Malaysian Journal of Mathematical Sciences 8(S): 37- 44 (2014)
Special Issue: International Conference on Mathematical Sciences and Statistics 2013 (ICMSS2013)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

Non-Mixture Cure Model for Interval Censored Data: Simulation Study

Fauzia Taweab, Noor Akma Ibrahim, Jayanthi Arasan and Mohd Rizam Abu Bakar

*Institute for Mathematical Research, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia*

E-mail: taweabf@yahoo.com.my

*Corresponding author

ABSTRACT

With the ongoing advance in the medical sciences, we may quite often encounter data sets where some patients have been cured from disease. Standard survival models are usually not appropriate for modeling such data because they simply do not take into account the possibility of cure. In this article, a non-mixture cure model is proposed based on lognormal distribution when the exact time of the event of disease is subject to interval censoring. The maximum likelihood estimation (MLE) method is implemented to estimate the parameters and a simulation study is conducted to assess the performance of the estimators under various conditions. The study results demonstrate that the bias, standard error, and root mean squared error values of the parameters estimates decrease with the increase in sample size and that the estimation method is more robust for data sets that have low censoring rates.

Keywords: Non-mixture cure model, interval censoring, maximum likelihood method, lognormal distribution.

1. INTRODUCTION

Due to advances in medical science, it is possible for a substantial proportion of patients not to experience the adverse event and these are thus considered as cured subjects. Two of the major means of modeling such data are the mixture cure model and the non-mixture cure model. The mixture cure model was first proposed by Boag (1949) to study cases where there a proportion of the patients receiving treatment for mouth cancer were cured. This model assumes that the entire population is composed of a mixture of the cured and uncured individuals. The overall survival function according to this model can be expressed as

$$S(t) = p + (1 - p)S_u(t) \tag{1}$$

where p is the probability of cure and $S_u(t)$ is the survival function for the uncured patients.

The mixture cure model has been investigated extensively by many researchers including Berkson *et al.* (1952), Kuk and Chen (1992), Peng and Dear (2000), Sposto (2002), and Kim and Jhun (2008), Shuangge (2010), peng and Xu (2011), among others. Although this model is common in survival data analysis, it might not fit some types of data, especially in cancer studies since it does not have the proportional hazards structure when the probability of cure is related to covariates and it does not closely describe the underlying biological process. Chen *et al.* (1999) developed an alternative useful model; the non-mixture cure model, for estimating the cure rate.

2. THE NON-MIXTURE CURE MODEL

This model was developed by Chen *et al* (1999) based on the assumption that the treatment leaves the patient with a number of cancer cells, N , which may grow slowly over time and produce a detectable recurrence of cancer. The number of these cancer cells is assumed to follow a Poisson distribution with a mean of θ.

Suppose that the i^{th} clonogen needs time, Z_i , to produce a cancer mass, then the recurrence of cancer can be defined by the random variable T such that $T = \min\{Z_i,\ 0 \le i \le N\}$, where Z_i are independent and identically distributed with $F(.)$. Then, according to the Poisson distribution and the distribution function of the time to detection of cancer relapse, the survival function for T is given by

$$\begin{aligned} S(t) &= P(no\, cancer\, by\, time\, t) \\ &= P(N = 0) + P(Z_1 > t, \ldots.., Z_N > t, N \ge 1) \\ &= \sum_{N=0}^{\infty} \frac{\theta^N e^{-\theta}}{N!} \left[1 - F(t)\right]^N \\ &= e^{-\theta F(t)} = p^{F(t)} \end{aligned} \tag{2}$$

where p is the probability of cure or the cure fraction which can be defined as

$$p = S(\infty) = \lim_{t \to \infty} e^{-\theta F(t)} = e^{-\theta} \tag{3}$$

Considering that censoring times are independent and non-informative, Weston and Thompson (2010) showed that the likelihood function for the model based on right censored data takes the form

$$L_i = \prod_{i=1}^{n} \left[-\log(p) f(t_i) \right]^{\delta_i} S(t_i) \tag{4}$$

where,

$$\delta_i = \begin{cases} 1 & if\ event\ at\ t_i, \\ 0 & if\ censored\ at\ t_i \end{cases}$$

The model can be further extended by using covariates X to model the probability of cure through θ. Moreover, a parametric model can be specified for the failure time. In this work, we consider that the mean of the cancer cells is related to covariates by $\theta = e^{X'\beta}$ and that the lognormal distribution for modeling the failure time of the uncured subjects,

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp[-\frac{(\ln t - \mu)^2}{2\sigma^2}]$$

$$F(t) = \Phi(\frac{\ln t - \mu}{\sigma}),$$

where Φ is the standard normal distribution function.

3. NON-MIXTURE CURE MODEL WITH INTERVAL CENSORED DATA

Interval censoring occurs if instead of observing the event time T only an interval $[L_i, R_i]$ is observed where $T_i \in [L_i, R_i]$, and $L_i \le R_i$. Here, L_i is the latest examination time before the event and R_i is the earliest examination time after the event. The subject is right censored if she/he has been eventfree at the last known time; $T_i \in [L_i, \infty)$. Based on these data, model (2) can be re-expressed as follows:

$$\begin{aligned} P(L_i \le T_i \le R_i) &= P(T_i \ge L_i) - P(T_i \ge R_i) \\ &= \exp(-e^{X'\beta} F(L_i^-)) - \exp(-e^{X'\beta} F(R_i^+)) \end{aligned} \qquad (5)$$

In the cure model, there are two possibilities for the i^{th} right subject $(R_i = \infty)$; the subject is either cured or she/he experiences the event of interest after the last examination time (uncured). Further details have been provided by Liu and Shen (2009). Then,

$$P(L_i \le T_i < \infty) = \exp(-e^{X'\beta} F(L_i^-))$$

Now, we reformat the censoring indicator δ as $\delta_i = I(R_i < \infty)$ for $T_i \in [L_i, R_i]$. Then, the log-likelihood function for the n observed interval data $(L_i, R_i, \delta_i, X_i),\ \ i = 1, 2, ..., n,$ can be written as:

$$L_c = \prod_{i=1}^{n} \left\{ \exp(-e^{X'\beta} F(L_i^-)) - \exp(-e^{X'\beta} F(R_i^+)) \right\}^{\delta_i} \times \left\{ \exp(-e^{X'\beta} F(L_i^-)) \right\}^{1-\delta_i} \qquad (6)$$

This can be simplified to

$$L_c = \prod_{i=1}^{n} \left\{ 1 - \exp\left[-e^{X'\beta} (F(R_i^+) - F(L_i^-)) \right] \right\}^{\delta_i} \times \exp(-e^{X'\beta} F(L_i^-)) \qquad (7)$$

Considering lognormal distribution, the log likelihood function of (7) is given by:

$$\begin{aligned} l = \sum_{i=1}^{n} \Bigg\{ & \delta_i \ln \left(1 - \exp\left[-e^{X'\beta} \left[\Phi\left(\frac{\ln R_i^+ - \mu}{\sigma} \right) - \Phi\left(\frac{\ln L_i^+ - \mu}{\sigma} \right) \right] \right] \right) \\ & - e^{X'\beta} \varphi\left(\frac{\ln L_i^+ - \mu}{\sigma} \right) \Bigg\} \end{aligned} \qquad (8)$$

4. SIMULATION STUDY AND RESULTS

A simulation study was conducted using 1000 samples each with sample sizes of 100, 200, and 300 for the interval-censored observations and one covariate X. The covariate values were generated from the Bernoulli distribution with 0.5. For a given X, a random variable was simulated from the Bernoulli distribution with $\exp[-\exp(X^{'}\beta)]$. If it is 1, then $T=\infty$ and if it is 0, then we generate the failure time values T from the lognormal distribution. The true value of β was chosen to be 0.01, corresponding to a cure rate of around 0.20. Two different values of the parameter σ were studied (0.7, 1) while the value of 1 was chosen as the true value of the parameter μ. The visiting or examination times were simulated independent of X and T following Goulin (2008), with different percentages of interval and right censoring as well. It was assumed that the number of visiting times is 10 visits for each subject and that the time between two visits has a uniform distribution on $(0,c)$, where c is a constant control censoring rate.

In each simulation, we assessed the bias, standard error (SE), and root mean square error (RMSE), $\sqrt{bias^{2}+SE^{2}}$, of the estimates and the results are collectively presented in Table 1.

In Table 1, we can see that the biases of estimates for β are very small even for the highest level of censoring and that these biases do not change much when the value of σ is changed from 0.7 to 1. On the other hand, the bias of the estimates for μ increase dramatically in both levels of censoring when σ is increased to 1. For the estimates of σ, the biases are small when $\sigma = 0.7$ and the censoring rate is low. A comparison of the values of the SE provides information on whether the estimates of the proposed method are underestimated or overestimated. Good agreement in SE between the parameters was obtained and the estimates for both the SE and RMSE decreased with sample size. With respect to the censoring rate, the performance of the suggested procedure is better under low levels of censoring than under heavy censoring.

TABLE 1: Bias, SE, and RMSE of the estimates for two censoring rates (moderate (20-30 per cent) and heavy (40-60 per cent))

			Censoring Rate			
		20-30			**40-60**	
	Bias	SE	RMSE	Bias	SE	RMSE
$n=100$						
$\sigma=0.7$						
β	0.0033	0.0083	0.0089	0.0049	0.0156	0.0164
σ	0.0636	0.1316	0.1461	0.0744	0.170	0.1852
μ	0.3640	0.2184	0.4245	0.4025	0.3882	0.5592
$\sigma=1$						
β	0.0036	0.0095	0.0103	0.0051	0.0147	0.0155
σ	0.1102	0.2517	0.2748	0.0896	0.2773	0.2914
μ	0.5466	0.3871	0.6697	0.5995	0.5522	0.8151
$n=200$						
$\sigma=0.7$						
β	0 .0024	0.0056	0.0061	0.0027	0.0103	0.1012
σ	0.0589	0.0904	0.1079	0.0616	0.1178	0.1329
μ	0.3352	0.1459	0.3655	0.3481	0.2523	0.4298
$\sigma=1$						
β	0.0033	0.0075	0.0082	0.0042	0.0104	0.0112
σ	0.1031	0.1749	0.2031	0.1155	0.1904	0.2227
μ	0.5209	0.2905	0.5964	0.5598	0.3751	0.6739
$n=300$						
$\sigma=0.7$						
β	0.0017	0.0041	0.0064	0.0025	0.0067	0.0073
σ	0.0514	0.0704	0.0872	0.0660	0.0881	0.1105
μ	0.3110	0.1057	0.3284	0.3438	0.1725	0.3846
$\sigma=1$						
β	0.0005	0.0048	0.0048	0.0035	0.0074	0.0082
σ	0.0893	0.1542	0.1782	0.1093	0.1515	0.1868
μ	0.5297	0.1780	0.2926	0.4660	0.2783	0.5983

5. CONCLUSION

In this paper, the maximum likelihood estimates for the parameters of the lognormal non-mixture cure model in presence of interval censored data were analyzed. It was found that the bias, SE, and RMSE of the estimates decrease when the sample size increases. Also, it was shown that the estimation method seems to provide consistent and better parameters estimates when the censoring rate is a bit low.

ACKNOWLEDGEMENTS

The authors gratefully acknowledged the financial support of Ministry of Higher Education Malaysia, and the institute for Mathematical Research, Universiti Putra Malaysia for their generous support of this study.

REFERENCES

Berkson, J and Gage, R. P. (1952). Survival curve for cancer patients following *treatment*. *Journal of the American Statistical*. **47**: 501–515

Boag, J.W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society.* Series B, **11**:15-44.

Chen, M. H, Ibrahim, J. G and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*. **94** (447): 909-919.

Claire, L. Weston and John, R, Thompson. (2010). Modeling survival in childhood cancer studies using two- stage non-mixture cure models. *Journal of Applied Statistics.* **37**(9): 1523–1535.

Hao Liu and Shen Yu. (2009). A semi parametric regression cure model for interval censored data. *J Am Stat Assoc.* **487**: 1168-1178.

Kim, Y. and Jhun, M. (2008). Cure rate model with interval censored data. *Statist Med*. **27**: 3-14.

Kuk, A.Y.C and Chen, C.H.P. (1992). A mixture Modeling Combining Logistic Regression with Proportional Hazard Regression. *Biometrika.***79**: 531-541.

Peng, Y. and Dear, K.B.G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics.* **56**: 237-243.

Sposto, R. (2002). Cure model analysis in cancer: An application to data from the children’s cancer group. *Statist. Med*. **21**: 293-312.

Shuangge, Ma. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica*. **20**: 1165-1181.

Yingwei Peng and Jianfeng Xu. (2011). An extended cure model and model selection. *Lifetime Data Anal*. DOI 10.1007/s10985-011-9213-1.

Zhao, Guolin, M. A. (2008). Nonparametric and *parametric survival analysis of censored data with possible violation of method assumptions.* PhD thesis, North Carolina University.