lep	rep	tep	lep	rep	tep	lep	rep	tep	lep	rep	tep
80	0.001	0.123	0.124	0.001	0.107	0.108	Fail	Fail	Fail	Fail	Fail	Fail
100	0.000	0.127	0.127	0.001	0.089	0.090	0.000	0.123	0.123	0.003	0.114	0.117
150	0.001	0.107	0.108	0.001	0.074	0.075	0.000	0.117	0.117	0.001	0.078	0.079
200	0.001	0.096	0.097	0.001	0.073	0.074	0.001	0.120	0.120	0.001	0.092	0.093

CONCLUSIONS

In conclusion, the log-normal survival model performed well under all settings and is robust against higher percentage of truncation and censoring present in the data. However, this parametric estimator performed best by producing estimates with lower values of bias, SE and RMSE when lower proportion of left-truncation and/or shorter width of inspection interval is present in the LTIC survival data. In other words, the estimation procedure generated more efficient and accurate parameter estimates with inclusion of cases from incidence cohort observed with frequent number of inspection times.

Following that, the midpoint imputation procedure applied to the LTIC data equally worked well under similar conditions indicated above, nonetheless, we recommend this method to be implemented when smaller percentage of left-truncation , e.g. $\leq \mathrm{T2}$ is observed for small sample data, e.g. $n \leq 80$, as the imputation method suffered convergence problems at higher percentage of truncation, e.g. $\geq \mathrm{T6}$.

On the other hand, the coverage probability study indicated that the Wald method performed well with the covariate parameter, β_1 at $\alpha = 0.05$. In contrary, the Wald method performed poorly with parameter σ and β_0. Additionally, the parameterization of $\log(\sigma)$ did not improve the performance of the Wald method for parameter σ as it generated mostly anticonservative and asymmetrical intervals. Thus, the Wald and PLS method is not recommended for parameter σ and β_0 as inference drawn from such intervals will be unreliable; e.g. there are higher possibility of a researcher rejecting the true value of a desired parameter when the intervals appear to be shorter in length or where the estimated error probabilities are higher than it needs to be Manoharan et.al, (2015). Therefore, there is a necessity to investigate the performance of confidence interval methods which relaxes on the assumption of normality and alternatively rely on the distribution of data in hand. On that basis, the bootstrap confidence interval method may work well with the parameters of the log-normal distribution in the presence of left-truncation and case-*k* interval censoring, however further study is required to verify the suitability of the proposed method.
The results from the simulation and coverage probability study is applicable with parameters of the log-logistic distribution in the presence of left-truncation and case-*k* interval censoring as this distribution shares similar hazard rate properties as the log-normal distribution.

ACKNOWLEDGEMENTS

We would like to extend our gratitude to Fundamental Research Grant Scheme (FRGS), VOT 5524226, University Putra Malaysian and Dr.Patricia Tai, University of Saskatchewan, Saskatoon, Canada.

REFERENCES

Balakrishnan, N. and Mitra, D. (2011), Likelihood inference for log-normal data with left truncation and right censoring with an illustration. *Journal of Statistical Planning and Inference*., **141(11)**: 3536– 3553.

Balakrishnan, N. and Mitra, D. (2012), Left truncated and right censored Weibull data and likelihood inference with an illustration. *Computational Statistics & Data Analysis*., **56(12)**: 4011– 4025.

Balakrishnan, N. and Mitra, D. (2014), Some further issues concerning likelihood inference for left truncated and right censored lognormal data. *Communications in Statistics - Simulation and Computation*., **43(2)**: 400– 416.

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(9)**: 1078– 1084.

Cox, D. R. and Oakes, D. (1984), Analysis of survival data. CRC Press.

Doganaksoy, N., and Schmee J. (1993), Comparison of approximate confidence interval distributions used in life-data analysis. *Technometrics*., **2**: 175– 184.

Guo, G. (1992), Event-history analysis for left-truncated data. *Sociological methodology*., **23**: 217 – 243.

Klein, J. P. and Moeschberger, M. L. (2003), Survival analysis: techniques for censored and truncated data. Springer.

Lawless, J . F. (1982). Statistical Model and Methods for Lifetime Data. Wiley: New York.

Lindsey, J. (1998), A study of interval censoring in parametric regression models. *Lifetime Data Analysis*., **4(4)**: 329– 354.

Manoharan, T. and Arasan, J. (2013), Assessing the performance of the log-normal distribution with left truncated survival data. *AIP Conference* Proceedings., **1557(1)**: 545– 549.

Manoharan, T., Arasan, J., Midi, H. and Adam, B. (2015), A coverage probability study 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)**: 127– 144.

Nardi, A. and Schemper, M. (2003), Comparing Cox and parametric models in clinical studies. *Statistics in medicine* ., **22(23)**: 3597– 3610.

Shen, P.-S. (2011), Estimation in the Cox proportional hazards model with doubly censored and truncated data. *Journal of Statistical Computation and Simulation*., **81(11)**: 1717– 1725.

Stovring, H. and Kristiansen, I. (2011), Simple parametric survival analysis with anonymized register data: A cohort study with truncated and interval censored event and censoring times. *BMC Research Notes*., **4(1)**: 1– 11.

Sun, J. (2007), The statistical analysis of interval-censored failure time data. Springer Science & Business Media.

Tai, P., Tonita, J., Yu, E., and Skarsgard, D. (2007), Disease-specific survival for limited-stage small-cell lung cancer affected by statistical method of assessment. *BMC Cancer*., **7(1)**: 31 – 39.

Tai, P., Yu, E., Vinh-Hung, V., Cserni, G. and Vlastos, G. (2004), Survival of patients with metastatic breast cancer: twenty-year data from two seer registries. *BMC cancer*., **4(1)**: 60 – 67.