computer methods and programs in biomedicine 112 (2013) 114–124

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

# The exponentiated exponential mixture and non-mixture cure rate model in the presence of covariates

*Josmar Mazucheli*[a,*], *Emílio A. Coelho-Barros*[b,c], *Jorge Alberto Achcar*[c]

[a] *Universidade Estadual de Maringá, Departamento de Estatística, DEs/UEM, Maringá, PR, Brazil*
[b] *COMAT, Universidade Tecnológica Federal do Paraná, Cornélio Procópio, PR, Brazil*
[c] *Universidade de São Paulo, Departamento de Medicina Social, FMRP/USP, Ribeirão Preto, SP, Brazil*

ARTICLE INFO

*Article history:*
Received 7 August 2012
Received in revised form
29 May 2013
Accepted 20 June 2013

*Keywords:*
Long-term data
Mixture lifetime model
Non-mixture lifetime model
Exponentiated Weibull distribution

ABSTRACT

This paper presents estimates for the parameters included in long-term mixture and non-mixture lifetime models, applied to analyze survival data when some individuals may never experience the event of interest. We consider the case where the lifetime data have a two-parameters exponentiated exponential distribution. The two-parameter exponentiated exponential or the generalized exponential distribution is a particular member of the exponentiated Weibull distribution introduced by [31]. Classical and Bayesian procedures are used to get point and confidence intervals of the unknown parameters. We consider a general survival model where the scale, shape and cured fraction parameters of the exponentiated exponential distribution depends on covariates.



## 1. Introduction

A long-term survivor mixture model, also known as standard cure rate model, assumes that the studied population is a mixture of susceptible individuals, who experience the event of interest and non-susceptible individuals that will never experience it. These individuals are not at risk with respect to the event of interest and are considered immune, non-susceptible or cured [28]. Different approaches, parametric and non-parametric, have been considered to model the proportion of immunes and interested readers can refer, for example, to Refs. [4,5,11,12,14–16,18,21,30,32,33,40]. Following [28], the standard cure rate model assumes that a certain fraction $p$ in the population is cured or never fail with respect to the specific cause of death or failure, while the remaining $(1-p)$ fraction of the individuals is not cured, leading to the survival function for the entire population written as:

$$S(t) = p + (1-p)S_0(t), \tag{1}$$

where $p \in (0, 1)$ is the mixing parameter and $S_0(t)$ denotes a proper survival function for the non-cured group in the population. Considering a random sample of lifetimes $(t_i, \delta_i, i=1, \ldots, n)$, the contribution of the ith individual for the likelihood function is:

$$L_i = [f(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i}, \tag{2}$$

where $\delta_i$ is a censoring indicator variable, that is $\delta_i = 1$ for an observed lifetime and $\delta_i = 0$ for a censored lifetime.

* *Corresponding author*. Tel.: +55 4430114338.
E-mail address: jmazucheli@uem.br (J. Mazucheli).
0169-2607/$ – see front matter 
http://dx.doi.org/10.1016/j.cmpb.2013.06.015

From the mixture survival function, (1), the probability density function is obtained from $f(t_i) = -(d/dt)S(t_i)$ and given by:

$$f(t_i) = (1-p) f_0(t_i), \tag{3}$$

where $f_0(t_i)$ is the probability density function for the susceptible individuals. Substitution of the mixture density (3) and survival function (1) in the standard likelihood function (2) yields the likelihood for the long-term survivor mixture model:

$$L_i = [(1-p) f_0(t_i)]^{\delta_i} [p + (1-p) S_0(t_i)]^{1-\delta_i}. \tag{4}$$

Thus, the log-likelihood considering all observations is given by:

$$l = r \log(1-p) + \sum_{i=1}^{n} \delta_i \log f_0(t_i) + \sum_{i=1}^{n} (1-\delta_i) \log[p + (1-p) S_0(t_i)], \tag{5}$$

where $r = \sum_{i=1}^{n} \delta_i$ is the number of uncensored observations. Common choices for the survival function $S_0(t)$, in (1), are the exponential and Weibull distributions. Ref. [34] investigated the use of a generalized Fisher–Snedecor distribution as baseline for $S_0(t)$. The generalized Fisher–Snedecor distribution is a supermodel that includes the most popular survival models as particular cases, such as the exponential, Weibull, log-normal, among others. Ref. [43] considered the generalized log-gamma distribution for the mixture cure rate model in the context of accelerated failure-time regression models. The Gompertz distribution was considered by Ref. [19], while the exponentiated Weibull and exponentiated exponential distributions were considered, respectively, by Refs. [6,23]. The Conway–Maxwell Poisson cure rate model was proposed by Ref. [35] as an alternative to the cure rate model discussed by Ref. [44].

An alternative to a long-term survivor mixture model is the long-term survivor non-mixture model suggested by Refs. [42,41,26] which defines an asymptote for the cumulative hazard and hence for the cure fraction. The survival function for a non-mixture cure rate model is defined as:

$$S(t) = p^{1-S_0(t)}, \tag{6}$$

where, like in (1), $p \in (0,1)$ is the mixing parameter and $S_0(t)$ denotes a proper survival function for the non-cured group. Observe that if the probability of cure is large, then the intrinsic survival function $S(t)$ is large—$S_0(t)$ will be large which implies in $F_0(t) = 1 - S_0(t)$ small. Larger values of $F_0(t)$ at a fixed time $t$ imply lower values of $S(t)$. The non-mixture model (6) or the promotion time cure fraction has been used by Refs. [27,26] to estimate the probability of cure fraction in cancer lifetime data.

From (6), the survival and hazard function for the non-mixture cure rate model can be written, respectively, as:

$$S(t_i) = \exp[\log(p) F_0(t_i)] \tag{7}$$

and

$$h(t_i) = -\log(p) f_0(t_i). \tag{8}$$

Since $f(t) = h(t)S(t)$, the contribution of the ith individual for the likelihood function is given by:

$$L_i = h(t_i)^{\delta_i} S(t_i) \tag{9}$$

that is:

$$L_i = [-\log(p) f_0(t_i)]^{\delta_i} \exp[\log(p) F_0(t_i)]. \tag{10}$$

Considering a random sample of lifetimes $(t_i, \delta_i, i = 1, \ldots, n)$ the log-likelihood is:

$$l = r \log[-\log(p)] + \sum_{i=1}^{n} \delta_i \log f_0(t_i) + \log(p) \sum_{i=1}^{n} [1 - S_0(t_i)], \tag{11}$$

where, as before, $r = \sum_{i=1}^{n} \delta_i$.

A Bayesian formulation of the non-mixture cure rate model is given in Ref. [9]. A model which includes a standard mixture model for cure rate was considered in Ref. [44]. Ref. [36] extended the long-term survival model proposed by Ref. [9].

In this paper, considering the exponentiated exponential distribution, we compare the performance of the mixture and non-mixture cure fraction formulation when the fraction $p$ and the other parameters may vary by covariates. The paper is organized as follows: in Section 2, we introduce the likelihood function assuming an exponentiated exponential distribution for the susceptible individuals; in Section 3, we present a Bayesian analysis assuming the mixture and non-mixture models in presence or not of covariates; in Section 4, we present an application with randomized trials of two different residential treatment programs aimed to reduce drug abuse and, consequently, AIDS high-risk behavior; finally in Section 5, we introduce some comments and remarks.

## 2. The exponentiated exponential distribution cure model

Ref. [20] introduced the generalized exponential (GE) distribution, also known as exponentiated exponential distribution, this distribution is a particular member of the exponentiated Weibull distribution introduced by [31], whose probability density function is given by:

$$f_0(t \mid \alpha, \lambda) = \alpha\lambda \exp(-\lambda t) [1 - \exp(-\lambda t)]^{\alpha-1}, \tag{12}$$

where $\alpha > 0$ and $\lambda > 0$ are the shape and scale parameters, respectively. For $\alpha = 1$ we have the exponential distribution as particular case. Similarly to a Weibull distribution, the hazard function of an exponentiated exponential distribution could be increasing, decreasing or constant depending on the shape parameter $\alpha$ [22,25].

From (12), we have that the distribution and survival functions can be written, respectively, as:

$$F_0(t \mid \alpha, \lambda) = [1 - \exp(-\lambda t)]^{\alpha} \quad \text{and} \quad S_0(t \mid \alpha, \lambda) = 1 - [1 - \exp(-\lambda t)]^{\alpha}. \tag{13}$$