©Simposium Kebangsaan Sains Matematik Ke-23 (SKSM23)
Pulai Springs Resort, Johor Bahru
24 – 26 November 2015

A Comparative Study of Cure Models

Oh Yit Leng[a] and Zarina Mohd Khalid[b]

[a]Faculty of Business, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia
[b]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, UTM Skudai 81310, Johor, Malaysia

Abstract. In survival analysis, there are two types of model, parametric and nonparametric. For parametric models the survival data is described by a known non negative distribution. Exponential, weibull, log-normal and log-logistic distributions are the popular distributions used in survival analysis. Most of the time, distributions with two parameters are used as it allowed for more flexibility than one parameter distribution. There are cases where a fraction of individual who are not at risk in the event of interest. This fraction of individual is known as cure fraction. Survival models that take into account the existing of a cure fraction are called as cure model. Cure model separates the target population into two subgroups, long-term and short-term survivor. The survival time of the short-term survivor is described by a proper survival function, such as exponential, weibull, and log-normal survival functions. Weibull cure model is the most popular cure model used in survival analysis However, in some cases weibull distribution is not able to describe the survival data well. As an alternative distribution with two parameters Log-normal cure model is discussed in this study. Weibull cure model and log-normal cure models are compared in term of consistency. Survival data with different sample sizes and cure fractions are simulated. These data are then analyzed using the two cure models.

Keywords: Cure models, Cure fraction, Weibull distribution, Log-normal distribution

1. INTRODUCTION

Survival analysis is time to event analysis. The time is measured from a well-defined origin to the occurrence of an event of interest. There are two types of survival model which are parametric and nonparametric. Parametric survival model assumes the survival data follows a known non negative distribution. The common distributions used to describe the survival time are exponential, weibull, log-normal and log-logistic distributions while the most common nonparametric estimator used is the Kaplan-Meier estimation approach.

From time to time new survival models have been developed to model survival data with different characteristics.Cure models [1,2] have been used to model survival data with a cure fraction which will be discussed later, frailty model [3] take into account the effect of unobserved factors and frailty cure models [1,4] have been used to model survival data with a cure fraction and random effect. Beside that, [5] used a finite mixture survival model to characterize risk groups of neuroblastoma, while [6,7] used bivariate model for modelling family association in diseases.

In some cases, the survival curve has a positive asymptote to the right of the curve, it infers that there are some subjects who are not at risk in the event of interest. These subjects are considered cured from the event of interest and they will not experience the event of interest and this fraction of subjects is called as a cure fraction. These cases often happen in medical survival data, for example the neuroblastom cancer, and cancer of colon.A survival model which incorporating a cure fraction is more appropriate to be used in cases that involve a cure fraction. Cure models are survival model used in cases with a cure fraction as it allowed for a cure fraction. The popular cure models used in survival analysis are Weibull cure model, log-normal cure model and exponential cure model [8, 9, 10]. However, most of the researchers used Weibull cure model in their studies [11]. It is because weibull distribution has two parameters, this allowed more flexibility than distribution with one parameter and it can be used to model survival curves that drop off rapidly or decreasing with a slow speed. In some situations, weibull distribution is not suitable to describe the survival data. Therefore, in this study another commonly used distribution with two parameters which is log-normal distribution is discussed as alternative distribution of cases where weibull distribution does not fit the survival data well. In this research, weibull and log-normal cure models are compared in term of consistency. Different sample sizes and cure fractions are used to compare these two cure models.

2. CURE MODEL

Cure model divides the target population into two groups, long-term survivor and short-term survivor. Long-term survivor refers to cure fraction while short-term survivor refers to the at risk groups which has a proper survival function . Hence, an individual is either cured or expected to experience the event of interest following the survival function of at risk group. If *t* is the survival time, the general model of cure model is given by

$$S(t) = (1-p) + pS_0(t). \tag{1}$$

where *(1-p)* is the probability of cure, *p* is the probability of at risk, and $S_0(t)$ is the survival function of the in risk group.

Two cure models, Weibull and log-normal cure models are discussed in this research. The two cure models are compared in term of consistency for varying sample sizes and cure fractions.

Model 1: Weibull cure model. The Weibull survival function is given as

$$S_W(t) = e^{-\left(\frac{t}{a}\right)^b} \tag{2}$$

where$a > 0$ and $b > 0$. By joining equation (1) and equation (2)

$$S(t) = (1-p) + pe^{-\left(\frac{t}{a}\right)^b} \tag{3}$$

Model 2: Log-normal cure model. The log-normal survival function is given as

$$S_{LN}(t) = \frac{1}{2} - \frac{1}{2} erf\left(\frac{\ln(t) - \mu}{\sqrt{2}\sigma}\right) \tag{4}$$

where $-\infty < \mu < \infty$ and $\sigma \geq 0$. The error function erf, isa functionencountered in integrating the normal distribution and it can be expressed as

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{5}$$

From equation (1) and equation (4), the log-normal cure model is given as

$$S(t) = (1-p) + p\left[\frac{1}{2} - \frac{1}{2} erf\left(\frac{\ln(t) - \mu}{\sqrt{2}\sigma}\right)\right] \tag{6}$$

For each cure model, 8 groups of survival data with sample sizes of *n*=100, 200, 300,400 and 500, cure fractions of 1-*p*=0.2, 0.3, 0.4 and 0.5 were simulated. Each group consists of 100 sets of survival data.

3. MAXIMUM LIKELIHOOD

Maximum likelihood is the most common used approach to estimate the unknown parameters [12, 13]. In this research, maximum likelihood is utilized to estimate the unknown parameters involved. An observed data for individual *i* consist of true failure timex_i , potential censoring timec_i , survival time$t_i = \min(x_i, t_i)$, and censoring

indicator d_i. If $x_i \leq c_i$, then $d_i = 0$ and if $x_i > c_i$, then $d_i = 1$. The likelihood function for cure model (1) can be express as

$$L(t|\theta) = \prod_{i=1}^{n} [f(t_i)]^{1-d_i} [S(t\ )]^{d_i} \tag{7}$$

where $f(t)$ is the probability density function and $S(t)$ is the survival function.

Model 1: Weibull cure model likelihood function. The probability density function of the Weibull cure model can be expressed as

$$f(t_i) = p\frac{b}{a}\left(\frac{t}{a}\right)^{b-1} e^{-\left(\frac{t}{a}\right)^{b}} \tag{8}$$

Therefore, from equation (3), equation (7), and equation (8) the likelihood function of weibull cure model is

$$L(t|\theta) = \prod_{i=1}^{n} \left[p\frac{b}{a}\left(\frac{t}{a}\right)^{b-1} e^{-\left(\frac{t}{a}\right)^{b}} \right]^{1-d_i} \left[(1-p) + pe^{-\left(\frac{t}{a}\right)^{b}} \right]^{d_i} \tag{9}$$

Model 2: Log-normal cure model likelihood function. The probability density function of log-normal cure model is given as

$$f(t_i) = \frac{p}{t\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^{2}} \tag{10}$$

Therefore, from equation (6), equation (7) and equation (10) the likelihood function of log-normal cure model is given as

$$L(t|\theta) = \prod_{i=1}^{n} \left[\frac{p}{t\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^{2}} \right]^{1-d_i} \left[(1-p) + p\left(\frac{1}{2} - \frac{1}{2} erf\left(\frac{\ln(t)-\mu}{\sqrt{2}\sigma} \right) \right) \right]^{d_i} \tag{11}$$

The maximum likelihood estimates are obtained by maximizing the likelihood function. In this study, the newton - Raphson approach is used to determine the maximum likelihood estimates.

4. RESULT AND DISCUSSION

In this study, weibull cure model and log-normal cure model are compared in term of consistency. Survival data with different sample sizes and cure fractions are simulated and the two cure models are used to model these data. The results are presented in the following tables

TABLE 1. Weibull cure model with different cure fractions

	Average							
	Weibull survival		**Weibull cure**			**Mean Square Error**		
Cure Fraction	***a***	***b***	***a***	***b***	***p***	***a***	***b***	***p***
(1-*p*)=0.2	1647.5757	4.7082	1528.5432	4.8256	0.8226	16103.5293	0.1890	0.0013
(1-*p*)=0.3	1713.2389	4.9069	1539.7928	4.7855	0.7301	33049.4351	0.2937	0.0018
(1-*p*)=0.4	1771.8407	5.2399	1531.5676	4.7935	0.6263	60868.9912	0.5442	0.0011
(1-*p*)=0.5	1827.4039	5.9091	1538.6347	4.8399	0.5312	89023.3294	1.8192	0.0016

From Table 1, the mean square errors for *a* and *b* are increased when the cure fraction increasedfrom 0.2 to 0.5. However, the mean square error of the cure fractiondoesn't show any pattern of increase or decrease. This indicates that the performance of the Weibull cure model in estimating the cure fraction is not affected by the value of the cure fraction.The difference between the weibull survival function and the survival function of the short-term survivor of weibull cure model is getting bigger when the cure fraction changing from 0.2 to 0.5.

TABLE 2. Weibull cure model with different sample sizes

	Average							
	Weibull survival		**Weibull cure**			**Mean Square Error**		
Sample size	***a***	***b***	***a***	***b***	***p***	***a***	***b***	***p***
n=100 (1-*p*=0.8)	1647.5757	4.7082	1528.5432	4.8256	0.8226	16103.5293	0.1890	0.0013
n=200 (1-*p*=0.8)	1640.2372	4.6533	1529.9520	4.6780	0.8338	12866.3253	0.0501	0.0009
n=300 (1-*p*=0.8)	1646.6961	4.6868	1528.7152	4.7620	0.8240	14617.9215	0.0796	0.0009
n=400 (1-*p*=0.8)	1643.6873	4.6600	1529.7681	4.6984	0.8293	13466.3932	0.0411	0.0008
n=500 (1-*p*=0.8)	1644.6321	4.6486	1525.3873	4.7037	0.8235	14582.5227	0.0297	0.0007

From Table 2, the mean square errorof cure fraction is decreasing when sample size increased from 100 to 500. This show that the performance of weibull cure model in estimating the cure fraction is better for a larger sample size. This also infers that the Weibull survival function is closer to the survival function of the short-term survivor of Weibull cure model for a larger sample size.

TABLE 3. Log-normal cure model with different cure fractions

	Average							
	Log-normal survival		**Log-normal cure**			**Mean Square Error**		
Cure Fraction	μ	σ	μ	σ	***p***	μ	σ	***p***
(1-*p*)=0.2	5.6082	1.4828	5.0123	0.9880	0.8007	0.3663	0.2503	0.0000
(1-*p*)=0.3	5.9369	1.5950	5.0307	0.9904	0.6951	0.8357	0.3712	0.0000
(1-*p*)=0.4	6.1696	1.6542	4.9964	0.9963	0.6101	1.4056	0.4395	0.0000
(1-*p*)=0.5	6.4900	1.6448	5.0199	0.9871	0.5076	2.1876	0.4439	0.0000

In Table 3, it can be seen that the mean square errors of μ and σ are increasedwhen the cure fraction changed from 0.2 to 0.5. It means that the difference between log-normal survival function and the survival function of short-term survivor of log-normal cure model is increased while the cure fraction increased.

TABLE 4. Log-normal cure model with different sample sizes

	Average							
	Log-normal survival		**Log-normal cure**			**Mean Square Error**		
Sample Size	μ	σ	μ	σ	p	μ	σ	p
n=100 (1-*p*=0.8)	5.6082	1.4828	5.0123	0.9880	0.8007	0.3663	0.2503	0.0000
n=200 (1-*p*=0.8)	5.6171	1.4875	5.0209	0.9981	0.8003	0.3646	0.2433	0.0000
n=300 (1-*p*=0.8)	5.6106	1.4983	5.0062	1.0013	0.7984	0.3692	0.2493	0.0000
n=400 (1-*p*=0.8)	5.5931	1.4971	4.9971	1.0033	0.8018	0.3586	0.2454	0.0000
n=500 (1-*p*=0.8)	5.5961	1.4937	5.0044	1.0030	0.8028	0.3521	0.2417	0.0000

As shown in Table 4, the mean square errors of μ and σ for sample size 100 to 500 are closed to each other, it means that the performance of log-normal cure model is not affected by the sample size. In Table 3 and Table 4, it can be seen that the mean square errors of cure fraction *1-p* are equal to zero. It shows that the cure fraction estimated using log-normal cure model is close to its true value and the performance of the log-normal cure model in estimating the cure fraction is not affected by the sample size and the cure fraction.

5. CONCLUSION

The results of this study show that the performance of weibull cure model in estimating the cure fraction is better for a larger sample size. However, the performance of log-normal cure model is not affected by the sample size and its performance in estimating the cure fraction is better than the weibull cure model. It is because its mean square errors of the cure fraction for sample sizes from 100 to 500 and cure fractions 0.2, 0.3, 0.4, and 0.5 are all equal to zero. Beside that, this study also showsthe difference between survival function without cure fraction and the survival function of the short-term survivor of cure model isgetting bigger when cure fraction increased from 0.2 to 0.5. This indicates that the inclusion of a cure fraction is becoming essential for larger cure fraction.

REFERENCES

1. D. L. Price and A. K. Manatunga, Modelling Survival Data with a Cured Fraction using Frailty Models. *Statistics in Medicine,* 20(9-10), 1515-1527 (2001).
2. Y. Li, R. C. Tiwai, and S. Guha, Mixture Cure Survival Models with Dependent Censoring. *J. R. Statist. Soc. B.,* 69(3), 285-306 (2007).
3. P. Hougaard, Frailty Models for Survival Data, *Lifetime Data Analysis,* 1(3), 255-273 (1995).
4. B. Yu, A Frailty Mixture Cure Model with Application to Hospital Readmission Cata. *Biometrical Journal,* 50(3, 386-394 (2008).
5. S. Hunsberger, P. S. Albert and W. B. London, A Finite Mixture Survval Model to Characterize Risk Groups of Neuroblastoma, *Statistics in Medicine,* 28(8), 1301-1314 (2009).
6. A. Wienke, P. Lichtenstein, and A. I. Yashin, A Bivariate Frailty Model with a Cure Fraction for Modeling Familial Correlations in Diseases. *Biometrics,* 59, 1178-1183 (2003).
7. N. Chatterjee and J. Shih, A Bivariate Cure-Mixture Approach for Modeling Familial Association in Diseases. *Biometrics,* 57, 779-786 (2001).
8. R. D. Angelis, R. Capocaccia, T. Hakulinen, B. Soderman and A. Verdecchia, Mixture Models for Cancer Survval Analysis: Application to Population-Based Data with Covariates. *Statistics in Medicine,* 18, 441-454 (1999).
9. R. Sposto, Cure Model Analysis in Cancer: An Application to Data from the Children's Cancer Group. *Statistics in Medicine,* 21, 293-312 (2002).
10. P. C. Lambert, Modeling of the Cure Fraction in Survival Studies. *The Stata Journal,* 7(3), 351-375 (2007).
11. C. A. Liu and T. M. Braun, Parametric Non-Mixture Cure Models for Schedule Finding of Therapeutic Agents. *Journal of the Royal Statistical Society,* 58(2), 225-236 (2009).

12 R. C. Gupta, O. Akman and S. Lvin, A Study of Log-Logistic Model in Survival Analysis. *Biometrical Journal*, 41(4), 431-443 (1999.

13. Y. Peng and J. M. G. Taylor, Mixture Cure Model with Random Effects for the Analysis of a Multi-Center Tonsil Cancer Study. *Statistics in Medicine*, 30, 211-223 (2011).