

A general long-term aging model with different underlying activation mechanisms: Modeling, Bayesian estimation, and case influence diagnostics

Adriano K. Suzuki, Gladys D. C. Barriga, Francisco Louzada & Vicente G. Cancho

To cite this article: Adriano K. Suzuki, Gladys D. C. Barriga, Francisco Louzada & Vicente G. Cancho (2017) A general long-term aging model with different underlying activation mechanisms: Modeling, Bayesian estimation, and case influence diagnostics, Communications in Statistics - Theory and Methods, 46:6, 3080-3098, DOI: [10.1080/03610926.2015.1053945](https://doi.org/10.1080/03610926.2015.1053945)

To link to this article: <https://doi.org/10.1080/03610926.2015.1053945>



Accepted author version posted online: 15 Apr 2016.
Published online: 15 Apr 2016.



Submit your article to this journal [↗](#)



Article views: 108



View Crossmark data [↗](#)



A general long-term aging model with different underlying activation mechanisms: Modeling, Bayesian estimation, and case influence diagnostics

Adriano K. Suzuki^a, Gladys D. C. Barriga^b, Francisco Louzada^a, and Vicente G. Cancho^a

^aDepartment of Applied Mathematics and Statistics, University of São Paulo, São Carlos, Brazil; ^bFaculty of Engineering at Bauru, São Paulo State University, Bauru, Brazil

ABSTRACT

In this paper we propose a general cure rate aging model. Our approach enables different underlying activation mechanisms which lead to the event of interest. The number of competing causes of the event of interest is assumed to follow a logarithmic distribution. The model is parameterized in terms of the cured fraction which is then linked to covariates. We explore the use of Markov chain Monte Carlo methods to develop a Bayesian analysis for the proposed model. Moreover, some discussions on the model selection to compare the fitted models are given, as well as case deletion influence diagnostics are developed for the joint posterior distribution based on the ψ -divergence, which has several divergence measures as particular cases, such as the Kullback–Leibler (K-L), J -distance, L_1 norm, and χ^2 -square divergence measures. Simulation studies are performed and experimental results are illustrated based on a real malignant melanoma data.

ARTICLE HISTORY

Received 4 May 2015
Accepted 5 May 2015

KEYWORDS

Competing risks; cure rate models; cured fraction; logarithmic distribution; long-term aging models.

MATHEMATICS SUBJECT CLASSIFICATION

62N99

1. Introduction

Cure rate models cover a large amount of practical situations where there are sampling units unsusceptible to the occurrence of the event of interest. The proportion of such units is usually termed as the cured fraction. The statistical literature on modeling lifetime data in presence of a cure fraction is by now vast and growing rapidly.

Perhaps the most popular type of cure rate model is the mixture model introduced by Boag (1949) and Berkson and Gage (1952). In this model, it is assumed that a certain proportion of the individuals are cured, in the sense that they do not present the event of interest during a long period of time and can found to be immune to the cause of failure under study. The mixture model has been considered by various authors, such as Farewell (1982, 1986); Ghitany and Maller (1994); Maller and Zhou (1995); Chen and Ibrahim (2001); Sposto (2002); Perperoglou et al. (2007), and Perdona and Louzada-Neto (2011), among others.

Following the proposition of comprehensive cure rate models, Yakovlev and Tsodikov (1996); Ibrahim et al. (2001); Chen et al. (2002, 1999); Tsodikov et al. (2003); Tournoud and Ecochard (2007); de Castro et al. (2009); Ortega et al. (2009); de Castro et al. (2010); Cancho et al. (2011b), and Kim et al. (2011) include in their formulation the possibility of

having a cured rate in the population, and assume the occurrence of the event of interest might be a result of many competing causes (Gordon, 1990), with the number of causes and the survival times associated with each cause (Cox and Oakes, 1984) being unknown, which leads to the so-called latent competing causes (Louzada-Neto, 1999), which are assumed to follow a discrete distribution such as the Poisson, the negative binomial, the geometric, the COM-Poisson, and the power series, among others. For instance, in clinical studies the event of interest may be the death of a patient (which can occur due to different competing causes) or a tumor recurrence (which can be attributed to metastasis-component tumor cells left active after an initial treatment). Another approach, by Cooner et al. (2006, 2007), forms an arranged stochastic sequence of latent causes, which induce the occurrence of the event of interest via an underlying activation mechanism that leads to the event of interest.

We deal with different underlying activation mechanisms that lead to the event of interest when the latent number of competing causes is modeled by a logarithmic distribution (Johnson et al., 2005), hereafter, called the logarithmic cure rate (LCR) model. We focus in the scan of a type of activation mechanism, beginning with the first activation possible, or the last one, or even a random mechanism. We assume there is no information about which cause was responsible for the individual death or disease recurrence.

Bayesian analysis for the cure rate model has appeared in the literature. For instance, we cite Chen et al. (1999) and Cancho et al. (2011b). However, the Bayesian approach to LRC model under a variety of activation mechanisms has never been considered. Thus, one purpose of this paper is to propose a Bayesian approach for drawing inferences in LCR models in the presence of covariates on the cure fraction.

After fitting a model, it is important to check its assumptions and conduct sensitivity studies in order to detect possible influential or extreme observations, which may cause distortions on the analysis results, leading to the so-called diagnostic methods. Following the pioneering work by Cook (1986), case-deletion and local influence diagnostics have been widely applied to many regression models. In order to examine outlying and influential observations, Louzada et al. (2013) present a Bayesian case deletion influence diagnostics based on the Kullback–Leibler divergence. Then, in the above context, other purpose of this paper is to develop case deletion influence diagnostics for the joint posterior distributions of the parameters of the GBScR model based on the ψ -divergence measure (Peng and Dey, 1995; Weiss, 1996). The ψ -divergence measure includes several divergence measures as particular cases, such as the Kullback–Leibler (K–L), J -distance, L_1 norm, and χ^2 -square divergence measures.

The paper is organized as follows. In Section 2 we formulate the model. We explore the use of Markov chain Monte Carlo (MCMC) methods to develop a Bayesian analysis in Section 3. A simulation study with different models is presented in Section 4. An application to a real data set is developed in Section 5. Finally, Section 6 ends up with some general remarks.

2. Model formulation

For an individual in the population, let M denote the unobservable number of causes of the event of interest for this individual. Assume that M follows a logarithmic distribution (Johnson et al., 2005) with probability mass function

$$P(M = m; \theta) = \frac{-1}{\log(1 - \theta)} \frac{\theta^{m+1}}{(m + 1)}, \quad m = 0, 1, 2, \dots, \quad 0 < \theta < 1 \quad (1)$$

The probability generating function of M is given by

$$G_M(s) = \frac{\log(1 - \theta s)}{s \log(1 - \theta)}, \text{ for } |s| < 1 \quad (2)$$

The time for the j th cause to produce the event of interest is denoted by Z_j , $j = 1, \dots, M$. We assume that, conditional on M , the Z_j are i.i.d. with cumulative distribution function $F(z)$ and surviving function $S(z) = 1 - F(z)$. Also, we assume that Z_1, Z_2, \dots are independent of M . The observable time to event is defined by the random variable $Y = Z_{(R)}$, where R depends on M , $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(R)} \leq \dots \leq Z_{(M)}$ are the order statistics, and $Y = \infty$ if $M = 0$. In many biological processes R can be interpreted as a resistance factor of the immune system of the individual. If the event of interest occurs (e.g., cancer relapse), then the random variable Y takes the value of the R th order statistics $Z_{(R)}$. In other words, as in Cooner et al. (2006, 2007), R out of M causes are required to produce the event of interest. The resistance factor can be a fixed constant, a function of M , or a random variable specified through a conditional distribution on M .

Using the terminology borrowed from Cooner et al. (2006, 2007), in this paper we deal with three specifications for R . First, we assume that given $M \geq 1$, the conditional distribution of R is uniform on $\{1, 2, \dots, M\}$ (random activation scheme). Under this setup, the surviving function for the population is given by

$$\begin{aligned} S_{\text{pop}}(y) &= P(Y > y) = P(M = 0) + \{1 - P(M = 0)\}S(y) \\ &= -\frac{\theta}{\log(1 - \theta)} + \left(1 + \frac{\theta}{\log(1 - \theta)}\right)S(y) \end{aligned} \quad (3)$$

which comes to be a mixture cure model with cured fraction $p_0 = P(M = 0) = \lim_{y \rightarrow \infty} S_{\text{pop}}(y) = -\theta / \log(1 - \theta)$. From (3) the density function is

$$f_{\text{pop}}(y) = -S'_{\text{pop}}(y) = \left(1 + \frac{\theta}{\log(1 - \theta)}\right) f(y) \quad (4)$$

where $f(y) = -S'(y)$ denotes the proper density function of the time to event Z . Furthermore, the corresponding hazard function is

$$h_{\text{pop}}(y) = -\left[\frac{\theta + \log(1 - \theta)}{\theta - (\theta + \log(1 - \theta))S(y)}\right] f(y)$$

As a second setup, the so-called first activation scheme, we suppose that the event of interest happens due to any one of the possible causes. Therefore, for $R = 1$, the time to event is $Y = Z_{(1)} = \min\{Z_1, \dots, Z_M\}$, implying that (Tsodikov et al., 2003)

$$S_{\text{pop}}(y) = G_M(S(y)) = \frac{\log(1 - \theta S(y))}{S(y) \log(1 - \theta)} \quad (5)$$

where $G_M(\cdot)$ is as in (2). The cured fraction is given by $p_0 = -\theta / \log(1 - \theta)$. The density function associated with (5) is given by

$$f_{\text{pop}}(y) = -\left[\frac{\theta S(y) + (1 - \theta S(y)) \log(1 - \theta S(y))}{(1 - \theta S(y))S^2(y) \log(1 - \theta)}\right] f(y) \quad (6)$$

with hazard function

$$h_{\text{pop}}(y) = -\left[\frac{\theta S(y) + (1 - \theta S(y)) \log(1 - \theta S(y))}{(1 - \theta S(y))S(y) \log(1 - \theta S(y))}\right] f(y)$$

In our third scenario, also known as the last activation scheme, the event of interest only takes place after all the M causes have been occurred, so that $R = M$ and the observed failure time is $Y = Z_{(M)} = \max\{Z_1, \dots, Z_M\}$. According to Cooner et al. (2007),

$$S_{\text{pop}}(y) = 1 + G_M(0) - G_M(F(y))$$

Hence, from (2) we have

$$S_{\text{pop}}(y) = 1 - \frac{\theta}{\log(1 - \theta)} - \frac{\log(1 - \theta F(y))}{F(y) \log(1 - \theta)} \tag{7}$$

so that the cured fraction is $p_0 = -\theta / \log(1 - \theta)$. The surviving function in (7) leads to the density function

$$f_{\text{pop}}(y) = - \left[\frac{\theta F(y) + (1 - \theta F(y)) \log(1 - \theta F(y))}{(1 - \theta F(y)) F^2(y) \log(1 - \theta)} \right] f(y) \tag{8}$$

with hazard function

$$h_{\text{pop}}(y) = - \left[\frac{\theta F(y) + (1 - \theta F(y)) \log(1 - \theta F(y))}{(1 - \theta F(y)) F^2(y) \log(1 - \theta) \left(1 - \frac{\theta}{\log(1 - \theta)} - \frac{\log(1 - \theta F(y))}{F(y) \log(1 - \theta)} \right)} \right] f(y)$$

From these results we realize that, whichever the activation scheme, the cured fraction is the same. The models differ by their surviving, density, and hazard functions. The relationship between the distribution functions in (3), (5), and (7) is described in Proposition 1.

Proposition 1. *Under conditions of models in (3), (5), and (7) and for any distribution function $F(y)$, we have $S_{\text{pop}}^{\text{first}}(y) \leq S_{\text{pop}}^{\text{random}}(y) \leq S_{\text{pop}}^{\text{last}}(y)$ for $y > 0$.*

The proof follows from the Theorem 2.1 from Kim et al. (2011).

The proper surviving function for non cured population, denoted by $S_{\text{nc}}(y)$, is computed by $S_{\text{nc}}(y) = P(Y > y | M \geq 1)$. We present in Table 1 the density and surviving functions for the non cured population under different activation schemes. We note that $S_{\text{nc}}(0) = 1$ and $S_{\text{nc}}(\infty) = 0$, so that it is a proper surviving function. From the distribution for the non cured population under the first activation, considering different choices for the distribution of the latent random variables Z_j 's, some recently proposed lifetime models can be obtained as a special case. For example, if $\{Z_j\}_{j \in \mathbb{N}}$ follows the exponential or Weibull distribution, the exponential logarithmic distribution (Tahmasbi and Rezaei, 2008) and Weibull logarithmic (Ciumara and Preda, 2009) distribution, respectively, are obtained. From Table 1 new families of distribution can be generated.

In many applications of long-term aging models the cured fraction plays a central role. With this concern in mind, we change the parameterization of the model in order to put the

Table 1. Non cured surviving function (S_{nc}) and density function (f_{nc}) for logarithmic models under different activation devices.

Activation	$S_{\text{nc}}(y)$	$f_{\text{nc}}(y)$
First	$\frac{\log(1 - \theta S(y))}{\log(1 - \theta)}$	$-\frac{\theta f(y)}{\log(1 - \theta)(1 - \theta S(y))}$
Last	$1 - \frac{\log(1 - \theta F(y))}{\log(1 - \theta)}$	$-\frac{\theta f(y)}{\log(1 - \theta)(1 - \theta F(y))}$
Random	$S(y)$	$f(y)$

Table 2. Surviving function (S_{pop}) and density function (f_{pop}) for Logarithmic cure rate models under different activation devices.

Activation	$S_{\text{pop}}(y)$	$f_{\text{pop}}(y)$
First	$-\frac{\log(1 - W_0 S(t))}{W_0 S(y)} p_0$	$\frac{W_0 S(y) + \{1 - W_0 S(y)\} \log(1 - W_0 S(y))}{\{1 - W_0 S(y)\} W_0 S(y)^2} p_0 f(y)$
Last	$1 + p_0 + \frac{\log(1 - W_0 F(y))}{W_0 F(y)} p_0$	$\frac{W_0 F(y) + \{1 - W_0 F(y)\} \log(1 - W_0 F(y))}{\{1 - W_0 F(y)\} W_0 F(y)^2} p_0 f(y)$
Random	$p_0 + (1 - p_0) S(y)$	$(1 - p_0) f(y)$

Remark: $W_0 = 1 + p_0 W(-e^{-1/p_0}/p_0)$, where $W(\cdot)$ is the Lambert W function (Corless et al., 1996).

cured fraction p_0 in the expressions. Since $p_0 = P(M = 0) = -\theta / \log(1 - \theta)$, we have that $p_0 = 1 + W(-\theta e^{-\theta})/\theta$, where $W(\cdot)$ stands for the Lambert W function (Corless et al., 1996). As we shall see more clearly in Section 3, this parameterization is advantageous. Using p_0 as parameter and expressions (3)–(8), we obtain the improper surviving and density functions presented in Table 2. The Lambert W function in Table 2 can be found in the R package `emdbook`.

Hereafter, we assume a Weibull distribution for the unobserved time Z with $F(z; \boldsymbol{\gamma}) = 1 - \exp(-z^{\gamma_1} e^{\gamma_2})$ and $f(z; \boldsymbol{\gamma}) = \gamma_1 z^{\gamma_1 - 1} \exp(\gamma_2 - z^{\gamma_1} e^{\gamma_2})$, for $z > 0$, $\gamma_1 > 0$, $\gamma_2 \in \mathbb{R}$, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^\top$. Of course, at least in principle, other lifetime distributions may be considered. Figure 1 portrays distinct behaviors of the surviving functions in Tables 1 and 2. These plots illustrate the flexibility afforded by our proposal.

3. Inference

Let us consider the situation when the failure time Y in Section 2 is not completely observed and is subject to right censoring. Let C_i denote the censoring time. In a sample of size n , we then observe $T_i = \min\{Y_i, C_i\}$ and $\delta_i = I(Y_i \leq C_i)$, where $\delta_i = 1$ if T_i is a failure time and $\delta_i = 0$ if it is right censored, for $i = 1, \dots, n$.

Let $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^\top$ denote the vector of covariates for the i th individual. Completing our model, we propose to relate the cured fraction to the covariates by the logistic link

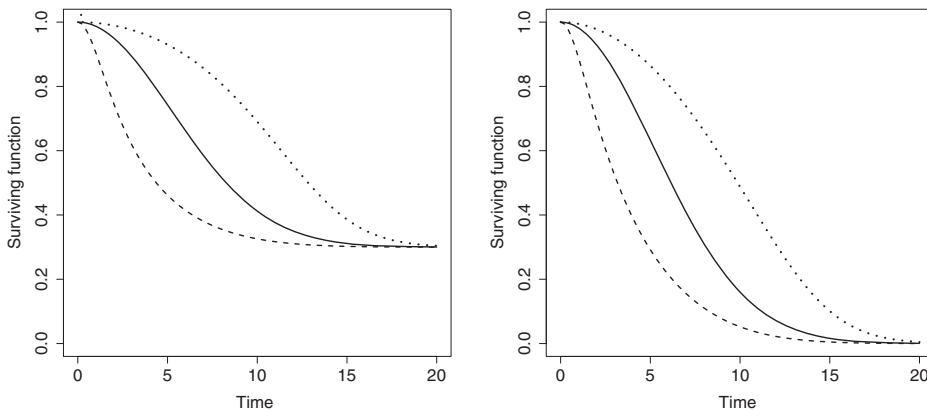


Figure 1. Surviving functions population (left panel) and non cured population (right panel) for the logarithmic models with $p_0 = 0.3$, and a Weibull distribution ($\gamma_1 = 2, \gamma_2 = -4$) under different activations (first: dashed, random: solid, and last: dotted).

$$\log\left(\frac{p_{0i}}{1 - p_{0i}}\right) = \mathbf{x}_i^\top \boldsymbol{\beta} \quad \text{or} \quad p_{0i} = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \tag{9}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ encapsulates the vector of regression coefficients, so that for each group of individuals represented by \mathbf{x}_i , we have a different cured fraction. With this link function the models are identifiable in the sense of Li et al. (2001).

We stress that regardless of the specific model in Table 2, covariates are associated with the cured fraction through a unique expression very familiar to practitioners. If we had adopted the parameterization in θ , for logarithmic distributions the cured fraction is $p_0 = -\theta / \log(1 - \theta)$, and θ would be linked to the covariates (e.g. logistic links). The connections between the cured fraction and the covariates would be much more clumsy in these expressions than in (9). Therefore, although the entries for the logarithmic model in Table 2 seem uneasy, we have a direct interpretation of the coefficients in (9).

With the expression (9) we can write the likelihood of $\boldsymbol{\vartheta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ under non informative censoring as

$$L(\boldsymbol{\vartheta}; \mathcal{D}) \propto \prod_{i=1}^n f_{\text{pop}}(t_i; \boldsymbol{\vartheta})^{\delta_i} S_{\text{pop}}(t_i; \boldsymbol{\vartheta})^{1-\delta_i} \tag{10}$$

where $\mathcal{D} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{x})$, $\mathbf{t} = (t_1, \dots, t_n)^\top$, $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$, and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$, whereas $f_{\text{pop}}(\cdot; \boldsymbol{\vartheta})$ and $S_{\text{pop}}(\cdot; \boldsymbol{\vartheta})$ are the improper density and surviving functions, respectively, in Table 2.

3.1. Prior and posterior

Now, some inferential tools are investigated under a Bayesian viewpoint. The normal distribution and gamma distribution with a as shape and b as scale (and mean a/b) are denoted by $N(\mu, \sigma^2)$ and $G(a, b)$. In this context we assume that $\boldsymbol{\beta}$, γ_1 , and γ_2 are *a priori* independent, that is,

$$\pi(\boldsymbol{\vartheta}) = \prod_{i=1}^p \pi(\beta_i) \pi(\gamma_1) \pi(\gamma_2) \tag{11}$$

where $\beta_j \sim N(0, \sigma_{\beta_j}^2)$, $j = 1, \dots, p$, $\gamma_1 \sim G(a_0, b_0)$, and $\gamma_2 \sim N_1(0, \sigma_{\gamma_2}^2)$. Here all the hyperparameters are specified in order to express non informative priors.

Combining the likelihood function (10) and the prior distribution in (11), the joint posterior distribution for $\boldsymbol{\vartheta}$ is obtained as $\pi(\boldsymbol{\vartheta}|\mathcal{D}) \propto L(\boldsymbol{\vartheta}; \mathcal{D}) \prod_{i=1}^p \pi(\beta_i) \pi(\gamma_1) \pi(\gamma_2)$. This joint posterior density is analytically intractable. So, we based our inference on the MCMC simulation methods. In particular, the Gibbs sampler algorithm (see Gamerman and Lopes, 2006) has proved to be a powerful alternative. In this direction, we observed that there is no closed form available for any of the full conditional distributions needed to implement Gibbs sampler. Thus we instead resort to the Metropolis–Hastings algorithm. We begin by making a change of variables to $\boldsymbol{\xi} = (\log(\gamma_1), \gamma_2, \boldsymbol{\beta})$. This transforms the parameter space to \mathcal{R}^{p+2} (necessary to work with Gaussian proposal densities). Accounting for the Jacobian of this transformation, our joint posterior density (or target density) is now

$$\pi(\boldsymbol{\xi}|\mathcal{D}) \propto L(\boldsymbol{\xi}; \mathcal{D}) \exp\left\{-\frac{1}{2} \sum_{j=1}^p \frac{\beta_j^2}{\sigma_{\beta_j}^2} - \frac{1}{2} \frac{\gamma_2^2}{\sigma_{\gamma_2}^2} - \frac{\exp(\xi_1)}{b_0} + a_0 \xi_1\right\} \tag{12}$$

To implement the Metropolis–Hastings algorithm, we proceed as follows:

- (1) Start with any point $\xi_{(0)}$, and stage indicator $j = 0$.
- (2) Generate a point ξ' according to the transitional kernel $Q(\xi', \xi_j) = N_{p+2}(\xi_j, \tilde{\Sigma})$, where $\tilde{\Sigma}$ is covariance matrix of ξ is same in any stage.
- (3) Update $\xi_{(j)}$ to $\xi_{(j+1)} = \xi'$ with probability $p_j = \min\{1, \pi(\xi'|\mathcal{D})/\pi(\xi_{(j)}|\mathcal{D})\}$, or keep $\xi_{(j)}$ with probability $1 - p_j$.
- (4) Repeat steps (2) and (3) by increasing the stage indicator until the process reaches a stationary distribution.

The computational program is available from the authors upon request.

3.2. Model comparison criteria

There exist a variety of methodologies to compare several competing models for a given data set and to select the one that best fits the data. One of the most used in applied works is derived from the conditional predictive ordinate (CPO) statistic. For a detailed discussion on the CPO statistic and its applications to model selection, see Gelfand et al. (1992) and Geisser and Eddy (1979). Let \mathcal{D} the full data and $\mathcal{D}^{(-i)}$ denoting the data with the i th observation be deleted. In our model, for an observed time to event ($\delta_i = 1$) we have from Section 3 that $g(y_i|\boldsymbol{\vartheta}) = f_{\text{pop}}(y_i; \boldsymbol{\vartheta})$ and, for a censored time, $g(y_i|\boldsymbol{\vartheta}) = S_{\text{pop}}(y_i; \boldsymbol{\vartheta})$. We denote the posterior density of $\boldsymbol{\vartheta}$ given $\mathcal{D}^{(-i)}$ by $\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})$, $i = 1, \dots, n$. For the i th observation, CPO_i can be written as

$$CPO_i = \int_{\boldsymbol{\vartheta} \in \Theta} g(y_i|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}|\mathcal{D}^{(-i)})d\boldsymbol{\vartheta} = \left\{ \int_{\boldsymbol{\vartheta}} \frac{\pi(\boldsymbol{\vartheta}|\mathcal{D})}{g(y_i|\boldsymbol{\vartheta})} d\boldsymbol{\vartheta} \right\}^{-1} \tag{13}$$

The CPO_i can be interpreted as the height of the marginal density of the time to event at y_i . Thus, large values of CPO_i imply a better fit of the model. For the proposed model a closed form of the CPO_i is not available. However, a Monte Carlo estimate of CPO_i can be obtained by using a single MCMC sample from the posterior distribution $\pi(\boldsymbol{\vartheta}|\mathcal{D})$. Let $\boldsymbol{\vartheta}^{(1)}, \dots, \boldsymbol{\vartheta}^{(Q)}$ be a sample of size Q of $\pi(\boldsymbol{\vartheta}|\mathcal{D})$ after the burn-in. A Monte Carlo approximation of CPO_i (Ibrahim et al., 2001) is given by

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{g(y_i|\boldsymbol{\vartheta}^{(q)})} \right\}^{-1}$$

For model comparison we use the log pseudo marginal likelihood (LPML) defined by $LPML = \sum_{i=1}^n \log(\widehat{CPO}_i)$. The larger is the value of $LPML$, the better is the fit of the model.

Other criteria like the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002), the expected Akaike information criterion (EAIC) Brooks (2002), and the expected Bayesian (or Schwarz) information criterion (EBIC) Carlin and Louis (2001) can be used. These criteria are based on the posterior mean of the deviance, which can be approximated by $\bar{d} = \sum_{q=1}^Q d(\boldsymbol{\vartheta}_q)/Q$, where $d(\boldsymbol{\vartheta}) = -2 \sum_{i=1}^n \log [g(y_i|\boldsymbol{\vartheta})]$. The DIC can be estimated using the MCMC output by $\widehat{DIC} = \bar{d} + \widehat{\rho}_d = 2\bar{d} - \widehat{d}$, with ρ_D as the effective number of parameters, which is defined as $E\{d(\boldsymbol{\vartheta})\} - d\{E(\boldsymbol{\vartheta})\}$, where $d\{E(\boldsymbol{\vartheta})\}$ is the deviance evaluated at the posterior mean and is estimated as

$$\widehat{D} = d \left(\frac{1}{Q} \sum_{q=1}^Q \boldsymbol{\beta}^{(q)}, \frac{1}{Q} \sum_{q=1}^Q \gamma_1^{(q)}, \frac{1}{Q} \sum_{q=1}^Q \gamma_2^{(q)} \right)$$

Similarly, the EAIC and EBIC can be estimated by means of $\widehat{EAIC} = \bar{d} + 2\#(\boldsymbol{\vartheta})$ and $\widehat{EBIC} = \bar{d} + \#(\boldsymbol{\vartheta}) \log(n)$, where $\#(\boldsymbol{\vartheta})$ is the number of model parameters.

3.3. Bayesian case influence diagnostics

Since regression models are sensitive to the underlying model assumptions, generally performing a sensitivity analysis is strongly advisable. Cook (1986) uses this idea to motivate his assessment of influence analysis. He suggests that more confidence can be put in a model which is relatively stable under small modifications. The best known perturbation schemes are based on case-deletion (Cook and Weisberg, 1982) in which the effects are studied of completely removing cases from the analysis. This reasoning will form the basis for our Bayesian global influence methodology and in doing so it will be possible to determine which subjects might be influential for the analysis.

Let $D_\psi(P, P_{(-i)})$ denote the ψ -divergence between P and $P_{(-i)}$, where P denotes the posterior distribution of $\boldsymbol{\vartheta}$ for full data, and $P_{(-i)}$ denotes the posterior distribution of $\boldsymbol{\vartheta}$ without the i th case. Specifically,

$$D_\psi(P, P_{(-i)}) = \int_{\boldsymbol{\vartheta} \in \Theta} \psi \left(\frac{\pi(\boldsymbol{\vartheta} | \mathcal{D}^{(-i)})}{\pi(\boldsymbol{\vartheta} | \mathcal{D})} \right) \pi(\boldsymbol{\vartheta} | \mathcal{D}) d\boldsymbol{\vartheta} \quad (14)$$

where ψ is a convex function with $\psi(1) = 0$. Several choices of ψ are given in Dey and Birmiwal (1994). For example, $\psi(z) = -\log(z)$ defines K-L divergence, $\psi(z) = (z-1)\log(z)$ gives J -distance (or the symmetric version of K-L divergence), $\psi(z) = 0.5|z-1|$ defines the variational distance or L_1 norm, and $\psi(z) = (z-1)^2$ defines the χ^2 -square divergence.

The relationship between the CPO (13) and the ψ -divergence measure is described in next proposition.

Proposition 2. *The ψ -divergence measure can be written as*

$$D_\psi(P, P_{(-i)}) = E_{\boldsymbol{\vartheta} | \mathcal{D}} \left[\psi \left(\frac{CPO_i}{g(y_i | \boldsymbol{\vartheta})} \right) \right] \quad (15)$$

where the expected value is taken with respect to the joint posterior distribution $\pi(\boldsymbol{\vartheta} | \mathcal{D})$.

Proof. From Bayes's theorem the posterior distribution of $\boldsymbol{\vartheta}$ is given by

$$\pi(\boldsymbol{\vartheta} | \mathcal{D}) = \frac{\pi(\boldsymbol{\vartheta}) \prod_{j \in \mathcal{D}} g(y_j | \boldsymbol{\vartheta})}{\int_{\boldsymbol{\vartheta} \in \Theta} \pi(\boldsymbol{\vartheta}) \prod_{j \in \mathcal{D}} g(y_j | \boldsymbol{\vartheta}) d\boldsymbol{\vartheta}}$$

where $\pi(\boldsymbol{\vartheta})$ and $\prod_{i=1}^n g(y_i | \boldsymbol{\vartheta})$ represent the prior distribution and the likelihood function $\boldsymbol{\vartheta}$, respectively. The ratio of the posterior distributions is given by

$$\begin{aligned} \frac{\pi(\boldsymbol{\vartheta} | \mathcal{D}^{(-i)})}{\pi(\boldsymbol{\vartheta} | \mathcal{D})} &= \frac{\pi(\boldsymbol{\vartheta}) \prod_{j \in \mathcal{D}^{(i)}} g(y_j | \boldsymbol{\vartheta})}{\int_{\boldsymbol{\vartheta} \in \Theta} \pi(\boldsymbol{\vartheta}) \prod_{j \in \mathcal{D}^{(i)}} g(y_j | \boldsymbol{\vartheta}) d\boldsymbol{\vartheta}} \times \frac{\int_{\boldsymbol{\vartheta} \in \Theta} \pi(\boldsymbol{\vartheta}) \prod_{j \in \mathcal{D}} g(y_j | \boldsymbol{\vartheta}) d\boldsymbol{\vartheta}}{\pi(\boldsymbol{\vartheta}) \prod_{j \in \mathcal{D}} g(y_j | \boldsymbol{\vartheta})} \\ &= \frac{1}{g(y_i | \boldsymbol{\vartheta})} \times \frac{\int_{\boldsymbol{\vartheta} \in \Theta} \pi(\boldsymbol{\vartheta}) \prod_{j \in \mathcal{D}} g(y_j | \boldsymbol{\vartheta}) d\boldsymbol{\vartheta}}{\int_{\boldsymbol{\vartheta} \in \Theta} \frac{1}{g(y_i | \boldsymbol{\vartheta})} \pi(\boldsymbol{\vartheta}) \prod_{j \in \mathcal{D}} g(y_j | \boldsymbol{\vartheta}) d\boldsymbol{\vartheta}} \\ &= \frac{\left(\int_{\boldsymbol{\vartheta} \in \Theta} \frac{1}{g(y_i | \boldsymbol{\vartheta})} \pi(\boldsymbol{\vartheta} | \mathcal{D}) d\boldsymbol{\vartheta} \right)^{-1}}{g(y_i | \boldsymbol{\vartheta})} = \frac{CPO_i}{g(y_i | \boldsymbol{\vartheta})} \quad \square \end{aligned}$$

From Proposition 2, the K-L divergence can be expressed by

$$\begin{aligned}
 D_{K-L}(P, P_{(-i)}) &= -E_{\boldsymbol{\vartheta}|\mathcal{D}} \{ \log(CPO_i) \} + E_{\boldsymbol{\vartheta}|\mathcal{D}} \{ \log [g(y_i|\boldsymbol{\vartheta})] \} \\
 &= -\log(CPO_i) + E_{\boldsymbol{\vartheta}|\mathcal{D}} \{ \log [g(y_i|\boldsymbol{\vartheta})] \}
 \end{aligned}
 \tag{16}$$

From (15) we can compute $D_\psi(P, P_{(-i)})$ by sampling from the posterior distribution of $\boldsymbol{\vartheta}$ via MCMC methods. Let $\boldsymbol{\vartheta}^{(1)}, \dots, \boldsymbol{\vartheta}^{(Q)}$ be a sample of size Q of $\pi(\boldsymbol{\vartheta}|\mathcal{D})$. Then, a Monte Carlo estimate of $K(P, P_{(-i)})$ is given by

$$\widehat{D}_\psi(P, P_{(-i)}) = \frac{1}{Q} \sum_{q=1}^Q \psi \left(\frac{\widehat{CPO}_i}{g(y_i|\boldsymbol{\vartheta}^{(q)})} \right)
 \tag{17}$$

From (17) a Monte Carlo estimate of K-L divergence $D_{K-L}(P, P_{(-i)})$ is given by

$$\widehat{D}_{K-L}(P, P_{(-i)}) = -\log(\widehat{CPO}_i) + \frac{1}{Q} \sum_{q=1}^Q \log [g(y_i|\boldsymbol{\vartheta}^{(q)})]
 \tag{18}$$

$D_\psi(P, P_{(-i)})$ can be interpreted as the ψ -divergence of the effect of deleting the i th case from the full data on the joint posterior distribution of $\boldsymbol{\vartheta}$. As pointed by Peng and Dey (1995) and Weiss (1996) (see also Cancho et al., 2010, 2011a), it may be difficult for a practitioner to judge the cutoff point of the divergence measure so as to determine whether a small subset of observations is influential or not. In this context, we will use the proposal given by Peng and Dey (1995) and Weiss (1996) by considering as follows. Consider a biased coin which has success probability p . Then the ψ -divergence between the biased and an unbiased coin is

$$D_\psi(f_0, f_1) = \int \psi \left(\frac{f_0(x)}{f_1(x)} \right) f_1(x) dx
 \tag{19}$$

where $f_0(x) = p^x(1-p)^{1-x}$ and $f_1(x) = 0.5, x = 0, 1$. Now if $D_\psi(f_0, f_1) = d_\psi(p)$ then it can be easily checked that d_ψ satisfies the following equation:

$$d_\psi(p) = \frac{\psi(2p) - \psi(2(1-p))}{2}
 \tag{20}$$

It is not difficult to see for the divergence measures considered, that is d_ψ increases as p moves away from 0.5. In addition, $d_\psi(p)$ is symmetric about $p = 0.5$ and d_ψ achieves its minimum at $p = 0.5$. In this point, $d_\psi(0.5) = 0$, and $f_0 = f_1$. Therefore, if we consider $p > 0.75$ (or $p \leq 0.25$) as a strong bias in a coin, and since $d_{L_1}(0.75) = d_{\chi^2}(0.75) = 0.25$, it implies that i th case is considered influential when $d_{L_1}(0.75) > 0.25$ or $d_{\chi^2} > 0.25$. Thus, if we use the K-L divergence, we can consider an influential observation when $d_{K-L} > 0.14$. Similarly, using the J-distance, an observation in which $d_J > 0.27$ can be considered as influential. \square

4. Simulation study

A simulation study was designed with two objectives in mind, the first being to evaluate the frequentist properties of the parameter estimates for the proposed models and a misspecification study was also performed in order to verify if we can distinguish between the cure models given in Table 2 in light of a data set based on the criterion described in Section 3. The second being to examine the performance of the proposed diagnostics measures, we considered simulated data sets with one or more of the generated cases perturbed. In this study we considered the logarithmic cure rate model under three (first, last, and random) activation schemes as given in Table 2 with the Weibull distribution for the event times (Z) with parameters $\gamma_1 = 2$

and $\gamma_2 = -3.2$. More specifically, for each individual $i, i = 1, \dots, n$, number of causes of the event of interest for this individual (M_i) is generated of distribution Logarithmic with parameter θ_i equal to $p_{0i} W(-e^{-1/p_{0i}}/p_{0i})$, where $W(\cdot)$ is the Lambert W function (Corless et al., 1996). In our simulations we have one binary covariate x with values drawn from Bernoulli distribution with parameter 0.5. In this way, $p_{0i} = \exp(\beta_0 + \beta_1 x_i)/(1 + \exp(\beta_0 + \beta_1 x_i))$. We took $\beta_0 = -0.5$ and $\beta_1 = 0.7$, so that the cured fraction for the two levels of x is $p_0^{(0)} = 0.378$ and $p_0^{(1)} = 0.550$, respectively. The censoring times were sampled from the uniform distribution on the interval $(0, \tau)$, where τ controlled the proportion of censoring of the uncured population. In this study the proportion of censored observation is taken approximately to be equal to 57%.

4.1. Frequentist properties

The goal of this simulation study is to show the good behavior of the Bayesian estimates, based on the frequentist mean squared error (MSE) and the frequentist mean (Mean), and the measurement used for model comparison. We performed the simulation generating samples of size $n = 300$ for the three different structures of activation. Therefore, three different simulation settings are performed, each one with 500 Monte Carlo-generated data sets, and once the data are simulated we fit the three models. The following independent priors are considered to perform the Metropolis–Hastings algorithm, for $\beta_j \sim N(0, 10^4) \ j = 0, 1, \gamma_1 \sim \text{Gamma}(1, 0.1)$ and $\gamma_2 \sim N(0, 10^4)$. Thus, our choice is to assume weakly but informative prior. Because our prior is still informative the posterior is always proper. After 20,000 sample burn-in, we used every tenth sample from the 40,000 MCMC posterior samples to reduce the autocorrelations and yield better convergence results. We monitored the convergence of the Metropolis–Hasting algorithm using the method proposed by Geweke (1992), as well as trace plots.

For each sample the posterior mean of the parameter and $DIC, EAIC, EBIC$, and $LPML$ are recorded. Simulation summary statistics for the parameters assuming the models described in Table 2, crossed with the three schemes of activation simulated, are given in Table 3. In this table, MC Mean denotes the arithmetic average of the 500 estimates given by $\sum_{j=1}^{500} \hat{\theta}_{kj}/500$, and MC MSE is the empirical mean squared error given by $\sum_{j=1}^{500} (\hat{\theta}_{kj} - \theta_k)^2/500$, where θ_k is the respective true value of the estimated parameter $\hat{\theta}_k$. Notice from Table 3 that the MC Mean values are very close to the true value and MC MSE is small when the fit was performed to the true activation scheme.

Table 3. Monte Carlo results based on each 300 simulated samples crossed with the three schemes of activation. MC mean and MC MSE are the mean average and the MSE average of the posterior mean estimates of cured fraction, respectively.

True Model	Cured Fraction	Fitted model					
		First activation		Last activation		Random activation	
		MC Mean	MC MSE	MC Mean	MC MSE	MC Mean	MC MSE
First	$p_0^{(0)}$	0.3687	0.0020	0.4464	0.0062	0.3598	0.0030
	$p_0^{(1)}$	0.5392	0.0030	0.5105	0.0072	0.5641	0.0026
Last	$p_0^{(0)}$	0.3615	0.0028	0.3628	0.0039	0.1418	0.0841
	$p_0^{(1)}$	0.3769	0.0341	0.5320	0.0052	0.2585	0.1323
Random	$p_0^{(0)}$	0.3965	0.0030	0.4235	0.0041	0.3565	0.0070
	$p_0^{(1)}$	0.4905	0.0074	0.5320	0.0053	0.5351	0.0057

Table 4. Percentages of samples in which the fitted model was indicated as the best one according to the *DIC*, *EAIC*, *EBIC*, and *LPML* criteria.

True model	Fitted model		
	First activation	Last activation	Random activation
First	89.8	10.0	0.2
Last	10.8	81.2	8.0
Random	23.8	20.0	56.2

Moreover, presented in Table 4 are the percentages of samples in which the generating distribution was indicated as the best one according to the *DIC*, *EAIC*, *EBIC*, and *LPML* criteria. We can observe that the true model from which the sample was generated has a higher percentage.

4.2. Influence of outlying observations

To examine the performance of the proposed diagnostic measures, we considered simulated data sets with one or more of the generated cases perturbed. In order to do this, we consider a sample of size 300 generated by the logarithmic cure rate model under the first activation scheme. In the simulated data, y_i ranged 0.07824–12.870 with median = 2.97, mean = 4.22, and standard deviation = 3.50. We selected cases 90, 100, and 285 for perturbation. To create influential observation in the data set, we choose one, two, or three of these selected cases and perturb the response variable as follows: $\tilde{y}_i = y_i + 4S_y$, $i = 4, 13, \text{ and } 45$, where S_y is the standard deviation of the y_i 's. The MCMC computations were done similar to those in Section 4.1 and further to monitor the convergence of the Gibbs samples we used the methods recommended by Cowles and Carlin (1996).

We fit the logarithmic cure rate models under the last and random activation schemes. Table 5 shows that the posterior inferences to cured fraction are not sensitive to the perturbation of the selected case(s). In Table 5, data set (a) denotes the original simulated data set with no perturbation and data sets (b)–(f) denote data sets with perturbed cases.

In Table 6 we report the Monte Carlo estimates of *DIC*, *EAIC*, *EBIC*, and *LPML* for each perturbed version of the original data set under first, last, and random schemes of activation. We can see that the logarithmic cure rate model under the first activation scheme presents a better fit in all cases.

Now we consider the sample from the posterior distributions of the parameters of the logarithmic cure rate model under the first activation scheme to calculate the ψ -divergence measures in (14) described in Section 3.3. The results in Table 7 show, before perturbation (data

Table 5. Mean and standard deviation (SD) for cured fraction estimates for each data set for fitting logarithmic cure rate models under the first, last, and random activation schemes.

Data set names	Perturbed case	First		Last				Random					
		$p_0^{(0)}$		$p_0^{(1)}$		$p_0^{(0)}$		$p_0^{(1)}$		$p_0^{(0)}$		$p_0^{(1)}$	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
a	None	0.393	0.042	0.573	0.050	0.452	0.039	0.558	0.054	0.377	0.050	0.593	0.050
b	90	0.366	0.042	0.529	0.057	0.440	0.043	0.549	0.056	0.365	0.048	0.584	0.054
c	100	0.363	0.052	0.524	0.071	0.441	0.042	0.553	0.053	0.357	0.049	0.580	0.053
d	285	0.358	0.042	0.536	0.058	0.448	0.041	0.539	0.055	0.361	0.051	0.580	0.052
e	{90, 100}	0.318	0.080	0.449	0.117	0.430	0.042	0.545	0.057	0.346	0.051	0.569	0.055
f	{90, 100, 285}	0.254	0.081	0.364	0.128	0.429	0.044	0.526	0.066	0.311	0.066	0.542	0.062

Table 6. Comparison between logarithmic cure rate models under the first, last, and random activation schemes fitting by using different Bayesian criteria.

Data set names	First				Last				Random			
	DIC	EAIC	EBIC	LPML	DIC	EAIC	EBIC	LPML	DIC	EAIC	EBIC	LPML
a	730.0	733.9	748.7	−364.8	747.5	751.7	766.5	−373.7	736.1	739.9	754.8	−367.9
b	747.1	751.4	766.2	−374.7	766.2	770.4	785.2	−383.8	754.4	758.4	773.2	−378.6
c	748.3	752.1	767.0	−375.7	766.7	770.7	785.5	−384.4	754.9	758.8	773.6	−378.4
d	746.7	750.7	765.5	−374.5	768.1	772.2	787.0	−385.0	755.2	759.1	773.9	−379.2
e	758.0	763.8	778.6	−381.0	781.0	785.1	799.9	−391.6	767.8	771.9	786.7	−384.6
f	762.2	770.0	784.8	−383.5	794.7	798.8	813.6	−398.8	779.4	783.1	798.0	−390.7

set (a)), that all the selected cases are not influential according to all ψ -divergence measures. However, after perturbation (data set, (b)–(f)), the measures increase indicating that the perturbed cases are influential.

In Figures 2 and 3, we have depicted the four ψ -divergence measures for the cases (a) and (b), respectively. Clearly we can see that all measures performed well in identifying influential case(s), providing larger ψ -divergence measures when compared to the other cases.

5. Malignant melanoma data

In this section we work out an example extracted from Scheike (2009) employing the models presented in Section 2. The data set includes 205 patients observed after operation for removal of malignant melanoma. The observed time (T) ranges from 10 to 5565 days (from 0.0274 to 15.25 years, with mean = 5.9 and standard deviation = 3.1 years) and refers to the time until the patient’s death or the censoring time. Patients dead from other causes, as well as patients still alive at the end of the study, are censored observations (72%). We take ulceration status (x_1) (absent, $n = 115$; present, $n = 90$) and tumor thickness (x_2) (in mm, mean = 2.92 and standard deviation = 2.96) as covariates. The Kaplan–Meier estimate of the surviving function given in Figure 4 levels off above 0.6. The presence of a plateau indicates that models that ignore the possibility of cure will not be suitable for these data.

First we fitted the models described in Table 2. For all models the following independent priors were adopted in the Bayesian computations $\beta_j \sim N(0, 10^4)$ $j = 0, 1, 2$, $\gamma_1 \sim \text{Gamma}(1, 0.1)$ and $\gamma_2 \sim N(0, 10^4)$. A total of 40,000 MCMC posterior samples were used

Table 7. ψ -Divergence measures for the simulated data fitting the logarithmic cure rate model under the first activation scheme.

Data set names	Case number	d_{K-L}	d_j	d_{L_1}	d_{χ^2}
a	90	0.007	0.015	0.050	0.016
	100	0.007	0.013	0.047	0.014
	285	0.016	0.032	0.072	0.032
b	90	1.569	3.417	0.657	23.309
	100	2.016	4.176	0.720	27.023
d	285	1.476	3.060	0.635	11.407
	e	90	0.661	1.446	0.453
100		0.744	1.639	0.478	5.497
f	90	0.273	0.565	0.311	0.788
	100	0.324	0.671	0.338	0.986
	285	0.333	0.690	0.344	1.002

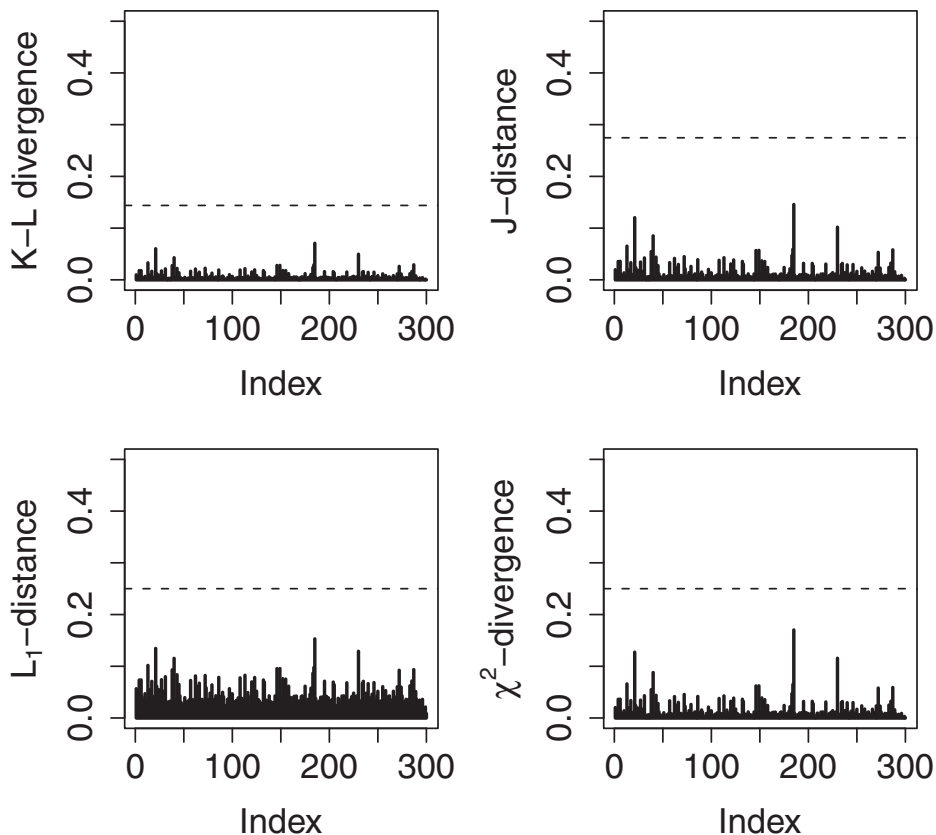


Figure 2. ψ -divergence measures from data set (a).

in this analysis after burn-in. MCMC computations. We used every tenth sample from the 40,000 MCMC posterior samples to reduce the autocorrelations and yield better convergence results.

To compare the logarithmic cure rate model under the first, last and random activation scheme fits, we also obtained the values of *DIC*, *EAIC*, *EBIC*, and *LPML*. These information criteria furnish the values given in [Table 8](#). According to all the criteria, the logarithmic cure rate model under the first activation scheme (Log-1st, say) stands out as the best one. Taking into account the criteria in [Table 8](#), we select the Log-1st model as our working model.

The posterior means, medians, standard deviations, and 95% highest posterior density (HPD) intervals are shown in [Table 9](#). The mean posterior of the shape parameter (γ_1) furnishes an evidence against the exponential distribution ($\gamma_1 = 1$) for the unobserved failure times. The covariates have a significant effect on the reduction of the cured fraction.

We now consider the sample from the posterior distributions of the parameters of the Log-1st models to calculate the ψ -divergence measures in (14) described in [Section 3.3](#). [Figure 5](#) shows the index plot of the four ψ -divergence measures, where we observed that cases 5 and 171 are possible influential observations in the posterior distribution.

[Table 10](#) presents the relative changes (RCs) in the estimates after dropping one of the two cases without standing influence and also when both are dropped at once (represented by the set $I = \{5, 171\}$). The 95% HPD intervals for the new estimates are given in parentheses.

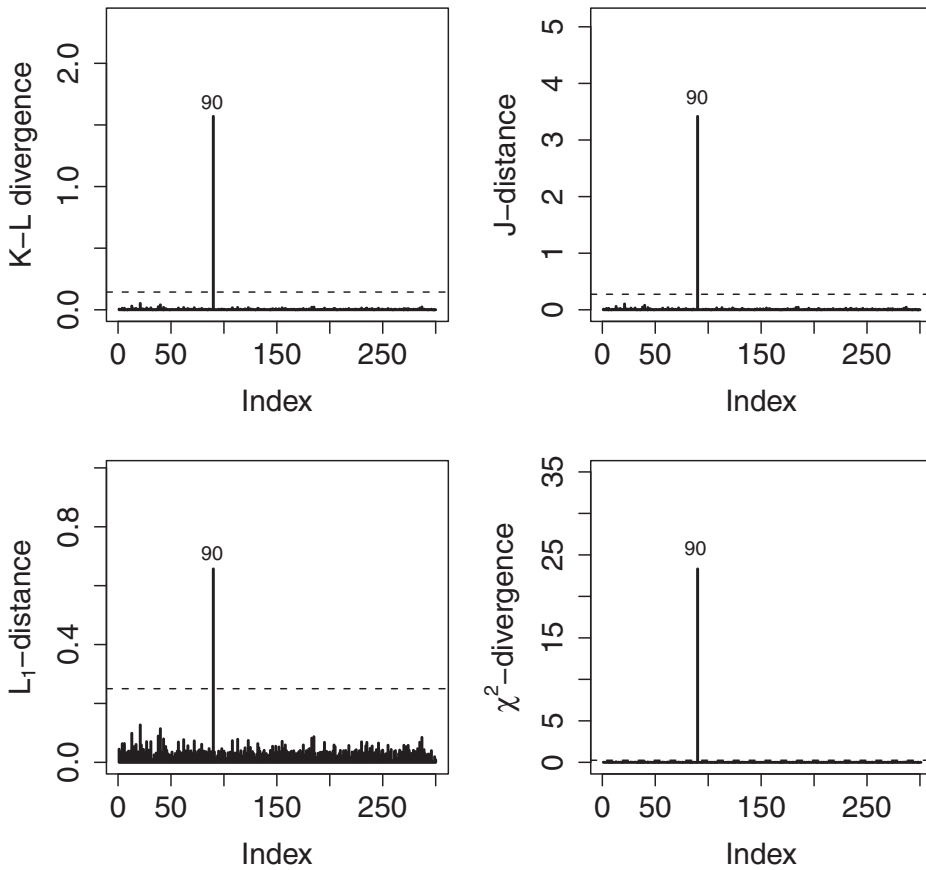


Figure 3. ψ -divergence measures from data set (b).

The RC (in percentage) of each estimated parameter is defined by $RC_{\theta_j} = |(\hat{\theta}_j - \hat{\theta}_{j(I)})/\hat{\theta}_j| \times 100\%$, where $\hat{\theta}_{j(I)}$ denotes the posterior mean of θ_j , with $j = 1, \dots, 5$, after the set I of observations has been removed. We notice that there are little changes in the inference for the coefficients. Particularly, β_0 is not significant at 5% after dropping the observation 5. In [Table 10](#)

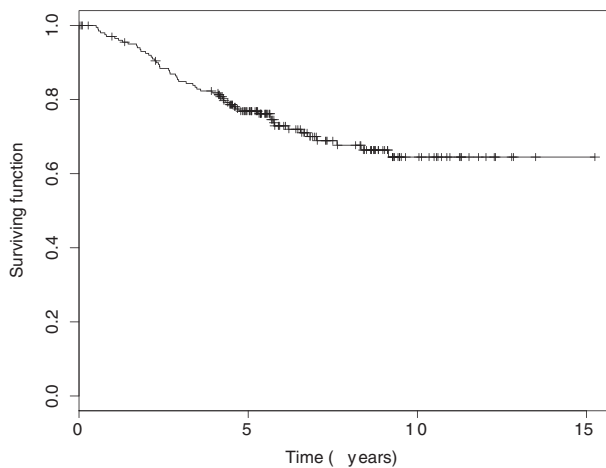


Figure 4. Kaplan–Meier estimate of the surviving function.

Table 8. Bayesian criteria for the fitted models.

Activation	Criterion			
	LPLM	DIC	EAIC	EBIC
First	-172.34	344.67	351.16	367.78
Last	-192.22	388.39	391.00	407.611
Random	-190.42	380.52	383.89	400.51

Table 9. Posterior summaries of the parameters for the Log-1st model.

Parameter	Mean	Median	Standard deviation	HPD interval (95%)	
				LI	LS
γ_1	2.115	2.106	0.2619	1.640	2.6542
γ_2	-4.500	-4.339	0.8969	-7.092	-3.2786
β_0	1.353	1.493	0.6471	-0.438	2.2654
β_1	-0.125	-0.125	0.0395	-0.203	-0.0477
β_2	-1.405	-1.392	0.3522	-2.101	-0.6882

we applied the selection criteria on each fitting. We observe a better fit when we drop the observation 171 compared with the fitting when we drop the observation 5, according to all criteria. Furthermore, when the observation 171 is removed all the regression coefficients are significant.

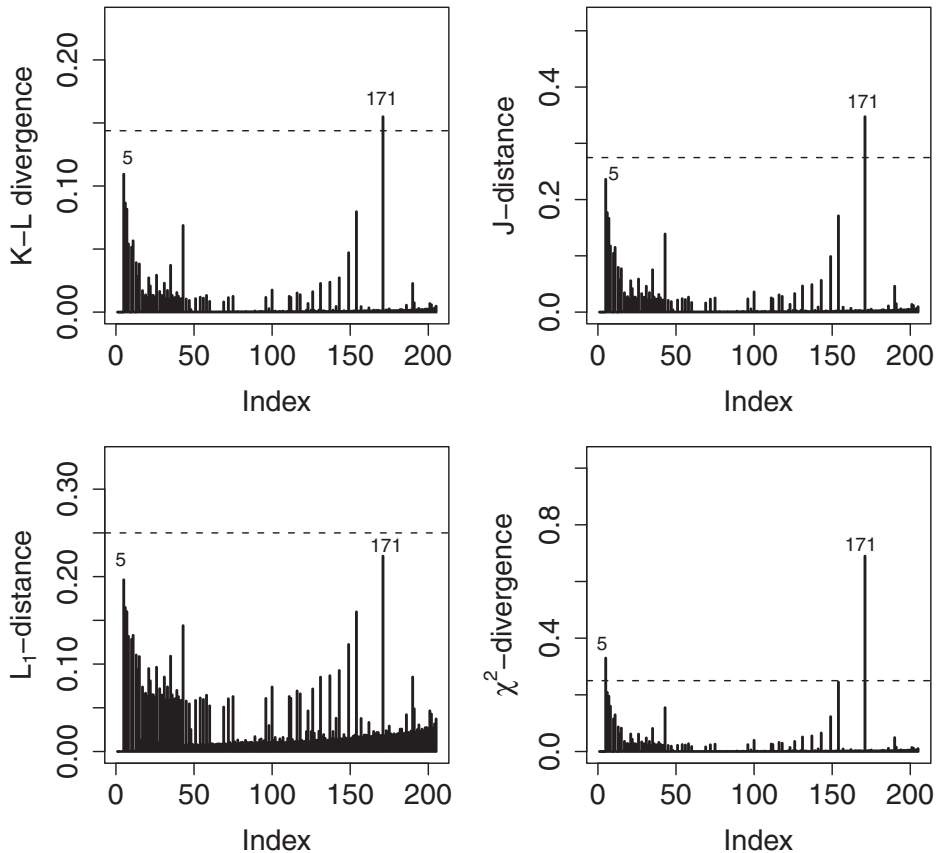


Figure 5. Index plots of ψ -divergence measures for the melanoma data.

Table 10. RCs (in %) and the corresponding 95% highest posterior density (HPD) intervals for fitting the Log-1st model to the melanoma data dropped influence case.

Model	Parameter	Dropped observation		
		5	171	Set I
Log-1st	γ_1	2.37 (1.610, 2.662)	0.93 (1.653, 2.719)	1.02 (1.668, 2.701)
	γ_2	0.75 (-7.007, -3.291)	3.49 (-6.191, -3.198)	5.64 (-5.602, -3.159)
	β_0	8.40 (-0.427, 2.272)	7.24 (0.012, 2.384)	8.73 (0.391, 2.376)
	β_1	13.08 (-0.194, -0.035)	3.08 (-0.212, -0.055)	7.69 (-0.202, -0.039)
	β_2	3.70 (-2.086, -0.709)	0.27 (-2.115, -0.787)	0.62 (-2.151, -0.803)
	LPML	-171.222	-167.895	-166.418
	DIC	343.382	337.419	333.752

We turn our attention to the role of the covariates on the cured fraction p_0 . Table 11 shows the posterior summaries of the cured fraction stratified by ulceration status with tumor thickness equal to 0.32, 1.94, and 8.32 mm, which correspond to the 5, 50, and 95 percentiles, under the Log-1st model with and without 171st observation. Clearly, the withdrawal of the influential observation 171 leads to an increase in the cured fraction. Figure 6 displays the surviving function (without 171 observation) stratified by ulceration status for patients with tumor thickness equal to 0.64, 1.94, and 8.32 mm. These plots highlight the combined impact of the covariates on the cured fraction. For each selected value of tumor thickness the intervals do not overlap.

We end up our application dealing with the estimation of the proportion of patients non cured who survived beyond a certain fixed time. For illustration we choose 5 years. This proportion is estimated from $S_{nc}(5)$ (model under the first activation scheme) given in Table 1. Considering posterior samples of parameter of the Log-1st model without 171st observation, the Monte Carlo estimates of $S_{nc}(5)$ stratified by ulceration status (absent, present) for non cured patients with tumor thickness equal to 0.32, 1.94, and 8.32 mm (joint with standard deviation posterior) are: (0.546[0.139], 0.416[0.110]), (0.538[0.138], 0.391[0.105]), and (0.470[0.135], 0.252[0.0825]).

Table 11. Posterior summaries of the cured fraction stratified by ulceration status and selected tumor thickness under the Log-1st model.

Data set	Tumor thickness	Ulceration	Mean	Median	Standard deviation	95% HPD interval
Full (D)	0.64	Absent	0.766	0.803	0.124	(0.382, 0.898)
		Present	0.470	0.485	0.112	(0.206, 0.651)
	1.94	Absent	0.739	0.775	0.124	(0.361, 0.879)
		Present	0.432	0.444	0.102	(0.192, 0.598)
	8.32	Absent	0.575	0.589	0.125	(0.261, 0.770)
		Present	0.257	0.257	0.065	(0.129, 0.388)
D-171	0.64	Absent	0.805	0.829	0.100	(0.491, 0.908)
		Present	0.512	0.524	0.104	(0.243, 0.676)
	1.94	Absent	0.780	0.803	0.101	(0.467, 0.889)
		Present	0.470	0.481	0.096	(0.223, 0.622)
	8.32	Absent	0.611	0.624	0.107	(0.331, 0.776)
		Present	0.276	0.277	0.063	(0.145, 0.400)

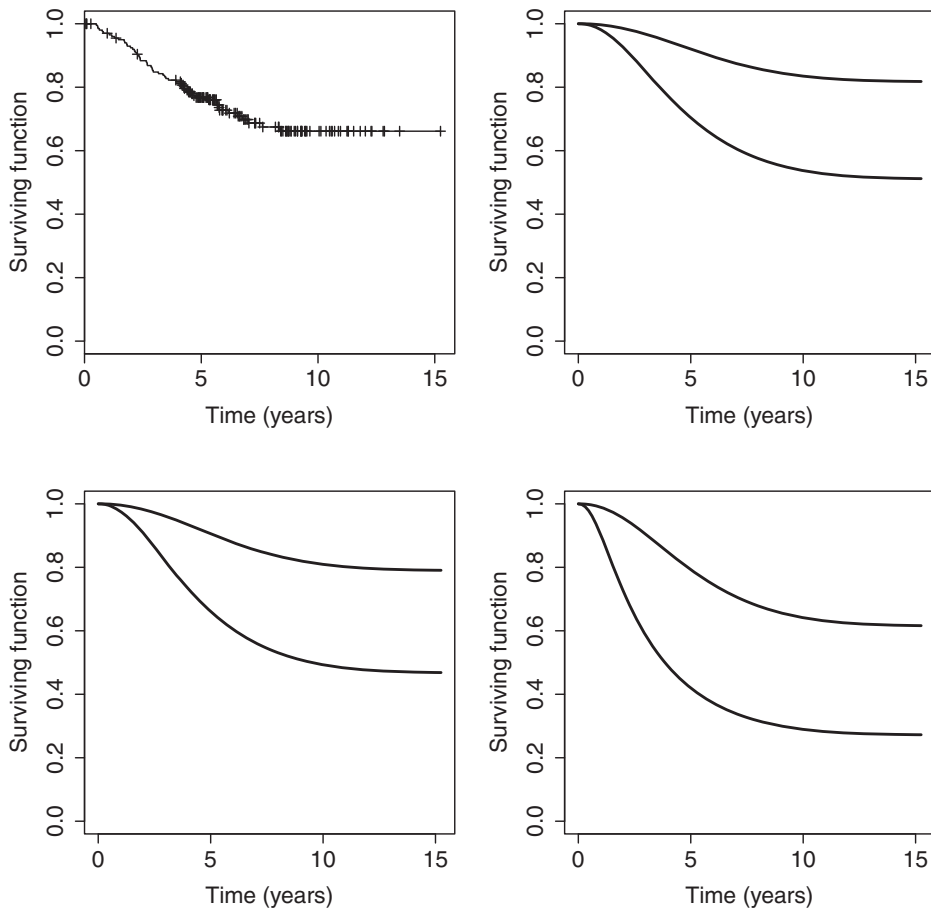


Figure 6. Upper left panel: Kaplan–Meier estimate of the surviving function. Surviving function under the Log-1st model (without 171 observation) stratified by ulceration status (upper: absent, lower: present) for patients with tumor thickness equal to 0.64 (upper right panel), 1.94 (lower left panel), and 8.32 mm (lower right panel).

6. Conclusions

In this paper we proposed the logarithmic cure rate model under different activations as an alternative model for modeling aging data with a cured fraction. The model can be tested for the best fitting in a straightforwardly way. Moreover, we propose an influence diagnostic approach from the Bayesian point of view, based on the ψ -divergence (Peng and Dey, 1995; Weiss, 1996) between the posterior distributions of the parameters of the proposed model. In the application to a melanoma data set, we discovered that the logarithmic cure rate model under the first activation scheme delivers the best fit. We observed that the surviving probability decreases more rapidly for patients with thicker tumors, and that the cured fraction is lower for patients with ulceration. The interpretation of the role of covariates is easy due to the parameterization in the cured fraction.

Although we assume M as logarithmic distributed, we believe that, at least in principle, the methodology presented here may be extended by considering other distributions, such as the negative binomial and power series ones, among others.

References

- Berkson, J., Gage, R.P. (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. B* 11:15–53.
- Brooks, S.P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde (2002). *J. Royal Stat. Soc. B* 64:616–618.
- Cancho, V., Ortega, E., Paula, G. (2010). On estimation and influence diagnostics for log-Birnbaum-Saunders student-t regression models: Full Bayesian analysis. *J. Stat. Plann. Inference* 140(9):2486–2496.
- Cancho, V., Dey, D., Lachos, V., Andrade, M. (2011a). Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Comput. Stat. Data Anal.* 55(1):588–602.
- Cancho, V., Rodrigues, J., de Castro, M. (2011b). A flexible model for survival data with a cure rate: A Bayesian approach. *J. Appl. Stat.* 38(1):57–70.
- Carlin, B.P., Louis, T.A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*. second edition. Boca Raton, FL: Chapman & Hall/CRC.
- Chen, M.-H., Ibrahim, J.G. (2001). Maximum likelihood methods for cure rate models with missing covariates. *Biometrics* 57(1):43–52.
- Chen, M.-H., Ibrahim, J.G., Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *J. Am. Stat. Assoc.* 94:909–919.
- Chen, M.H., Ibrahim, J.G., Sinha, D. (2002). Bayesian inference for multivariate survival data with cure fraction. *J. Multivariate Anal.* 89:101–126.
- Ciumara, R., Preda, V. (2009). The Weibull-logarithmic distribution in lifetime analysis and its properties. In *The XIIIth International Conference Applied Stochastic Models and Data Analysis (ASMDA-2009)*, pages 395–399. Institute of Mathematics and Information/Vilnius Gediminas Technical University, Vilnius, Lithuania.
- Cook, R.D. (1986). Assessment of local influence. *J. Royal Stat. Soc. Ser. B.* 48:133–169.
- Cook, R.D., Weisberg, S. (1982). *Residuals and Influence in Regression*. Boca Raton, FL: Chapman & Hall/CRC.
- Cooner, F., Banerjee, S., McBean, A.M. (2006). Modelling geographically referenced survival data with a cure fraction. *Stat. Methods Med. Res.* 15:307–324.
- Cooner, F., Banerjee, S., Carlin, B.P., Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *J. Am. Stat. Assoc.* 102:560–572.
- Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E. (1996). On the Lambert W function. *Adv. Comput. Math.* 5:329–359.
- Cowles, M.K., Carlin, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Am. Stat. Assoc.* 91:883–904.
- Cox, D., Oakes, D. (1984). *Analysis of Survival Data*. London, New York: Chapman & Hall.
- de Castro, M., Cancho, V.G., Rodrigues, J. (2009). A Bayesian long-term survival model parametrized in the cured fraction. *Biomet. J.* 51:443–455.
- de Castro, M., Cancho, V.G., Rodrigues, J. (2010). A hands-on approach for fitting long-term survival models under the GAMLSS framework. *Comput. Methods Prog. Biomed.* 97:168–177.
- Dey, D., Birmiwál, L. (1994). Robust Bayesian analysis using divergence measures. *Stat. Probab. Lett.* 20(4):287–294.
- Farewell, V. (1982). The use of mixture models for the analysis of survival data with long term survivors. *Biometrika* 38:1041–1046.
- Farewell, V. (1986). Mixture models in survival analysis: Are they worth the risk? *Can. J. Stat.* 14:257–262.
- Gamerman, D., Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, second edition. Boca Raton, FL: Chapman & Hall/CRC.
- Geisser, S., Eddy, W. (1979). A predictive approach to model selection. *J. Am. Stat. Assoc.* 74(365):153–160.

- Gelfand, A.E., Dey, D.K., Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Smith, A.F.M. eds. *Bayesian Statistics 4*. Oxford University Press, Oxford, London pp. 147–167.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J. M., Berger, J.O., Dawid, A.P., and Smith, A.F.M., eds. *Bayesian Statistics 4*. Oxford: Oxford University Press. Vol. 1(4).
- Ghitany, M., Maller, R.A., Zhou, S. (1994). Exponential mixture-models with long-term survivors and covariates. *J. Multivariate Anal.* 49:218–241.
- Gordon, N.H. (1990). Application of the theory of finite mixtures for the estimation of ‘cure’ rates of treated cancer patients. *Stat. Med.* 9:397–407.
- Ibrahim, J.G., Chen, M.-H., Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer.
- Johnson, N.L., Kemp, A.W., Kotz, S. (2005). *Univariate Discrete Distributions*, third edition. Hoboken, NJ: Wiley.
- Kim, S., Chen, M., Dey, D. (2011). A new threshold regression model for survival data with a cure fraction. *Lifetime Data Anal.* 17(1):101–122.
- Li, C.S., Taylor, J.M., Sy, J.P. (2001). Identifiability of cure models. *Stat. Probab. Lett.* 54:389–395.
- Louzada, F., Suzuki, A.K., Cancho, V.G. (2013). The fgm long-term bivariate survival copula model: Modeling, Bayesian estimation, and case influence diagnostics. *Commun. Stat. – Theory. Methods* 42(4):673–691.
- Louzada-Neto, F. (1999). Poly-hazard regression models for lifetime data. *Biometrics* 55:1121–1125.
- Maller, R.A., Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics* 51(4):1197–1205.
- Ortega, E.M.M., Cancho, V.G., Paula, G.A. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Anal.* 15:79–106.
- Peng, F., Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *Can. J. Stat.* 23(2):199–213.
- Perdona, G., Louzada-Neto, F. (2011). A general hazard model for lifetime data in the presence of cure rate. *J. Appl. Stat.* 38:1395–1405.
- Perperoglou, A., Antonis, K., Houwelingen, H.C.v. (2007). Approaches in modelling long-term survival: An application to breast cancer. *Stat. Med.* 26:2666–2685.
- Scheike, T. (2009). *Timereg Package*. R package version 1.1-0. With contributions from T. Martinussen and J. Silver. R package version 1.1-6.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Royal Stat. Soc. B* 64:583–639.
- Sposto, R. (2002). Cure model analysis in cancer: An application to data from the children’s cancer group. *Stat. Med.* 21:293–312.
- Tahmasbi, R., Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Comput. Stat. Data Anal.* 52(8):3889–3901.
- Tournoud, M., Ecohard, R. (2007). Application of the promotion time cure model with time-changing exposure to the study of HIV/AIDS and other infectious diseases. *Stat. Med.* 26:1008–1021.
- Tsodikov, A.D., Ibrahim, J.G., Yakovlev, A.Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *J. Am. Stat. Assoc.* 98:1063–1078.
- Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *J. Royal Stat. Soc. Ser. B (Methodol.)* 739–750.
- Yakovlev, A.Y., Tsodikov, A.D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. Singapore: World Scientific.