Journal of
Applied Sciences

science
alert

ANSInet
an open access publisher
http://ansinet.com

An Analytical Approach on Cure Rate
Estimation Based on Uncensored Data

M.T. Uddin, M.N. Islam and Q.I.U. Ibrahim

Department of Statistics, Shahjalal University of Science and Technology, Sylhet, Bangladesh

Abstract: This study deals with the analysis of cure rate estimation under uncensored data. Cure models have been proposed for cure rate estimation. We have tried to estimate the parameter of the model using Maximum Likelihood Method (MLE). The analysis showed that the cure rate estimator converges to the true parameter when considered both cured and non-cured group. The analysis also showed that the estimating equation converges to the true equation of parameter when considered only non-cured group.

Key words: Uncensored data, maximum likelihood method, cure model, parametric model, survival function, cure rate

INTRODUCTION

In some clinical trials, a substantial proportion of patients who respond favorably to treatment subsequently appear to be free of any signs or symptoms of the diseases and may be considered as cured, while the remaining patients may eventually relapse. The objective of these clinical trials is to estimate the proportion of patients who are cured and the failure time distribution of proportion of patients who are not cured.

CURE MODEL

The survival models incorporating a cure fraction are called cure rate models. Now a days these models are becoming popular in analyzing data from cancer (or other diseases) clinical trials. The cure rate model has been widely used for modeling time to event data for various types of cancers for which a significant proportion of patients are cured. For example, Breast cancer, non-hodgkins lymphoma, leukemia, prostate cancer, melanoma, head and neck cancer etc. In fact cure model is a special case of frailty model, whether or not a patient is cured is an 'unobserved explanatory variable'. Because of the mixture nature of the patients in the clinical trials, the most popular type of cure model is the mixture model introduced by Berkson and Gage[1].

$$S(t) = \pi + (1 - \pi)S^*(t) \qquad (1)$$

In this model assume that T be a non-negative random variable denoting the failure time. A certain proportion π of population is 'cured'(cured patients) and the remaining $1-\pi$ are not cured (uncured patients). For this model, S(t) denote the survival function for the whole population (cured and uncured patients) and S*(t) denote the survival function for the uncured patients in the population.

Note that the survival function of the failure time distribution of cured patients is set equal to one for all finite values of the survival time because it is assumed that cured patients will never experience a relapse or death to the disease under investigation.

To estimate S*(t) in this model, we must specify the failure time distribution of uncured patients. The specification can be parametric or non parametric, which leads to parametric and semi-parametric cure model respectively. In parametric cure models, we assume a particular distribution for the failure time distribution of uncured patients. The most typical distributions that have been used for this purpose are: the Exponential distribution[2,3], the Weibull distribution[4,5], the Lognormal distribution[6], the Gompertz distribution[7], the Log-logistic distribution, the Gamma distribution, the Normal distribution, the Exponential power distribution, the Inverse Gaussian distribution and the Pareto distribution. More comprehensive distribution families such as the Extended Generalised Gamma (EGG) distribution and the Generalised F (GF) distribution[8] have also been proposed recently to accommodate other forms of failure time distributions for uncured patients. The model (1) is called Standard Cure Rate Model. It has been extensively used in the statistical literature by many researchers[2,4,5,7-18]. The

---

Corresponding Author: M.T. Uddin, Department of Statistics, Shahjalal University of Science and Technology, Sylhet, Bangladesh

model (1) is also called Mixture Cure Rate Model by several authors. In this model, the fraction of non-cured patients ultimately will experience recurrence or other treatment failure according to a statistical distribution function, F(t). Note that in model (1),

$$S^*(t) = 1 - F(t) \qquad (2)$$

And F(0) = 0, F(∞) = 1, so that $S^*(0) = 1, S^*(\infty) = 0$ and $S_1(\infty) = \pi$, the plateau value. The hazard function for this model is

$$h(t) = \frac{(1-\pi)f(t)}{S_1(t)} \qquad (3)$$

$$= \frac{f_1(t)}{S_1(t)} \qquad (4)$$

Where, f(t) is the density corresponding to F(t) and $f_1(t) = (1-\pi) f(t)$.

Originally, when treatments for cancer comprised surgery or radiation therapy alone administered over a short period, Mixture Cure Model has a practical interpretation[1,6] . Patients either were cured by treatment or they were not and the later would experience a recurrence after some time. Now a days with combined modality treatment that can last up to three years in some children cancers, this interpretation does not apply, since eradication of disease, if it occurs, can occurs at any time during treatment. The time to cure can not reliably be observed current technology. Hence even though mixture cure models often fit cancer data well, they usually can not be viewed literally as describing a mixture of cured and uncured patients. The literal interpretation of the cure model is meaningful also in some non-cancer applications[19]. Cure models were first proposed by Boag[6] and have since received regular attention in the statistical literature. However, they have not attained wide use or acceptance in the medical research literature, perhaps in part because of their reliance on particular parametric forms. Although the cure rate model appears to be attractive and is widely used, it has some drawbacks. Chen *et al.*[14] has been identified the following drawbacks of model (1).

Firstly, in the presence of covariates, it can not have a proportional hazard structure which is a desirable property for any survival model, because many asymptotic and computational results require a proportional hazard structure.

Secondly, when including covariates through the parameter π via a standard regression model, (1) yields improper posterior distributions for many types of non-informative improper priors, including the uniform prior for the regression coefficients. This is the most crucial drawback of model (1), because Bayesian inference with the model (1) essentially requires a proper prior.

Thirdly, the model (1) does not appear to describe the underlying biological process generating the failure time, at least in the context of cancer relapse, where cure rate model are frequently used.

However Chen[14] proposed different types of cure rate model, which overcomes the drawbacks just mentioned above. The model that they propose is quite attractive in several respects. The model is derived from a natural biological motivation. The proposed model has a mathematical relationship with the standard cure rate model. Specifically one can show that any standard cure rate model can be written as the proposed model and vice versa. However the proposed cure rate model can be written as:

$$S_p(t) = \exp(-\theta F(t)) \qquad (5)$$

The model (5) is not a proper survival function. Because $S_p(\infty) = \exp(-\theta)$. The model incorporates parameters bearing clear biological meaning. The model (5) is suitable for any type of failure data that has a surviving fraction. Thus the model can be useful for modeling various types of failure time data, including time to relapse, time to death, time to first infection and so forth. We also observe that the cure rate π is given by:

$$S_p(\infty) = \exp(-\theta) \qquad (6)$$

As $\theta \to \infty$, the cure rate tends to zero, where as $\theta \to 0$, the cure rate tends to 1.i.e., the cure rate lies between 0 and 1.
Note that by taking first derivative of (5), we get,

$$S_p'(t) = -\theta f(t) \exp(-\theta F(t))$$

$$\text{So, } -S_p'(t) = \theta f(t) \exp(-\theta F(t)$$

$$\text{Since, } -S_p'(t) = f_p(t)$$

Therefore, the density function corresponding to model (5) is given by:

$$f_p(t) = \theta f(t) \exp(-\theta F(t)) \qquad (7)$$

We observe that $S_P(t)$ is not a proper survival function, because $S_P(\infty) \neq 0$. Therefore, $f_p(t)$ is not a proper probability density function. But f(t) in model (7) is a proper density function. The hazard function is given by:

$$h_p(t) = \frac{f_p(t)}{S_p(t)} = \frac{\theta f(t) \exp(-\theta F(t))}{\exp(-\theta F(t))}$$

$$= \theta f(t) \qquad (8)$$

We observe that $h_p(t)$ is not a proper hazard function corresponding to a probability distribution, because $S_p(t)$ is not a proper survival function. The survival function for the 'non-cured' group is given by:

$$S*(t) = \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)} \qquad (9)$$

Note that, in (9), $S^*(0) = 1$ and $S^*(\infty) = 0$, so that $S^*(t)$ is a proper survival function. Thus the survival density corresponding to (9) is given by:

$$f^*(t) = -\frac{d}{dt} S^*(t) = (\frac{\exp(-\theta F(t))}{1 - \exp(-\theta)}) \theta f(t) \qquad (10)$$

Here, $f^*(t)$ is a proper density function. Since $S^*(t)$ is a proper survival function. The hazard function for the non-cured group is given by:

$$h^*(t) = \frac{f^*(t)}{S^*(t)} = (\frac{\exp(-\theta F(t))}{\exp(-\theta F(t)) - \exp(-\theta)}) h_p(t) \qquad (11)$$

There is an attractive mathematical relationship between the models in (1) and (5). From (9), we can write:

$$\exp(-\theta F(t)) = (1 - \exp(-\theta)) S^*(t) + \exp(-\theta) \qquad (12)$$

Using (12) in (5), we obtain the following model as:

$$S_p(t) = \exp(-\theta) + (1 - \exp(-\theta)) S^*(t) \qquad (13)$$

Where, $S^*(t)$ is given by (9). Therefore, $S_p(t)$ is a standard cure rate model with cure rate equal to $\pi = e^{-\theta}$ and the survival function for the non-cured population given by $S^*(t)$. This shows that every model defined by (5) can be written as a standard cure rate model. This result also means that every standard cure model corresponds to some model of the form (5) for some θ and distribution function $F(\cdot)$.

ESTIMATION AND RESULTS

Since the cure model is a special kind of survival model including cured portion and non-cured portion. Suppose that X be the life time of a patient. Then,

$P(X = \infty) = \lim_{t \to \infty} P(X > t) = e^{-\theta}$, which is considered as

cure rate. On the other hand, $P(X < t) = 1 - e^{-\theta}$, $0 \leq t < \infty$, which is the non-cured rate.

Since $X \in [0, \infty) \cup \{\text{"cured"}\}$ i.e., $X \in [0, \infty) \cup \{\text{"X} = \infty\text{"}\}$, hence the probability density function of x can be written as

$$f_\theta(t) = \theta f_0(t) e^{-\theta F_0(t)} 1_{(t < \infty)} + e^{-\theta} 1_{(t = \text{"cured"})}$$

with respect to the measure

$$\mu(A) = \int_{A \cap [0,\infty)} dt + 1_{(\text{"cured"} \in A)} \text{ for } A \subseteq [0, \infty) \cup \{\text{"cured"}\}$$

Thus $P\{X \in A\} = \int_{A \cap [0,\infty)} \theta f_0(t) e^{-\theta F_0(t)} dt + e^{-\theta} 1_{(\text{"cured"} \in A)}$

By using Maximum likelihood method we can estimate the parameter of the cure model.

For uncensored data, we consider the following cases:

Case–(a): $f_0(t)$, $F(t)$ are known, θ unknown and also both cured and non-cured observed. So, this case is fully parametric. Suppose that we have the data in the form (x_i, ϵ_I), $i = 1, 2, ..., n$. where x_i denotes the survival time for the ith patient. ϵ_I is the cured indicator with 1 if x_i is not cured and 0 otherwise. Also, $x \in [0, \infty) \cup \{\text{"cured"}\}$, $1 \leq i \leq n$.

Likelihood function: The likelihood function is given by:

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} (\theta f_0(x_i) e^{-\theta F_0(x_i)})^{\epsilon_i} (e^{-\theta})^{1 - \epsilon_i} \qquad (14)$$

where, $\epsilon_i = 1_{(X_i < \infty)}$ and $1 - \epsilon_i = 1_{(X_i = \text{"cured"})}$

Therefore, the MLE of θ is

$$\hat{\theta} = \frac{\frac{1}{n} \sum_{i=}^{n} \epsilon_i}{\frac{1}{n} \sum_{i=1}^{n} \epsilon_i F_0(x_i) + \frac{1}{n} \sum_{i=1}^{n} (1 - \epsilon_i)} \qquad (15)$$

Which is the required estimate of θ. Thus, the estimate of cure rate is $\pi = e^{-\theta}$. Now, we will show that $\hat{\theta}$ converges to the true parameter θ.

Convergence of $\hat{\theta}$: From the law of convergence, we can write,

$$\frac{1}{n} \sum_{i=1}^{n} \epsilon_i \xrightarrow{a.s} E(\epsilon) = P(X < \infty) = 1 - e^{-\theta} \qquad (16)$$

$$\frac{1}{n} \sum_{i=1}^{n} (1 - \epsilon_i) \xrightarrow{a.s} E(1 - \epsilon) = P(X > \infty) = e^{-\theta} \qquad (17)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\in_i F_0(x_i) \xrightarrow{a.s} E(\in F_o(x))$$

$$= E(1_{(X < \infty)} F_0(x))$$

$$= \int_0^\infty \theta f_0(t) F_0(t) e^{-\theta F_o(t)} dt$$

$$= \int_0^1 \theta z e^{-\theta z} dz \text{ [putting } F_0(t) = z]$$

$$= \int_0^1 \theta z e^{-\theta z} dz + \int_0^1 e^{-\theta z} dz$$

$$= -[z e^{-\theta z}]_0^1 - [\frac{1}{\theta} e^{-\theta z}]_0^1$$

$$= -e^{-\theta} - \frac{1}{\theta}(e^{-\theta} - 1)$$

$$= \frac{1 - e^{-\theta}}{\theta} - e^{-\theta} \qquad (18)$$

Using (16), (17) and (18) in (15), we obtain the following expression:

$$\hat{\theta} = \frac{1 - e^{-\theta}}{(\frac{1 - e^{-\theta}}{\theta}) - e^{-\theta} + e^{-\theta}} \to \theta$$

Therefore, $\hat{\theta} \to \theta$

Case (b): $f_0(t)$ and $F_0(t)$ are known, θ unknown and only non-cured are observed. So this case is also fully parametric. We also consider here only non-cured group i.e., we observe X if $X < \infty$. ($X_i \in [0, \infty)$. The p.d.f of X given $X < \infty$ can be written as:

$$f_\theta^{nc}(x) = \frac{\theta f_o(x) e^{-\theta F_0(x)}}{1 - e^{-\theta}}, X < \infty \qquad (19)$$

Likelihood function: The likelihood function is given by:

$$L^{nc}(\theta) = \prod_{i=1}^{n} f_\theta^{nc}(x)$$

$$= \prod_{i=1}^{n} \frac{\theta f_0(x_i) e^{-\theta F_0(x_i)}}{1 - e^{-\theta}}$$

Therefore, the estimate of θ can be obtained from the following estimating equation:

$$\frac{1}{\hat{\theta}} - \frac{1}{n}\sum_{i=1}^{n} F_0(x_i) - \frac{e^{-\hat{\theta}}}{1 - e^{-\hat{\theta}}} = 0 \qquad (20)$$

or

$$\frac{1}{\hat{\theta}} - \frac{e^{-\hat{\theta}}}{1 - e^{-\hat{\theta}}} = \frac{1}{n}\sum_{i=1}^{n} F_0(x_i) \qquad (21)$$

Comment: The Eq. 20 is an estimating equation of θ and this equation can not be solved analytically but it may be solved numerically. Therefore, the numerical solution of this equation is the required estimate of θ.

Convergence of the equation: From the law of convergence we can write:

$$\frac{1}{n}\sum_{i=1}^{n} F_0(x_i) \xrightarrow{a.s} E\langle F_0(x)|x < \infty\rangle$$

$$= \frac{\int_0^\infty \theta F_0(t) f_0(t) e^{-\theta F_0(t)} dt}{P(x < \infty)} \qquad (22)$$

Using (16) and (18) in (22), we obtain the following expression as:

$$\frac{1}{n}\sum_{i=1}^{n} F_0(x_i) = \frac{\frac{1 - e^{-\theta}}{\theta} - e^{-\theta}}{1 - e^{-\theta}}$$

$$= \frac{1}{\theta} - \frac{e^{-\theta}}{1 - e^{-\theta}} \qquad (23)$$

Therefore,

$$\frac{1}{\hat{\theta}} - \frac{e^{-\hat{\theta}}}{1 - e^{-\hat{\theta}}} = \frac{1}{n}\sum_{i=1}^{n} F_0(x_i) \xrightarrow{a.s} \frac{1}{\theta} - \frac{e^{-\theta}}{1 - e^{-\theta}}$$

And hence,

$$\frac{1}{\hat{\theta}} - \frac{e^{-\hat{\theta}}}{1 - e^{-\hat{\theta}}} \xrightarrow{a.s} \frac{1}{\theta} - \frac{e^{-\theta}}{1 - e^{-\theta}} \qquad (24)$$

CONCLUSIONS AND FURTHER RESEARCH

In this study, we are tried to find an estimator to estimate the cure rate by considering uncensored data. In uncensored data, eventually we have considered several cases. When we have assumed $f_0(\cdot)$ and $F_0(\cdot)$ are known and also assumed non-cured and cured group, we have found an analytic solution for the cure rate parameter θ. That is, we have found an estimator of θ which converges to the true value of the parameter. On the other hand, when we assume only non-cured group, we could not find an analytic solution of θ but we have found an estimating equation for θ which might be solved numerically and the numerical solution of the estimating equation would be the estimate of the parameter. Also in that situation we have found that our estimating equation converges to the true equation of parameter.

For further research one could perform the simulation study of the cure rate model based on uncensored data. The fact that one could apply numerical method in order to get estimate of the parameter of cure model.

REFERENCES

1. Berkson, J., R.P. Gage, 1952. Survival curves for cancer patients following treatments. J. Am. Stat. Assoc., 47: 501-515.
2. Goldman, A.I., 1984. Survivorship analysis when cure is a possibility: A Monte-Corlo study. Statistics in Medicine, 3: 153-163.
3. Ghitany, M.E., R.A. Maller and S. Zhou, 1994. Exponential mixture models with long term survivors and covariates. J. Mutlivariate Analysis, 49: 218-241.
4. Farewell, V.T., 1982. The use of mixture models for the analysis of survival data with long term survivors. Biometrics, 38: 1041-1046.
5. Farewell, V.T., 1986. Mixture model in survival analysis: Are they worth the risk? Can. J. Stat., 14: 257-262.
6. Boag, J.W., 1949. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J. Royal Stat. Soc., Series B, 11: 15-44.
7. Cantor, A.B. and J.J. Shuster, 1992. Parametric versus non-parametric methods for estimation cure rates bases on censored survival data. Statistics in Medicine, 11: 931-937.
8. Peng, Y., K.B. Dear and J.W. Denham, 1998. A generalized F mixture model for cure rate estimation. Statistics in Medicine, 17: 813-830.
9. Halpern, J. and B.W. Brown, 1987. Cure rate models: Power of the log rank and generalized wilcoxon tests. Statistics in Medicine, 6: 483-489.
10. Kuk, A.Y.C. and C.H. Chen, 1992. A mixture model combining logistic regression with proportional hazard regression. Biometrika, 79: 531-541.
11. Laska, E.M. and M.J. Meisner, 1992. Non-parametric estimation and testing in a cure model. Biometrics, 48: 1223-1234.
12. Sposto, R., H.N. Sather and S.A. Baker, 1992. A comparison of tests of the difference in the proportion of patients who are cured. Biometrics, 48: 87-99.
13. Peng, Y. and K.C. Carriere, 2002. An empirical comparison of parametric and semi parametric cure model. Biometrical J., 44: 1002-1014.
14. Chen, M.H., J.G. Ibrahim and D. Sinha, 1999. A new Bayesian model for surviving data with a surviving fraction. J. Am. Stat. Assoc., 94: 909-919.
15. Peng, Y., 2003. Fitting semi-parametric cure models. Computational Statistics and Data Analysis, 41: 481-490.
16. Yakovlev, A.Y., 1994. Letter to the editor. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. Statistics in Medicine, 13: 983-985.
17. Sposto, R., 2002. Cure model analysis in cancer: An application to data from the children's cancer group. Statistics in Medicine, 21: 293-312.
18. Tsodikov, A., 2001. Estimation of Survival based on proportional hazards when cure is a possibility. Mathematical and Computer Modeling, 33: 1227-1236.
19. Hauck, W.W., L.T. Mckee and B.J. Turner, 1997. Two-part survival models applied to administrative data for determining rate and predictors for maternal-child transmission of HIV. Statistics in Medicine, 16: 1683-1694.