



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Cure Rate Models: A Review of Recent Progress with a Study of Change-point Cure Models when Cured is Partially Known

^{1,3}Fauzia Taweab and ^{1,2}Noor Akma Ibrahim

¹Laboratory of Applied and Computational Statistics, Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia

²Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Malaysia

³Department of Statistics, Faculty of Science, University of Tripoli, Libya

Abstract: In medicine and public health researches, survival cure models are widely used to analyse time-to-event data in which some subjects are reasonably believed to be medically cured. In general, there are two types of models for estimation of the cure fraction. The first one is the Mixture Cure Model (MCM), which was developed by Boag in 1949. This type of models assumes that the whole population is composed of susceptible subjects and cured subjects. The second cure model type was proposed by Yakovlev *et al.* (1993) based on the assumption that the treatment leaves the patient with a number of cancer cells, which may grow slowly over time and produce a detectable recurrence of cancer. It is known as the Non-Mixture Cure Model (NMCM). These two models are related and the NMCM can be transformed into the MCM, when the cure fraction specially specified. Different parametric and semi-parametric estimation methods for model parameters in both types have been proposed and many applications of these models have been reported. The extensions of the cure model focus on study of change point effects on the cure or hazard rate. A change point cure model is proposed when cured is partially known.

Key words: Cure rate models, mixture cure model, non-mixture cure model, change point models

INTRODUCTION

A common assumption in survival data analysis is that all of the study subjects will eventually experience the event of interest if they are followed long enough. However, in reality, the event may not occur with some subjects, even after a very long period of time. For example, in clinical trials, there exist a proportion of subjects who will not experience such an event. In this case, the patients are not censored in the traditional sense and are hence confidently assumed to be cured. Therefore, traditional survival models such as the Cox proportional hazard model or the accelerated failure time are not appropriate for such cured subjects. Consequently, cure models have been developed for manipulation and analysing survival data with cure fraction.

Cure fraction models basically focus on the proportion of patients who survive long-term following disease. Additionally, these models focus on the probability of survival of the uncured patients up to a given point in time. The most widely-used cure model is the Mixture Cure Model (MCM) which is also known as the standard cure rate. This model was introduced by

Boag (1949) to study cases where there was a proportion of cured patients among those who had been receiving treatment for mouth cancer. He proposed lognormal normal distribution to model the failure time of the susceptible group and assumed the cure probability to be constant. This model was further developed three years later by Berkson and Gage (1952) and later studied extensively by several authors, e.g., Farewell (1986), Goldman (1984), Kuk and Chen (1992), Maller and Zhou (1996), Taylor (1995), Peng and Dear (2000) and Banerjee and Carlin (2004), among many others.

The other cure model, the Non-Mixture Cure Model (NMCM), was first proposed by Yakovlev *et al.* (1993) and was further discussed by Chen *et al.* (1999), Ibrahim *et al.* (2001), Chen *et al.* (2002) and Tsodikov (2002). This model was motivated by the underlying biological mechanism and developed based on assumption that the number of cancer cells that remain active after cancer treatment follow Poisson distribution (Yakovlev *et al.*, 1993; Chen *et al.*, 1999; Gutierrez, 2002; Uddin *et al.*, 2006a).

Both cure fraction models have been actively compared (Sposito *et al.*, 1992; Broet *et al.*, 2001; Achcar *et al.*, 2012). Several extensions of these models

have also been investigated. For example, Yin and Ibrahim (2005) proposed a general class of cure models based on Box-Cox transformation of the population survival function. Later to this, a flexible class of cure rate models under latent activation schemes was developed by Cooner *et al.* (2007). Castro *et al.* (2010) described an application of the mixture and non-mixture models within location, scale and shape (GAMLSS) framework to the fitting of long-term survival models. The objective of the present paper is to provide a quick review of cure models and to propose a parametric approach that analyses the change-point cure model when cured is partially known.

MIXTURE CURE MODEL (MCM)

For decades, the MCM has been a popular method in analysing time-to-event data in which some subjects are reasonably believed to be cured. This model assumes that a proportion, η , of the subjects will be cured and that these subjects are not at risk of experiencing the re-occurrence of the event. The other proportion $(1-\eta)$ is for the individuals who are expected to experience the event in some future time eventually. The MCM can be derived as following: Suppose that T denotes the occurrence time of a disease with population survival function $S(t)$ and that v expresses a binary random variable taking the values 1 and 0 with probability $(1-\eta)$ (event rate) and η (cure rate), respectively, where $\eta = P_r(T = \infty)$. Furthermore, let $S_u(t)$ and $f_u(t)$ be the survival and density function for uncured group. So, the population survival function $S(t)$ can be represented as:

$$S(t) = \eta + [1-\eta]S_u(t) \quad (1)$$

and the density function corresponding to Eq. 1 is $f(t) = (1-\eta)f_u(t)$.

Different parametric distributions have been used to model the density function $f_u(t)$, including Exponential distribution (Ghitany *et al.*, 1994), Weibull distribution (Farewell, 1986), Log normal (Boag, 1949; Gamel *et al.*, 1990). Nonparametric approaches for $f_u(t)$ have also been considered in the literature (Taylor, 1995; Kuk and Chen, 1992; Sy and Taylor, 2000; Peng and Dear, 2000).

A generalized link function of logistic distribution was suggested to access the effects of covariates for the cured probability with using parametric functions for fitting $S_u(t)$. Farewell (1982) used logistic regression for modelling the cured probability and a Weibull distribution

for fitting the survival function for uncured subjects. Maller and Zhou (1996) provided a comprehensive treatment to the cure model based on different parametric failure time regression models and they also investigated one-sample nonparametric failure time models. Recently, Zhang and Peng (2009) who considered separate modeling of the covariate effects on the cure probability and the distribution of the failure time of uncured subjects. However, unlike the traditional mixture cure models, the model of Zhang and Peng (2009) allows the effects of covariates on the failure time distribution of the uncured patients to be negligible at time zero and to grow with time. Such model is particularly useful in certain cancer treatments when the effects of treatment gradually increase from zero.

Various non-parametric and semi-parametric methods have been proposed to estimate the survival function when the distribution of failure time is not specified. Tsodikov *et al.* (2003) provided a useful summary of the work of Maller and Zhou (1996) who employed a nonparametric approach on a homogenous sample. Kuk and Chen (1992) proposed semi-parametric cure model in which the failure times were estimated using the proportional hazards regression models while the cure probability was determined by the logistic regression model. They developed a Monte Carlo approximation algorithm to estimate the model parameters. The proportional hazards cure model was further studied by Peng and Dear (2000), Sy and Taylor (2000), Lu and Ying (2004), Lam *et al.* (2005) and Corbiere *et al.* (2009), among others, in order to develop alternative methods for computing the joint parametric-nonparametric likelihood function. Their approaches are largely based on the semi-parametric expectation-maximization (EM) algorithm which computes estimates for both the parametric and the nonparametric components.

More studies have been conducted based on the MCM. Examples include Kim and Jhun (2008) who proposed the MCM for interval-censored data. They developed the likelihood function based on an approximate approach suggested by Goetghebeur and Ryan (2000). In addition, they introduced a frailty model to characterize the association between the cure probability and failure time. In another example, Kim *et al.* (2009) proposed a new MCM via latent cure rate markers for survival data with a cure fraction. The latent cure rate markers were modeled via a multinomial logistic regression and patients who share the same cure rate were classified into the same risk group. Seppa *et al.* (2010) applied a mixture cure fraction model with random effects

to cause-specific survival data of female breast cancer patients. They used two sets of random effects to capture the regional variation in the cure fraction and in the survival of the uncured patients. Peng and Taylor (2010) studied the MCM with random effects and proposed several estimation methods based on Gaussian quadrature, rejection sampling and importance sampling to obtain the maximum likelihood estimates of the model for clustered survival data with a cure fraction. Ma (2010) presented a semi-parametric cure rate model for mixed case interval-censored data, in which a generalized linear model was used to describe the probability of cure and a Cox model was employed to estimate the failure time of uncured subjects.

NON-MIXTURE CURE MODEL (NMCM)

The second type of cure models is the NMCM developed by Yakovlev *et al.* (1993) as an alternative to the MCM. The NMCM assumes that after treatment a patient is left with N cancer cells that may grow rapidly and produce a detectable cancer disease later on. The number of cancer cells is assumed to have a Poisson distribution with mean λ . Let $Z_i, i = 1, 2, \dots, N$ denote the random time for the i th cancer cell to produce a detectable cancer mass, where, Z_i are assumed to be independently and identically distributed with $F(t) = 1 - S(t)$. Then, the time to relapse of cancer can be defined by the random variable $T = \min \{Z_i = 1, 2, \dots, N\}$. The survival function for the population is given by:

$$\begin{aligned} S(t) &= P[\text{No cancer by time } t] \\ &= P[N=0] + P[Z_1 > t, Z_2 > t, \dots, Z_N > t, N \geq 1] \\ &= \exp(-\lambda) + \sum_{N=1}^{\infty} (s(t))^N \left[\frac{\exp(-\lambda)\lambda^N}{N!} \right] \\ &= \exp(-\lambda F(t)) \end{aligned} \tag{2}$$

The cure fraction in this model is defined as:

$$\eta = \lim_{t \rightarrow \infty} S_i(t) = P(N=0) = \exp(-\lambda)$$

The NMCM has been investigated by many researchers (Tsodikov, 1998; Chen *et al.*, 1999; Ibrahim *et al.*, 2001; Yin, 2005; Zeng *et al.*, 2006). Most of the existing studies on the NMCM are in the Bayesian context due to its special form. The parametric and semi-parametric approaches of estimation have been discussed by Ibrahim *et al.* (2001). Tsodikov *et al.* (2003) provided a nice review of these non-mixture cure modelling techniques in cure rate estimation and associated statistical problems. They have highlighted the

following three distinct advantages of the NMCM over the MCM. First, the NMCM has proportional hazard structure when the covariates are involved through the cure rate parameter. Second, the derivation of the NMCM is linked closely to the underlying biological process and it is hence expected to present a much more meaningful interpretation of the results for the data analysis in some studies. Third, the NMCM is very attractive in computations since it has a simple structure for the survival function which can provide a naturally technical structure of maximum likelihood estimation procedures.

With rapid developments in medical and health sciences the NMCM has been widely studied and applied in medical research. For example, Herring and Ibrahim (2002) introduced a parametric estimation approach for random effects NMCM. They also proposed a methodology to account for non-ignorable missing covariates in this class of models. Brown and Ibrahim (2003) extended the NMCM to include longitudinal covariates. Uddin *et al.* (2006a, b) proposed two approaches; non-parametric and parametric, to cure rate estimation based on the NMCM and uncensored data. Liu and Shen (2009) introduced a semi-parametric NMCM for the analysis of interval-censored data. They developed semi-parametric maximum likelihood estimation for the model using the EM method. Lopesa and Bolfarine (2012) investigated the NMCM with random effects. The estimation was carried out by classical and Bayesian methods.

The majority of the statistical literature presumes that the number of cancer cells follows a Poisson distribution with mean λ (Chen *et al.*, 1999). Other distributions that may fit the number of cancer cells such as Bernoulli and negative binomial distributions have been suggested by Rodrigues *et al.* (2009).

CHANGE-POINT PROBLEM

Change point problems arise in many survival data analysis, for example, in clinical trials, one might suspect that undesirable side effects may cause different hazard rate after a threshold time. Another example, if a researcher wants to study the effectiveness of new treatment on the failure time rate, a change-point model may be useful in the analysis for an interest model; where it is often reasonable to assume that the effect of the treatment is not immediate and that the treatment may affect the risk of failure only after a lag time.

Several researchers have introduced change-point models into the field of survival analysis, beginning with

Matthews and Farewell (1982). These researchers have assumed that the hazard function is constant, with the exception of a jump. Muller and Wang (1990) proposed a non-parametric method for the estimation of the changes in the hazard rate. Sen (1993) and Pons (2003) each considered a Cox model with change-point in accordance with an unknown threshold of a covariate. However, these studies have mainly focused on modelling hazard function and do not consider a proportion of cure.

Cure fraction models may well exist in change-point scenarios. For example, in assessing the cure probability for a patient under a certain treatment depending on a subject's biomarker, one suspects that the treatment works more or less effectively after a threshold of patient's biomarker value (Ma, 2011). Recently, only two studies focused on change-point problems with survival cure data. Zhao *et al.* (2009) proposed a mixture cure model with a change-point at an unknown threshold of failure time. Othman *et al.* (2012) investigated the MCM based on a covariate threshold.

CURE RATE MODEL WITH A CHANGE-POINT

Let (t_i, δ_i, x_i) denote the observed data for the i th subject $i = 1, 2, \dots, n$, where t_i is the observed survival time of T for the i th individual, δ_i is a censoring indicator with $\delta_i = 1$ for uncensored t_i and $\delta_i = 0$ for censored t_i and x_i is the observed value of x for the i th individual. The covariate X is assumed to have a change point effect and at this point the hazard value or cure fraction takes a sudden jump or fall. Let τ refer to the change-point for X . If $X \leq \tau$, let $\eta(X) = p_1$ and $\mu(X) = \mu_1$. However, if $X > \tau$, then $\eta(X) = p_2$ and $\mu(X) = \mu_2$. The value of v for the i th patient is denoted as $v_i = 1$ if the patient is not cured and $v_i = 0$ if he/she is cured. Obviously for a censored subject ($\delta_i = 0$), v is a latent variable and its value is not observable and it can be either one or zero. Therefore, the EM algorithm is used to perform Maximum Likelihood Estimation (MLE) for the model parameters. Given δ_i and v_i , the complete likelihood function may be written as:

$$L_c(\theta) = \prod_{i=1}^n \left\{ \left[(1-p_1) f_{u1}(t_i) \right]^{\delta_i} \left[(p_1)^{1-v_i} \left((1-p_1) S_{u1}(t_i) \right)^{v_i} \right]^{1-\delta_i} \right\}^{I(X_i \leq \tau)} \times \left\{ \left[(1-p_2) f_{u2}(t_i) \right]^{\delta_i} \left[(p_2)^{1-v_i} \left((1-p_2) S_{u2}(t_i) \right)^{v_i} \right]^{1-\delta_i} \right\}^{I(X_i > \tau)} \quad (3)$$

Here, the lognormal distribution is considered for modelling failure time of uncured subject, with the density function as:

$$f_u(t_i) = \left(\frac{1}{t_i \sigma \sqrt{2\pi}} \right) \exp \left(-\frac{1}{2} \left(\frac{\ln t_i - \mu(X)}{\sigma} \right)^2 \right)$$

and the survival function:

$$S_u(t_i) = 1 - \Phi \left(\frac{\ln t_i - \mu(X)}{\sigma} \right)$$

where, $\Phi(\cdot)$ is the distribution function of the standard normal. Then, the function (3) would be rewritten as:

$$\prod_{i=1}^n \left\{ \left[\left(\frac{1-p_1}{t_i \sigma_1 \sqrt{2\pi}} \right) \exp \left(-\frac{1}{2} \left(\frac{\ln t_i - \mu_1}{\sigma_1} \right)^2 \right) \right]^{\delta_i} \left[(p_1)^{1-v_i} \left((1-p_1) \left[1 - \Phi \left(\frac{\ln t_i - \mu_1}{\sigma_1} \right) \right] \right)^{v_i} \right]^{1-\delta_i} \right\}^{I(X_i \leq \tau)} \times \left\{ \left[\left(\frac{1-p_2}{t_i \sigma_2 \sqrt{2\pi}} \right) \exp \left(-\frac{1}{2} \left(\frac{\ln t_i - \mu_2}{\sigma_2} \right)^2 \right) \right]^{\delta_i} \left[(p_2)^{1-v_i} \left((1-p_2) \left[1 - \Phi \left(\frac{\ln t_i - \mu_2}{\sigma_2} \right) \right] \right)^{v_i} \right]^{1-\delta_i} \right\}^{I(X_i > \tau)} \quad (4)$$

Before the implementation of the EM algorithm, we need to address the problem of non-smoothness of the likelihood function with respect to the unknown change-point parameter τ . Specifically, the indicator function $I(X \leq \tau)$ is not differentiable with τ and consequently, standard Taylor series methods cannot be used. To circumvent this issue, define a continuous function $K(\cdot)$ which satisfies $\lim_{u \rightarrow -\infty} K(u) = 0$ and $\lim_{u \rightarrow \infty} K(u) = 1$. Let:

$$K_n(u) = K \left(\frac{u}{h_n} \right)$$

where, h_n is a small positive constant that depends on the sample size in which $\lim_{n \rightarrow \infty} h_n = 0$. A common and useful choice for this class of functions is the logistic function where:

$$K_n(u) = \frac{\exp \left[\frac{u}{h_n} \right]}{1 + \exp \left[\frac{u}{h_n} \right]}$$

Then, to estimate the change point as well as the model parameters $p_1, p_2, \mu_1, \mu_2, \sigma_1, \sigma_2$ the smoothed likelihood function denoted by $L_c(\theta)$ is proposed:

$$L_c(\theta) = \prod_{i=1}^n \left\{ \left[\left(\frac{1-p_1}{t_i \sigma_1 \sqrt{2\pi}} \right) \exp \left(-\frac{1}{2} \left(\frac{\ln t_i - \mu_1}{\sigma_1} \right)^2 \right) \right]^{\delta_i} \left[(p_1)^{1-v_i} \left((1-p_1) \left[1 - \Phi \left(\frac{\ln t_i - \mu_1}{\sigma_1} \right) \right] \right)^{v_i} \right]^{1-\delta_i} \right\}^{K_n(X_i - \tau)} \times \left\{ \left[\left(\frac{1-p_2}{t_i \sigma_2 \sqrt{2\pi}} \right) \exp \left(-\frac{1}{2} \left(\frac{\ln t_i - \mu_2}{\sigma_2} \right)^2 \right) \right]^{\delta_i} \left[(p_2)^{1-v_i} \left((1-p_2) \left[1 - \Phi \left(\frac{\ln t_i - \mu_2}{\sigma_2} \right) \right] \right)^{v_i} \right]^{1-\delta_i} \right\}^{1-K_n(X_i - \tau)} \quad (5)$$

The likelihood function (5) is differentiable with respect to the unknown parameters θ and the EM algorithm can be employed to obtain the MLEs of the parameters.

The expectation (E) step of the EM algorithm computes $E [l_c(\theta) | \theta^m]$, the conditional expectation of the

log likelihood function (5) with respect to v_i , given the current estimates $\theta^{(m)} = (p_1^{(m)}, \mu_1^{(m)}, \sigma_1^{(m)}, \tau^{(m)}, p_2^{(m)}, \mu_2^{(m)}, \sigma_2^{(m)})$:

$$E[l_i(\theta) | \theta^{(m)}] = \sum_{i=1}^n [K_n(\tau - X_i) l_1(\theta, W_i) + 1(1 - K_n(\tau - X_i)) l_2(\theta, W_i)] \quad (6)$$

where, $W_i = (t_i, x_i, \delta_i, v_i)$:

$$l_1(\theta) = \sum_{i=1}^n \left\{ [\delta_i + g_i^{(m)}(1 - \delta_i)] \log(1 - p_j) + (1 - \delta_i)(1 - g_i^{(m)}) \log(p_j) \right\} + \sum_{i=1}^n \left\{ -\delta_i \left[\log(t_i \sigma_j \sqrt{2\pi}) + \frac{(\ln t_i - \mu_j)^2}{2\sigma_j^2} \right] + (1 - \delta_i) g_i^{(m)} \log \left(1 - \Phi \left(\frac{\ln t_i - \mu_j}{\sigma_j} \right) \right) \right\}$$

for $j = 1, 2$ and $g_i^{(m)}$ is the expectation of v_i on the current estimates of θ , given by:

$$g_i = E(v_i | \theta^{(m)}) = \delta_i + (1 - \delta_i) \left[\frac{(1 - p^{(m)}) S_u(t_i | \theta^{(m)})}{p^{(m)} + (1 - p^{(m)}) S_u(t_i | \theta^{(m)})} \right] \quad (7)$$

The M-step of the EM algorithm maximizes $E[l_i(\theta) | \theta^{(m)}]$ with respect to the unknown parameters θ for fixed g_i . Maximization in this step can be carried out using the Newton-Raphson algorithm.

Simulation studies: To investigate the performance of the proposed estimation method, simulation studies were conducted. Here, we considered two different choices of the smoothing function; logistic function and the cumulative distribution function of the standard normal distribution. Under each choice, two simulations scenarios were considered. The first scenario uses a Uniform (0,1) random variable with a change-point at 0.5. The second uses a truncated Normal (tN(1, 1, 0, 2)) random variable with a change-point at 1. The cure indicator is generated using uniform distribution to determine whether someone is cured. The event time (if not cured) is generated from lognormal model considering the change point. Censoring times follow a lognormal distribution with (2, 0.15), which induces a censoring rate of about 35%.

The studies showed the bias, standard error and mean square error of the parameter estimate under different sample sizes 200, 400 and 800.

The simulation results based on 500 replicates are summarized in Tables 1a, b and 2a, b. The results demonstrate that the proposed method has a good performance. The biases of estimates are small. The estimation of cure probability and change-point are quite accurate throughout all settings. The estimates for the standard error, as well as the mean squared error, decrease

Table 1a: Simulation results for the standard normal smoothed function
X~Uniform (0,1)

Parameters	θ_0	Est	Bias	SE	MSE×1000
N = 200					
P_1	0.4	0.389	0.011	0.048	2.425
μ_1	0.3	0.312	-0.012	0.018	0.468
σ_1	0.15	0.159	-0.009	0.014	0.277
P_2	0.3	0.314	-0.014	0.039	1.717
μ_2	0.4	0.386	0.014	0.021	0.637
σ_2	0.2	0.195	0.005	0.015	0.250
τ	0.5	0.459	0.041	0.077	7.610
N = 400					
P_1	0.4	0.392	0.008	0.032	1.088
μ_1	0.3	0.309	-0.009	0.013	0.250
σ_1	0.15	0.156	-0.006	0.010	0.136
P_2	0.3	0.313	-0.013	0.030	1.069
μ_2	0.4	0.388	0.012	0.016	0.400
σ_2	0.2	0.196	0.004	0.011	0.137
τ	0.5	0.460	0.040	0.055	4.625
N = 800					
P_1	0.4	0.395	0.005	0.023	0.554
μ_1	0.3	0.307	-0.007	0.009	0.130
σ_1	0.15	0.155	-0.005	0.007	0.074
P_2	0.3	0.311	-0.011	0.021	0.562
μ_2	0.4	0.391	0.009	0.011	0.202
σ_2	0.2	0.197	0.003	0.008	0.073
τ	0.5	0.466	0.034	0.047	3.365

θ_0 : True values of parameters, Est is the mean of estimates, SE is the mean of standard errors, MSE is the mean square errors for the MLE estimators

Table 1b: Simulation results for the standard normal smoothed function.
X~tN(1,1,0,2)

Parameters	θ_0	Est	Bias	SE	MSE×1000
N = 200					
P_1	0.4	0.394	0.006	0.049	20.437
μ_1	0.3	0.305	-0.005	0.020	00.425
σ_1	0.15	0.153	-0.003	0.015	00.234
P_2	0.3	0.314	-0.014	0.043	20.045
μ_2	0.4	0.390	0.010	0.024	00.676
σ_2	0.2	0.196	0.004	0.016	00.272
τ	1	0.927	0.073	0.131	22.49
N = 400					
P_1	0.4	0.396	0.004	0.035	1.241
μ_1	0.3	0.304	-0.004	0.014	0.212
σ_1	0.15	0.153	-0.003	0.010	0.109
P_2	0.3	0.310	-0.010	0.031	1.061
μ_2	0.4	0.392	0.008	0.016	0.320
σ_2	0.2	0.197	0.003	0.011	0.130
τ	1	0.945	0.055	0.087	10.594
N = 800					
P_1	0.4	0.396	0.004	0.024	0.592
μ_1	0.3	0.303	-0.003	0.009	0.090
σ_1	0.15	0.153	-0.003	0.007	0.058
P_2	0.3	0.306	-0.006	0.023	0.565
μ_2	0.4	0.394	0.006	0.011	0.157
σ_2	0.2	0.198	0.002	0.008	0.068
τ	1	0.967	0.033	0.065	5.314

θ_0 : True values of parameters; Est is the mean of estimates; SE is the mean of standard error; MSE is the mean square errors for the MLE estimators

with increasing sample sizes for all considered parameters in all the scenarios. Moreover, the results show that the estimated values are similar for the two different $K(\cdot)$ and

Table 2a: Simulation results for the logistic smoothed function

Parameters	X~ Uniform(0,1)				
	θ_0	Est	Bias	SE	MSE×1000
N = 200					
ρ_1	0.4	0.380	0.020	0.042	2.164
μ_1	0.3	0.322	-0.022	0.017	0.773
σ_1	0.15	0.166	-0.016	0.012	0.400
ρ_2	0.3	0.325	-0.025	0.039	2.146
μ_2	0.4	0.380	0.020	0.020	0.800
σ_2	0.2	0.193	0.007	0.014	0.245
τ	0.5	0.456	0.044	0.082	8.660
N = 400					
ρ_1	0.4	0.384	0.016	0.031	1.217
μ_1	0.3	0.317	-0.017	0.013	0.458
σ_1	0.15	0.163	-0.013	0.009	0.250
ρ_2	0.3	0.319	-0.019	0.028	1.145
μ_2	0.4	0.380	0.020	0.016	0.656
σ_2	0.2	0.195	0.005	0.011	0.146
τ	0.5	0.448	0.052	0.060	6.304
N = 800					
ρ_1	0.4	0.390	0.010	0.022	0.584
μ_1	0.3	0.313	-0.013	0.008	0.233
σ_1	0.15	0.161	-0.011	0.006	0.157
ρ_2	0.3	0.319	-0.019	0.020	0.761
μ_2	0.4	0.385	0.015	0.010	0.325
σ_2	0.2	0.196	0.004	0.007	0.065
τ	0.5	0.455	0.045	0.047	4.234

θ_0 : True values of parameters, Est: The mean of estimates, SE: The mean of standard errors, MSE: The mean square errors for the MLE estimators

Table 2b: Simulation results for the logistic smoothed function

Parameters	X~ tN(1,1,0,2)				
	θ_0	Est	Bias	SE	MSE×1000
N = 200					
ρ_1	0.4	0.392	0.008	0.046	2.180
μ_1	0.3	0.311	-0.011	0.019	0.482
σ_1	0.15	0.158	-0.008	0.013	0.233
ρ_2	0.3	0.315	-0.015	0.043	2.074
μ_2	0.4	0.384	0.016	0.022	0.740
σ_2	0.2	0.196	0.004	0.015	0.241
τ	1	0.925	0.075	0.128	22.009
N = 400					
ρ_1	0.4	0.392	0.008	0.032	1.088
μ_1	0.3	0.308	-0.008	0.014	0.260
σ_1	0.15	0.156	-0.006	0.009	0.117
ρ_2	0.3	0.315	-0.015	0.028	1.009
μ_2	0.4	0.389	0.011	0.014	0.317
σ_2	0.2	0.196	0.004	0.010	0.116
τ	1	0.929	0.071	0.091	13.322
N = 800					
ρ_1	0.4	0.395	0.005	0.023	0.554
μ_1	0.3	0.307	-0.007	0.010	0.149
σ_1	0.15	0.156	-0.006	0.007	0.085
ρ_2	0.3	0.311	-0.011	0.022	0.605
μ_2	0.4	0.389	0.011	0.011	0.242
σ_2	0.2	0.198	0.002	0.008	0.068
τ	1	0.950	0.050	0.072	7.684

θ_0 : True values of parameters; Est is the mean of estimates; SE is the mean of standard error; MSE is the mean square errors for the MLE estimators

thus the accuracy of the estimates are not sensitive to the choice of the smoothing function.

CONCLUSION

Although, survival cure models have been broadly studied for decades and many applications have been

reported, the scientific literature lacks studies that investigate or propose methods for specifying the effects of covariates on the survival rate and cure rate. However, cured subjects may exist in change-point situations. The change-point cure rate model, which was proposed by Othman *et al.* (2012) as an extension to the change-point hazard model proposed by Matthews and Farewell (1982) provides a useful addition to the literatures of the cure model. In this study, a smoothed likelihood function-based on EM algorithm for efficient estimation of the cure fraction in the MCM has been proposed. Further work is needed and it should take into account investigation of cure models with the change point phenomena. Instead of the mixture modeling approach, a non-mixture model could be considered. Such models may have nice biological interpretations and it is therefore expected that the scientific community will be better able to present a much more meaningful interpretation of the modeling and estimation results than what the NCMs allow for. Interval censoring, which is commonly encountered in medical studies, can be a useful extension, especially in follow-up studies.

REFERENCES

Ahcar, A.J., E.A. Coelho-Barros and J. Mazuchel, 2012. Cure fraction models using mixture and non-mixture models. *Tatra Mt. Math. Publ.*, 51: 1-9.

Banerjee, S. and B.P. Carlin, 2004. Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*, 60: 268-275.

Berkson, J. and R.P. Gage, 1952. Survival curve for cancer patients following treatment. *J. Am. Stat. Assoc.*, 47: 501-515.

Boag, J.W., 1949. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. R. Stat. Soc.*, 11: 15-53.

Broet, P., Y.D. Rycke, P. Tubert-Bitter, J. Lellouch, B. Asselain and T. Moreau, 2001. A Semiparametric approach for the two-sample comparison of survival times with long-term survivors. *Biometrics*, 57: 844-852.

Brown, E.R. and J.G. Ibrahim, 2003. Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics*, 59: 686-693.

Castro, M.D., V.D. Cancho and J. Rodrigues, 2010. A hands-on approach for fitting long-term survival models under the GSMLSS framework. *Comput. Methods Prog. Med.*, 97: 168-177.

Chen, M.H., J.G. Ibrahim and D. Sinha, 1999. A new Bayesian model for survival data with a surviving fraction. *J. Am. Statist. Assoc.*, 94: 909-919.

- Chen, M.H., J.G. Ibrahim and S.R. Lipsitz, 2002. Bayesian methods for missing covariates in cure rate models. *Lifetime Data Anal.*, 8: 117-146.
- Cooner, F., S. Banerjee, B.P. Carlin and D. Sinha, 2007. Flexible cure rate modeling under latent activation schemes. *J. Am. Stat. Assoc.*, 102: 560-572.
- Corbiere, F., D. Commenges, J.M.G. Taylor and P. Joly, 2009. A penalized likelihood approach for mixture cure models. *Stat. Med.*, 28: 510-524.
- Farewell, V.T., 1982. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38: 1041-1046.
- Farewell, V.T., 1986. Mixture models in survival analysis: Are they worth the risk? *Can. J. Stat.*, 14: 257-262.
- Gamel, J.W., I.W. McLean and S.H. Rosenberg, 1990. Proportion cured and mean log survival time as functions of tumour size. *Statist. Med.*, 9: 999-1006.
- Ghitany, M.E., R.A. Maller and S. Zhou, 1994. Exponential mixture models with long term survivors and covariates. *J. Multivariate Anal.*, 49: 218-241.
- Goetghebeur, E. and L. Ryan, 2000. Semiparametric regression analysis of interval-censored data. *Biometrics*, 56: 1139-1144.
- Goldman, A.I., 1984. Survivorship analysis when cure is a possibility: A monte carlo study. *Stat. Med.*, 3: 153-163.
- Gutierrez, R.G., 2002. Parametric frailty and shared frailty survival models. *Stata J.*, 2: 22-44.
- Herring, A.H. and J.G. Ibrahim, 2002. Maximum likelihood estimation in random effects cure rate models with nonignorable missing covariates. *Biostatistics*, 3: 387-405.
- Ibrahim, J.G., M.H. Chen and D. Sinha, 2001. *Bayesian Survival Analysis*. Springer, New York, USA., ISBN-13: 9780387952772, Pages: 479.
- Kim, S., Y. Xi and M.H. Chen, 2009. A new latent cure rate marker model for survival data. *Ann. Applied Stat.*, 3: 1124-1146.
- Kim, Y. and M. Jhun, 2008. Cure rate model with interval censored data. *Stat. Med.*, 27: 3-14.
- Kuk, A.Y.C. and C.H. Chen, 1992. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79: 531-541.
- Lam, K.F., D.Y.T. Fong and O.Y. Tang, 2005. Estimating the proportion of cured patients in a censored sample. *Stat. Med.*, 24: 1865-1879.
- Liu, H. and Y. Shen, 2009. A semi parametric regression cure model for interval censored data. *J. Am. Stat. Assoc.*, 487: 1168-1178.
- Lopes, C.M.C. and H. Bolfarine, 2012. Random effects in promotion time cure rate models. *Comput. Stat. Data Anal.*, 56: 75-87.
- Lu, W. and Z. Ying, 2004. On semiparametric transformation cure models. *Biometrika*, 91: 331-343.
- Ma, S., 2010. Mixed case interval censored data with a cured subgroup. *Stat. Sinica*, 20: 1165-1181.
- Ma, Y., 2011. Testing change-point in logistic models with covariate measurement error. *J. Stat. Res.*, 45: 131-138.
- Maller, R.A. and X. Zhou, 1996. *Survival Analysis with Long-Term Survivors*. Wiley, New York, ISBN: 9780471962014, Pages: 278.
- Matthews, D.E. and V.T. Farewell, 1982. On testing for constant hazard against a change point alternative. *Biometrics*, 38: 463-468.
- Muller, H.G. and J.L. Wang, 1990. Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. *Biometrika*, 77: 305-314.
- Othu, M., Y. Li and R. Tiwari, 2012. Change-point cure models with application to estimating the change-point effect of age of diagnosis among prostate cancer patients. *J. Applied Stat.*, 39: 901-911.
- Peng, Y. and J.M.G. Taylor, 2010. Mixture cure model with random effects for the analysis of a multi-center tonsil cancer study. *Stat. Med.*, 30: 211-223.
- Peng, Y. and K.B.G. Dear, 2000. A nonparametric mixture model for cure rate estimation. *Biometrics*, 56: 237-243.
- Pons, O., 2003. Estimation in a cox regression model with a change-point according to a threshold in a covariate. *Ann. Stat.*, 31: 442-463.
- Rodrigues, J., V.G. Cancho, M. de Castro and F. Louzada-Neto, 2009. On the unification of long-term survival models. *Stat. Probab. Lett.*, 79: 753-759.
- Sen, P.K., 1993. Some change-point problems in survival analysis: Relevance of nonparametric in applications. *Applied Change Point Problems in Statistics*, Baltimore, MD., pp: 325-336.
- Seppa, K., T. Hakulinen, H.J. Kim and E. Laara, 2010. Cure fraction model with random effects for regional variation in cancer survival. *Stat. Med.*, 29: 2781-2793.
- 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.
- Sy, J.P. and J.M.G. Taylor, 2000. Estimation in a cox proportional hazards cure model. *Biometrics*, 56: 227-236.
- Taylor, J.M., 1995. Semi-parametric estimation in failure time mixture models. *Biometrics*, 51: 899-907.
- Tsodikov, A., 1998. A proportional hazard model taking account of long-term survivors. *Biometrics*, 54: 1508-1516.

- Tsodikov, A., 2002. Semi-parametric models of long-and short-term survival: An application to the analysis of breast cancer survival in Utah by age and stage. *Stat. Med.*, 21: 895-920.
- Tsodikov, A.D., J.G. Ibrahim and A.Y. Yakovlev, 2003. Estimating cure rates from survival data: An alternative to two-component mixture models. *J. Am. Stat. Assoc.*, 98: 1063-1078.
- Uddin, M.T., A. Sen, M.S. Noor, M.N. Islam and Z.I. Chowdhury, 2006a. An analytical approach on non-parametric estimation of cure rate based on uncensored data. *J. Applied Sci.*, 6: 1258-1264.
- Uddin, M.T., M.N. Islam and Q.I.U. Ibrahim, 2006b. An analytical approach on cure rate estimation based on uncensored data. *J. Applied Sci.*, 6: 548-552.
- Yakovlev, A.Y., B. Asselain, V.J. Bardou, A. Fourquet, T. Hoang, A. Rochefediere and A.D. Tsodikov, 1993. A Simple Stochastic Model of Tumor Recurrence and its Applications to Data on Pre-Menopausal Breast Cancer. In: *Biometrics and Analysis Dormees Spatio-Temporal*, Asselain, B., M. Boniface, C. Duby, C. Lopez, J.P. Masson and J. Tranchefort (Eds.). French Society of Biometrics, France, pp: 66-82.
- Yin, G. and J.G. Ibrahim, 2005. Cure rate models: A unified approach. *Can. J. Stat.*, 33: 559-570.
- Yin, G., 2005. Bayesian cure rate frailty models with application to a root canal therapy study. *Biometrics*, 61: 552-558.
- Zeng, D., G. Yin and J.G. Ibrahim, 2006. Semiparametric transformation models for survival data with a cure fraction. *J. Am. Stat. Assoc.*, 101: 670-684.
- Zhang, J. and Y. Peng, 2009. Accelerated hazard mixture cure model. *Lifetime Data Anal.*, 15: 455-467.
- Zhao, X., X. Wu and X. Zhou, 2009. A change-point model for survival data with long-term survivors. *Stat. Sinica*, 19: 377-390.