Test (2012) 21:355–368
DOI 10.1007/s11749-011-0247-1

ORIGINAL PAPER

A multiple time scale survival model with a cure fraction

Francisco Louzada · Juliana Cobre

Received: 30 April 2009 / Accepted: 19 April 2011 / Published online: 25 May 2011
© Sociedad de Estadística e Investigación Operativa 2011

Abstract Many recent survival studies propose modeling data with a cure fraction, i.e., data in which part of the population is not susceptible to the event of interest. This event may occur more than once for the same individual (recurrent event). We then have a scenario of recurrent event data in the presence of a cure fraction, which may appear in various areas such as oncology, finance, industries, among others. This paper proposes a multiple time scale survival model to analyze recurrent events using a cure fraction. The objective is analyzing the efficiency of certain interventions so that the studied event will not happen again in terms of covariates and censoring. All estimates were obtained using a sampling-based approach, which allows information to be input beforehand with lower computational effort. Simulations were done based on a clinical scenario in order to observe some frequentist properties of the estimation procedure in the presence of small and moderate sample sizes. An application of a well-known set of real mammary tumor data is provided.

Keywords Cure fraction modeling · Berkson–Gage model · Recurrent events · Bayesian approach

Mathematics Subject Classification (2000) 62N99

1 Introduction

Models which consider that part of the population may not become susceptible to a certain event of interest have been widely developed recently. These models are

Communicated by: Domingo Morales.

F. Louzada (✉) · J. Cobre
SME-ICMC, Universidade de São Paulo, Av. Trabalhador São-Carlense, 400, CP. 668, São Carlos, SP 13565-905, Brazil
e-mail: louzada@icmc.usp.br

called cure fraction models or long-term survival models and were firstly proposed by Boag (1949) and Berkson and Gage (1952). Evidence that a cure fraction exists in the studied population may be given by a large percentage of censoring in the data, as in traditional survival studies, individuals that do not present the event of interest until a certain moment are censored (Goldman 1984), ignoring the possibility that the individual is not present in the event of interest.

In many cases, the event of interest may occur more than once with the same individual. Survival data with recurrent events exist in many areas, such as oncology studies with a cancerous tumor reappearing, geological studies with the recurrence of a seismic shock, or in industrial studies with recurrent failures in a component. Cox (1972) recurrent event modeling considers the total time of study and the time between occurrences. A Poisson process representation is considered by Lawless (2003), which models the total time of study. Lawless and Thiagarajah (1996) considered modulated renewal and Poisson processes, which accommodate the two time scales, focusing on estimation for a specific model. McDonald and Rosina (2001) and Yu (2008) consider mixed models to analyze situations in which the recurrence of the event and the possibility of cure can be observed. Yu (2008) analyzes the possibility of curing patients with colorectal cancer after certain surgical intervention.

In this paper, we propose a multiple time scale survival model (MSS) combined with the Berkson and Gage (1952) framework for modeling recurrent event data with a cure fraction. The main idea is to extend the modeling presented in Louzada-Neto and Cobre (2010) and Cobre and Louzada-Neto (2009), allowing us to discover if the efficiency of certain intervention may lead to a possible cure, in terms of the covariates and censoring, which is impossible by considering the former modeling. The proposed model accommodates several models from the literature including the Poisson, renewal, and count models as special cases. The idea is to combine the two time scales, total time and interval time (intervals between successive events), and the event counts in a hybrid model and to decide their appropriateness according to the data. In the model version with covariates, we assume a proportional baseline function. We envisage applications in which a moderate or large number of individuals are observed and the number of events per individual may be quite small.

The paper is organized as follows. The proposed model is described in Sect. 2. The inferential procedure, model assessment, and simulation results based on a clinical study are presented in Sect. 3. A real data analysis on a mammary tumor data set is presented in Sect. 4. Final comments in Sect. 5 conclude the paper.

2 Model formulation

In traditional survival studies, individuals who do not present the event of interest until a certain time are considered censored. Therefore, censorship can mean that the individual was no longer susceptible to the event of interest. Thus, it seems appropriate to consider a model which includes heterogeneous characteristics, such as two distribution mixture models. In this modeling, one distribution represents the failure or survival times of individuals, which are susceptible to a certain event (at risk individuals, AR), while the other distribution represents the survival times of the individuals, which are not susceptible to the event (out of risk individuals, OR), with this

latter distribution allowing for infinite survival times (Maller and Zhou 1996). The term "long duration" refers to the individuals who are not susceptible to the event of interest. In the medical area, it is common to use the term "cured" to refer to the population who is no longer at risk. To simplify the language, we will use the term cured.

The mixture models were firstly approached by Boag (1949) and by Berkson and Gage (1952). Both pieces of research consider that the individual belongs or not to the group at risk with a certain probability. In terms of the survival function, we can write the mixture model as:

$$S_{\text{pop}}(t) = \rho + (1 - \rho)S(t), \tag{1}$$

where $S(t)$ is a proper survival function, i.e., $\lim_{t\to\infty} S(t) = 0$, and ρ is the probability of an individual not belonging to the group at risk. Consequently, we have $\lim_{t\to\infty} S_{\text{pop}}(t) = \rho$, and therefore the survival function (not conditional) is improper, and its limit corresponds to the individuals' proportion OR.

In the literature, there are various proposals for the survival time distribution of an individual AR. The Weibull family was approached by Farewell (1982), Ghitany and Maller (1992), Ghitany et al. (1994), Ng et al. (2004). Peng et al. (1998) propose the use of the generalized F distribution. In the case of a recurrent event, we can mention the nonhomogeneous Poisson process and renewal the pure process that depends on the total time of study (Lawless 2003, p. 532). The renewal process (Prentice et al. 1981) models the interval time between the various occurrences and the last occurrence. The semi-Markov process considers each occurrence as being a stratum whereby the individual remains in that stratum until the next occurrence or censoring happens (Prentice et al. 1981). Another class of models is the renewal Poisson process (Cox 1972) and the hybrid scale models (Louzada-Neto 2004, 2008; Louzada-Neto and Cobre 2010), which incorporate two time scales, total time and interval times. Cobre and Louzada-Neto (2009) develop a sampling based approach for a hybrid scale intensity model. Semi-parametric models were approached by Kuk and Chen (1992), Sy and Taylor (2000), and Yu (2008).

We propose an MSS model to describe the survival time distribution of the individuals AR, which combines the two time scales, total time and interval time, and the event counts.

2.1 Multiple time scale survival model

In the MSS, the total time modeling considers that the risk for the occurrence of each event begins at the same time. Interval time modeling, however, is appropriate for situations where the risk for the next event does not begin until after the previous event has occurred. The event counts can influence the risk of the event. The parameters of the model reflect the relative risk of the next event from the time of the previous event.

The data on the ith individual consists of the total number, m_i, of events (lifetimes) observed over the time period $(0, \tau_i]$ and the ordered epochs of the m_i lifetimes at times $0 \leq t_{i1} < t_{i2} < \cdots < t_{im_i} \leq \tau_i$. Then, we firstly define a partial survival function

given by

$$S_j(t_{i_j}|\cdot) = \exp\left\{-\int_{t_{i_{j-1}}}^{t_{i_j}} h_j(u|\cdot)\,du\right\}, \tag{2}$$

where h_j is an intensity function defined on $[t_{i_{j-1}}, t_{i_j})$. Then the overall survival function is given by

$$S(t_{m_i}|\cdot) = \exp\left\{-\sum_{j=1}^{m_i}\int_{t_{i_{j-1}}}^{t_{i_j}} h_j(u|\cdot)\,du\right\}. \tag{3}$$

We propose a multiple time scale survival model (MSS) which considers both interval and total times, and also the number of events for each individual by assuming that

$$h_j(t_{i_j}|\cdot) = q_1(x_{i_j};\theta_1)q_2(t_{i_j};\theta_2)q_3(j;\theta_3)g\left(\boldsymbol{\alpha}^T z_i\right), \tag{4}$$

where $q_1(\cdot)$, $q_2(\cdot)$, and $q_3(\cdot)$ are positive functions denoting the parametric functions on the interval time, t_{i_j} denotes the total time from the origin, $x_{i_j} = t_{i_j} - t_{i_{j-1}}$ is the backward recurrence time, j denotes the event counts with respectively unknown parameter vectors θ_1, θ_2, and θ_3, z_i is the covariates vector, $g(\cdot)$ is a known positive function that equals one when its argument is zero, and α is a vector of unknown regression parameters. The covariates are assumed to be fixed and therefore not affected by the event process.

Model (3) with (4) covers a wide spectrum of survival models, but comprehensive classes of model can be obtained as particular cases with a reduced number of parameters. For instance, (3) reduces to a general nonhomogeneous Poisson process model, in its parametric version, if $q_1(\cdot) = q_3(\cdot) = 1$. For $q_2(\cdot) = q_3(\cdot) = 1$, we obtain a renewal process model. A hybrid Poisson/renewal intensity model is obtained if $q_3(\cdot) = 1$.

In this paper we shall examine the fully parametric approach by considering a particular flexible parameterization of (4), where $q_1(x;\boldsymbol{\theta_1}) = q_1(x;\gamma) = \gamma x^{\gamma-1}$, $q_2(t;\boldsymbol{\theta_2}) = q_2(t;\phi) = (1+\phi t)$, $q_3(j;\boldsymbol{\theta_3}) = q_3(j;\psi) = \psi^{j-1}$, and $g(\boldsymbol{\alpha}^T z) = \exp(\boldsymbol{\alpha}^T z)$. From (2), (3), and (4), the partial survival function and overall survival function of our MSS model are given respectively by

$$S(t_{i_j}|\cdot) = \exp\left\{-\psi^{j-1}x_{i_j}^{\gamma}\left(1+\phi t_{i_{j-1}} + \phi\gamma\frac{x_{i_j}}{\gamma+1}\right)e^{\boldsymbol{\alpha}^T z_i}\right\} \tag{5}$$

and

$$S(t_{m_i}|\cdot) = \exp\left\{-\sum_{j=1}^{m_i}\psi^{j-1}x_{i_j}^{\gamma}\left(1+\phi t_{i_{j-1}} + \phi\gamma\frac{x_{i_j}}{\gamma+1}\right)e^{\boldsymbol{\alpha}^T z_i}\right\}, \tag{6}$$

where ϕ, γ, and ψ are positive parameters, $\boldsymbol{\alpha}$ is the parameter vector associated to observed covariates z_i, $t_{i_1}, \ldots, t_{i_{m_i}}$ are the occurrence times of the studied event for the ith individual, $x_{i_j} = t_{i_j} - t_{i_{j-1}}$ is the interval time between successive events with $t_{i_0} = 0$, and $\boldsymbol{\alpha}^T z_i = \alpha_1 z_1 + \cdots + \alpha_K z_K$ has no intercept term.

An advantage of this parameterization is its relatively easy interpretation. While the parameters γ and ϕ denote the specific effect of each time scale (total and interval times, respectively) in the survival, the parameter ψ denotes the effect of the number of events in the formulation. Moreover, the renewal component, $q_1(\cdot)$, is driven by a Weibull-type model, while the Poisson component, $q_2(\cdot)$, works as a time-dependent Poisson process part. The event count function, $q_3(\cdot)$, penalizes large numbers of events. An exponentially proportional covariate effect $g(\cdot)$ completes the formulation. Apart from α, the three parameters γ, ϕ, ψ represent departures from the Poisson intensity model.

The proposed model incorporates some existing models as particular cases in the survival literature. The MSS model with $\phi = 0$ and $\gamma = 1$ reduces to a counting process (Cox 1972). For $\phi = 0$ and $\psi = 1$, we have an ordinary Weibull renewal model for the interval times (see, e.g., Yannaros 1994). If $\gamma = 1$ and $\psi = 1$, we obtain a nonhomogeneous Poisson intensity process. For $\phi = 0$, the MSS reduces to an ordinary Weibull model with a counting parameter, then called an ordinary Weibull and counting model (see, e.g., McShane et al. 2008). Fixing $\psi = 1$, we obtain the renewal Poisson process (Prentice et al. 1981; Lawless and Thiagarajah 1996). If only $\gamma = 1$, we obtain a nonhomogeneous Poisson process with a counting parameter, which we called nonhomogeneous Poisson and counting processes (see, e.g., Massey et al. 1996).

2.2 MSS model with long time survivors

Following Berkson and Gage (1952), from (1) and (6) the MSS model with long time survivors is given by

$$S_{\text{pop}}(t|\boldsymbol{\lambda}^*, \rho) = \rho + (1 - \rho)S(t|\boldsymbol{\lambda}^*), \tag{7}$$

where $\boldsymbol{\lambda}^* = (\phi, \gamma, \psi, \boldsymbol{\alpha})$. It is important to point out that if $\rho = 0$, then $S_{\text{pop}}(t|\boldsymbol{\lambda}^*, \rho) = S(t|\boldsymbol{\lambda}^*)$, i.e., of course, (7) includes the usual MSS model. We can incorporate the covariates from both groups in which the population is divided, AR and OR. Following Farewell (1982), we described the proportion OR in terms of the covariates via a logistic function,

$$\rho_i = \frac{\exp(\boldsymbol{\beta}^T z_i)}{1 + \exp(\boldsymbol{\beta}^T z_i)}, \tag{8}$$

and the covariates of the proportion AR are considered in the survival function $S(t|\boldsymbol{\lambda}^*)$. Taking this into account, the cure probability is different for each individual, varying from 0 to 1. Then the improper survival function and the improper density function for the ith individual are given, respectively, by

$$S_{\text{pop}}(t_i|\boldsymbol{\lambda}) = \rho_i + (1 - \rho_i)S(t_i|\boldsymbol{\lambda}^*) \tag{9}$$

and

$$f_{\text{pop}}(t_i|\boldsymbol{\lambda}) = f(t_i|\boldsymbol{\lambda}^*)(1 - \rho_i), \tag{10}$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda}^*, \boldsymbol{\beta})$ is the parameter vector of interest, $S(t_i|\boldsymbol{\lambda}^*)$ is the time occurrence of the event survival function, $f(t_i|\boldsymbol{\lambda}^*)$ is the density function, and ρ_i denotes the

cure probability for each individual. In our modeling, $S(t_i|\boldsymbol{\lambda}^*)$ is given by (1), and $f(t_i|\boldsymbol{\lambda}^*) = -\frac{d}{dt}S(t|\boldsymbol{\lambda}^*)|_{t_i}$.

3 Inference

For inference, we adopt a full Bayesian approach. The likelihood function, prior distributions for the parameters in the model, details of the Markov chain Monte Carlo (MCMC) algorithm and the model comparison are described below.

3.1 Likelihood function

Considering that n individuals were observed, the data set consists of three vectors $\boldsymbol{t} = (\{t_1\}, \{t_2\}, \ldots, \{t_n\})$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$, and $\boldsymbol{z} = (\{z_1\}, \ldots, \{z_n\})$, where, for the ith individual, $\{t_i\}$ is the set of occurrence times of the event of interest for individual i, δ_i denotes the censoring indicator that equals zero if the individual is right-censored and one otherwise, and $\{z_i\}$ is the covariates set. Let $\boldsymbol{D} = (\boldsymbol{t}, \boldsymbol{z}, \boldsymbol{\delta})$ be the observed data set, the contribution of each individual to the likelihood, $L_i(\boldsymbol{\lambda}|\boldsymbol{D})$, is given by the density function if the individual presents the event of interest and by the survival function if the individual is censored. Then, we have

$$L_i(\boldsymbol{\lambda}|\boldsymbol{D}) = f_{\text{pop}}(t_{m_i}|\boldsymbol{\lambda}, \boldsymbol{D})^{\delta_i}\ S_{\text{pop}}(t_{m_i}|\boldsymbol{\lambda}, \boldsymbol{D})^{1-\delta_i}, \tag{11}$$

which enables us to conclude that the likelihood function considering all observed individuals is given by

$$L(\boldsymbol{\lambda}|\boldsymbol{D}) = \prod_{i=1}^{n} f_{\text{pop}}(t_{m_i}|\boldsymbol{\lambda}, \boldsymbol{D})^{\delta_i}\ S_{\text{pop}}(t_{m_i}|\boldsymbol{\lambda}, \boldsymbol{D})^{1-\delta_i}. \tag{12}$$

3.2 Sampling-based inference

The target distribution for inference is the posterior of the parameters of interest $\boldsymbol{\lambda} = (\phi, \gamma, \psi, \boldsymbol{\alpha}, \boldsymbol{\beta})$. For this, we need to obtain the marginal posterior densities of each parameter, which are obtained by integrating the joint posterior density with respect to each parameter. The posterior distribution is proper considering proper prior distribution (Ibrahim et al. 2001). However, irrespective of the prior distribution chosen, the joint posterior distribution for the proposed model is analytically intractable. As an alternative, we use Markov chain Monte Carlo methods (MCMC), e.g., the Gibbs Sampling and Metropolis–Hastings algorithm (see, e.g., Chib and Greenberg 1995).

Although it is not necessary, for simplicity, we assume that the parameters are independent a priori, and they have prior distribution according to the parametric space of each one, which means that

$$\begin{aligned}\pi(\phi, \gamma, \psi, \boldsymbol{\beta}) = {} & f_{\Gamma}(\phi|a_\phi, b_\phi) f_{\Gamma}(\gamma|a_\gamma, b_\gamma) f_{\Gamma}(\psi|a_\psi, b_\psi) \\ & \times \prod_{k=1}^{K} f_{\mathcal{N}}\big(\alpha_k|0, \sigma^2_{\alpha_k}\big) f_{\mathcal{N}}\big(\beta_k|0, \sigma^2_{\beta_k}\big),\end{aligned} \tag{13}$$

where $f_\Gamma(x|a,b) \propto x^{a-1}\mathrm{e}^{-bx}$, $x > 0$, is the density function of a Gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$, with mean a/b and variance a/b^2, and $f_\mathcal{N}(\cdot|0,\sigma^2)$ is the density function of a normal distribution with mean 0 and variance σ^2. Then the posterior distribution is proportional to

$$\pi(\boldsymbol{\lambda}|\boldsymbol{D}) \propto L(\boldsymbol{\lambda}|\boldsymbol{D}) \times f_\Gamma(\phi|a_\phi,b_\phi) f_\Gamma(\gamma|a_\gamma,b_\gamma) f_\Gamma(\psi|a_\psi,b_\psi) \times \prod_{k=1}^{K} f_\mathcal{N}\big(\alpha_k|0,\sigma^2_{\alpha_k}\big) f_\mathcal{N}\big(\beta_k|0,\sigma^2_{\beta_k}\big). \tag{14}$$

The algorithm needs the complete conditional densities of each parameter which are presented in Appendix A.

The conditional densities given in Appendix A do not refer to any known distribution. However, the Metropolis–Hastings algorithm can generate a sample of $\phi, \gamma, \psi, \alpha_k$, and β_k using complete conditional distributions of unknown parameters. The steps are described below. We start with $\boldsymbol{\lambda}^{(0)} = (\phi^{(0)}, \gamma^{(0)}, \psi^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)})$ and generating $\widetilde{\phi}$ from the prior $\pi(\phi) = f_\Gamma(\phi|a_\phi,b_\phi)$ described previously and u from uniform distribution $U(0,1)$. We then make the following comparison: if $u \leq \min\{1, \pi(\widetilde{\phi}|\gamma^{(0)},\psi^{(0)},\boldsymbol{\alpha}^{(0)},\boldsymbol{\beta}^{(0)},\boldsymbol{D})/\pi(\phi^{(0)}|\gamma^{(0)}),\ \psi^{(0)},\ \boldsymbol{\alpha}^{(0)},\boldsymbol{\beta}^{(0)},\boldsymbol{D})\}$, then update $\phi^{(1)}$ by $\widetilde{\phi}$. Otherwise stay with $\phi^{(0)}$, i.e., $\phi^{(1)} = \phi^{(0)}$. Next, we make a similar procedure to obtain $\gamma^{(1)}, \psi^{(1)}, \alpha_k^{(1)}$, and $\beta_k^{(1)}, k = 1,\ldots,K$, always updating the start value. We repeat the algorithm steps until a stationary sample can be obtained.

In order to verify the convergence diagnostic of simulated samples, Gelfand and Smith (1990) suggest graph techniques, Gelman and Rubin (1992), Geweke (1992), and Cowless and Carlin (1996) propose statistics analysis of a generated sample. The Gelman–Rubin criterion is implemented in the R systems (R Development Core Team 2006) and shall be used with Geweke's graphic analysis here.

3.3 Model comparison

Model selection is a very important issue. The MSS model (6) has various particular cases, and our interest lies in verifying if a simpler model could be considered. Therefore, we may test the hypotheses $H_0 : \psi = 1$, $H_0 : \phi = 0$, $H_0 : \gamma = 1$, $H_0 : \gamma = 1, \psi = 1$, $H_0 : \phi = 0, \psi = 1$, and $H_0 : \phi = 0, \gamma = 1$, which lead to various particular cases of (6).

In the literature, there are various methodologies which intend to analyze the suitability of a model and to select the best fit among a collection of models. For instance, among several existing techniques (see, e.g., Paulino et al. 2003, p. 348), we consider the Bayesian information criterion (BIC) which is defined by $-2\log(\widehat{\boldsymbol{\lambda}_r}) + q\log(n)$, where $\widehat{\boldsymbol{\lambda}_r}$ is the maximum likelihood estimate $\boldsymbol{\lambda}_r$ under model r, q is the number of parameters estimated under r model, and n is the sample size. The best model corresponds to the lower BIC value. Furthermore, there is the Deviance Information Criterion (DIC), proposed by Spiegelhalter et al. (2002). The DIC is an approximation of the Bayes factor, and its objective is to incorporate the complexity of the model into the selection criterion. Define $\mathrm{DIC}_r(\boldsymbol{\lambda}_r) = -2\ln(f(D|\boldsymbol{\lambda}_r)/h(D))$, where $h(D)$ is a function of the data, with parameter vector $\boldsymbol{\lambda}_r$, which does not interfere in the choice

of the model. As a suitable measure of the model is proposed, the posterior mean of $\mathrm{DIC}_r(\boldsymbol{\lambda}_r)$, and, associated to the model complexity, a penalized factor is proposed, p_{DIC_r}, given by $p_{\mathrm{DIC}_r} = \mathrm{E}_{(\boldsymbol{\lambda}_r|D)}[\mathrm{DIC}_r(\boldsymbol{\lambda}_r)] - \mathrm{DIC}_r[\mathrm{E}_{(\boldsymbol{\lambda}_r|D)}(\boldsymbol{\lambda}_r)]$. Finally, the DIC is given by

$$\mathrm{DIC}_r = \mathrm{E}_{(\boldsymbol{\lambda}_r|D)}\big[\mathrm{DIC}_r(\boldsymbol{\lambda}_r)\big] + p_{\mathrm{DIC}_r} = 2\mathrm{E}_{(\boldsymbol{\lambda}_r|D)}\big[\mathrm{DIC}_r(\boldsymbol{\lambda}_i)\big] - \mathrm{DIC}_r\big[\mathrm{E}_{(\boldsymbol{\lambda}_r|D)}(\boldsymbol{\lambda}_r)\big]. \tag{15}$$

The better model corresponds to a lower DIC.

In our case, the mean values needed for the DIC calculation are not obtained in an analytical form. However, we can obtain numerical values using computational methods such as MCMC, in which, to obtain the mean $\mathrm{E}_{(\boldsymbol{\lambda}_r|D)}[\mathrm{DIC}_r(\boldsymbol{\lambda}_r)]$, it is enough to obtain a sample $\boldsymbol{\lambda}_r^{(*)} = \{\boldsymbol{\lambda}_r^{(1)}, \ldots, \boldsymbol{\lambda}_r^{(L)}\}$ of the posterior, and then, with this, we have $\mathrm{E}_{(\boldsymbol{\lambda}_r|D)}[\mathrm{DIC}_r(\boldsymbol{\lambda}_r)] \approx (1/L)\sum_{l=1}^{L} \mathrm{DIC}_r(\boldsymbol{\lambda}_r^{(l)})$. Similarly, we have $\mathrm{E}_{(\boldsymbol{\lambda}_r|D)}[\lambda_r] \approx (1/L)\sum_{l=1}^{L} \boldsymbol{\lambda}_r^{(l)}$.

3.4 Coverage probability

Simulation studies were carried out with the objective of analyzing the frequentist properties of the estimation procedure based on resamples. To examine the frequentist properties, we constructed the credible intervals for all the parameters and calculated their coverage probabilities (CP). The parameter values were chosen based on a clinical experiment, in which the effectiveness of a treatment is analyzed using a control group and a treatment one. The vector parameter to be estimated is given by $\boldsymbol{\lambda} = (\phi, \gamma, \psi, \beta)$. We consider two different sets of parameter values: (i) $\phi = 0.6$, $\gamma = 1.2$, $\psi = 0.4$, and $\beta = -0.3$, and (ii) $\phi = 1.3$, $\gamma = 0.8$, $\psi = 0.7$, and $\beta = -1.5$, with 40% of censoring. In order to differentiate the control group from the treatment group, we used a binary covariate z, so that z is equal to -1 or 1, respectively. The Gamma distribution $\Gamma(0.9, 0.3)$, with mean 3 and variance 10, is considered as the prior distribution of ϕ, γ, and ψ. A normal distribution with mean 0 and variance 100 is considered for β. Overall, sixteen setups were considered, defined according to the two different sets of parameter values, four different sample sizes ($n = 30$, 50, 70, and 100) and two different numbers of events per individual ($m_i = m = 3$ and $m_i = m = 5$, $i = 1, \ldots, n$). For each setup, 1,000 artificial data sets were generated. We considered two chains of 55,000 iterations. The first 5,000 were ignored to avoid the influence of the first values. The remaining ones were selected using thinning by 50 to avoid a series correlation. The R systems (R Development Core Team 2006) was used in the whole study. The convergence of the chains were monitored using the method proposed by Gelman and Rubin (1992) and the graphic analysis proposed by Geweke (1992).

In order to obtain the CP of the credible intervals, for all samples, we calculated the parameters 95% credible intervals and observed if they contained the true parameter values. The empirical CP results for different sets of parameter values, different sample sizes, and different numbers of recurrence are summarized in Table 1. It can be concluded that the empirical coverage probabilities are closer to 95% level for a larger sample size, but small and moderate numbers of recurrence do not harm the empirical coverage probabilities. These results are similar for both sets of parameter values.

4 Mammary tumor data

In this section, the methodology is illustrated in a well-known real data set. We consider the data from Table 1 of Gail et al. (1980), which presents the times to develop mammary tumors for 48 rats in a carcinogenicity experiment. Initially, a conductive for a mammary cancer carcinogenic compound was injected into all rats. Afterward, the rats were induced to remain tumor free during the first 60 days. Then, twenty-three rats were assigned randomly to a treatment group, and the remaining 25 to a control group, and they were observed for 122 more days. The rats were observed constantly to check the appearance of tumors. The times of recurrence of tumors make up the data set. We considered that an individual is cured if it does not have tumor recurrence after 70 days of observation. Then we have almost 40% and 12% of censoring in the treatment group and control group, respectively. These percentages of censoring reveal the possible existence of a cure fraction.

The MSS model with a cure fraction fitted the data in the sampling-based approach. We assumed that the covariate which denotes the group indicator is directly linked to the cure fraction, $\beta^T z_i = \beta_0 + \beta_1 z_i$, where z_i is a centralized covariate, so that $z_i = -1$ if the ith individual is in the control group and $z_i = 1$, otherwise. We considered as prior distributions: $\phi \sim \Gamma(1, 0.01)$, $\gamma \sim \Gamma(1, 0.01)$, $\psi \sim \Gamma(1, 0.01)$ with $\mathrm{E}(\phi) = \mathrm{E}(\gamma) = \mathrm{E}(\psi) = 1,000$ and $\mathrm{Var}(\phi) = \mathrm{Var}(\gamma) = \mathrm{Var}(\psi) = 100,000$; $\beta_0 \sim \mathcal{N}(0, 100)$ and $\beta_1 \sim \mathcal{N}(0, 100)$ with $\mathrm{E}(\beta_0) = \mathrm{E}(\beta_1) = 0$ and $\mathrm{Var}(\beta_0) = \mathrm{Var}(\beta_1) = 100$. The hyperparameter values were chosen ensuring noninformativeness. Two chains of 100,000 iterations were considered. The first 20,000 were ignored to avoid the influence of the first values. The remaining ones were selected using thinning by 40 to avoid a series correlation. Implementation was made in OpenBugs (Spiegelhalter et al. 1999), and the codes can be obtained on request by emailing the authors. The chain convergence was monitored using Gelman–Rubin statistic (Gelman and Rubin 1992). Table 2 shows the posterior means and the corresponding 95% credible intervals (in parentheses) of the parameters. Table 3 shows the values of BIC and DIC criterion values. The results provide positive evidence for the complete model, showing the importance of taking into account the event counts, and the two time scales in the analysis. Figures 1 and 2 in Appendix B show the history of the

Table 1 Coverage probabilities of the credible intervals for different samples sizes and different numbers of recurrent events per individual, $m = 3$ (*left*) and $m = 5$ (*right*)

Sample size	30	50	70	100
$\phi = 0.6$	0.819/0.823	0.849/0.850	0.852/0.864	0.890/0.912
$\gamma = 1.2$	0.864/0.848	0.855/0.871	0.841/0.876	0.849/0.865
$\psi = 0.4$	0.913/0.917	0.909/0.922	0.932/0.960	0.948/0.962
$\beta = -0.3$	0.945/0.966	0.983/0.972	0.990/0.986	0.995/0.991
$\phi = 1.3$	0.823/0.845	0.875/0.892	0.856/0.878	0.870/0.894
$\gamma = 0.8$	0.852/0.868	0.874/0.889	0.861/0.885	0.859/0.877
$\psi = 0.7$	0.906/0.919	0.915/0.938	0.946/0.954	0.914/0.967
$\beta = -1.5$	0.917/0.941	0.992/0.953	0.997/0.994	0.994/0.999

Table 2 Posterior means and corresponding 95% credible intervals (*in parenthesis*)

Model	Parameter				
	γ	ϕ	ψ	β_0	β_1
MSS	1.552 (1.245,1.897)	0.566 (0.015,2.172)	1.387 (1.254,1.527)	−2.550 (−5.967,−0.423)	2.996 (0.359,6.094)
Ordinary Weibull and counting model	1.510 (1.198,1.869)	0	1.404 (1.266,1.547)	−2.725 (−6.257,−0.493)	2.892 (0.138,5.974)
Renewal Poisson process	0.988 (0.775,1.248)	1.545 (0.104,4,301)	1	−2.048 (−4.873,−0.071)	3.297 (1.339,6.177)
Nonhomogeneous Poisson and counting process	1	0.266 (0.007,0.971)	1.195 (1.124,1.263)	−2.089 (−5.159,−0.130)	3.303 (1.417,6.354)
Counting process	1	0	1.211 (1.142,1.274)	−2.196 (−5.139,−0.236)	3.319 (1.348,6.288)
Ordinary Weibull model	0.853 (0.695,1.038)	0	1	−2.323 (−5.322,−0.351)	3.177 (1.068,6.228)
Nonhomogeneous Poisson process	1	1.515 (0.207,3.299)	1	−2.092 (−5.075,−0.092)	3.305 (1.337,6.258)

Table 3 BIC and DIC criterion values

Model	BIC	DIC
MSS	24.71	9.14
Ordinary Weibull and counting model	26.92	9.26
Renewal Poisson process	54.68	35.98
Nonhomogeneous Poisson counting process	30.47	12.42
Counting process	24.74	15.79
Ordinary Weibull model	51.19	42.27
Nonhomogeneous Poisson process	49.98	36.01

chains and approximate posterior marginal densities, respectively. From Table 3, the statistical significance of the γ, ϕ, and ψ parameters imply that it is important to consider the total time, the interval time, and the event counts in the modeling. The group cure probability is associated with the vector parameter $\boldsymbol{\beta}$ by the logistic function (8). Then the $\boldsymbol{\beta}$ estimation leads to a treatment cure fraction equal to 0.627, while in the control group this probability is 0.017. That is, individuals in the treatment group have approximately 63% of chance of not presenting a recurrence of the event of interest, while individuals in the control group have only approximately 2%. Therefore, there is clear evidence of the treatment benefit. This result is corroborated by the statistical significance of the parameter β_1, which is considered here for addressing

the important issue of testing the equality of the cure proportions in the control and treatment groups.

5 Final comments

Long time survival data appears in various areas, particularly in scenarios where individuals are subject to the recurrence of an event of interest. The proposed MSS model with long time survivors allows for two time scales (interval time and total time), the event counts, and covariates while keeping flexibility to accommodate a cure fraction. The model provides various particular cases which can be tested straightforwardly. Parameter estimates are obtained using a sampling-based approach, which allows for information to be input beforehand with lower computational effort. The results of a simulation study showed the effectiveness of the parameter estimation approach even for small and moderate sized samples even in the presence of censoring.

Although the specific parametric forms for $q_1(\cdot)$, $q_2(\cdot)$, and $q_3(\cdot)$ in (6) are analytically convenient and are appealing because they can be interpreted, they are not critical for the overall approach. Alternative forms should be considered in order to study possible model misspecification. Specifying MSS models with a nonproportional regression structure may have a physical appeal. For instance, we could consider an accelerated failure time model (Cox and Oakes 1984). This would, however, add extra difficulties to the analysis and requires further work.

Acknowledgements The authors thank the reviewers for their comments and suggestions, which led to a substantial improvement of the manuscript. The research was partially supported by the Brazilian Organizations CNPQ and CAPES.

Appendix A: The conditional posteriors for the model parameters are given below

$$\pi(\phi|\gamma,\psi,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{D})$$
$$\propto \phi^{a_\phi-1}\mathrm{e}^{-b_\phi\phi}\left\{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\log(1+\phi)-\left(1+\phi t_{i_j}+\phi\frac{\gamma}{\gamma+1}x_{i_j}\right)x_{i_j}^{\gamma}\psi^{j-1}\mathrm{e}^{\alpha^T z_i}\right]\right\},$$

$$\pi(\gamma|\phi,\psi,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{D})\propto\gamma^{a_\gamma-1}\exp\left\{-b_\gamma\gamma+\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\log\gamma+(\gamma-1)\log x_{i_j}\right.\right.$$
$$\left.\left.-\left(1+\phi t_{i_j-1}+\phi\frac{\gamma}{\gamma+1}x_{i_j}\right)x_{i_j}^{\gamma}\psi^{j-1}\mathrm{e}^{\alpha^T z_i}\right]\right\},$$

$$\pi(\psi|\phi,\gamma,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{D})\propto\psi^{a_\psi-1}\exp\left\{-b_\psi\psi+\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[(j-1)\log\psi\right.\right.$$
$$\left.\left.-\left(1+\phi t_{i,j-1}+\phi\frac{\gamma}{\gamma+1}x_{i_j}\right)x_{i_j}^{\gamma}\psi^{j-1}\mathrm{e}^{\alpha^T z_i}\right]\right\},$$

$$\pi(\alpha_k|\phi,\gamma,\psi,\boldsymbol{\alpha}_{-k},\boldsymbol{\beta},\boldsymbol{D})$$
$$\propto \exp\left\{-\frac{1}{2\sigma^2_{\alpha_k}}(\alpha_k-\mu_{\alpha_k})^2+\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\boldsymbol{\beta}^T z_i-\left(1+\phi t_{i_j-1}+\phi\frac{\gamma}{\gamma+1}x_{i_j}\right)x_{i_j}^{\gamma}\psi^{j-1}\mathrm{e}^{\boldsymbol{\alpha}^T z_i}\right]\right\},\quad \text{and}$$

$$\pi(\beta_k|\phi,\gamma,\psi,\boldsymbol{\alpha},\boldsymbol{\beta}_{-k},\boldsymbol{D})$$
$$\propto \exp\left\{-\frac{1}{2\sigma^2_{\beta_k}}(\beta_k-\mu_{\beta_k})^2+\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\boldsymbol{\beta}^T z_i-\left(1+\phi t_{i_j-1}+\phi\frac{\gamma}{\gamma+1}x_{i_j}\right)x_{i_j}^{\gamma}\psi^{j-1}\mathrm{e}^{\boldsymbol{\alpha}^T z_i}\right]\right\},$$

where a and b, indexed by the parameters, are the shape parameters and the scale parameters of the Gamma density of prior distributions of ϕ,γ, and ψ; μ_{β_k} and σ_{β_k} are, respectively, the prior means and standard deviation of each β_k; and $\boldsymbol{\beta}_{-k}=(\beta_0,\ldots,\beta_{k-1},\beta_{k+1},\ldots,\beta_k)$, that is, the parameter vector $\boldsymbol{\beta}$ without the kth component.

Appendix B: Plots of the chain histories and empirical marginal posterior densities for the mammary tumor data set

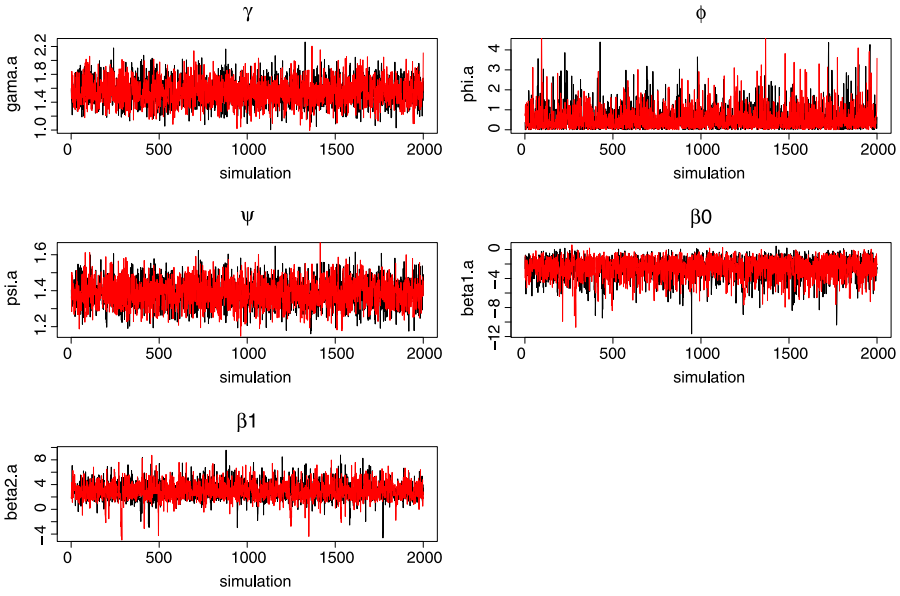

Fig. 1 History of the chains

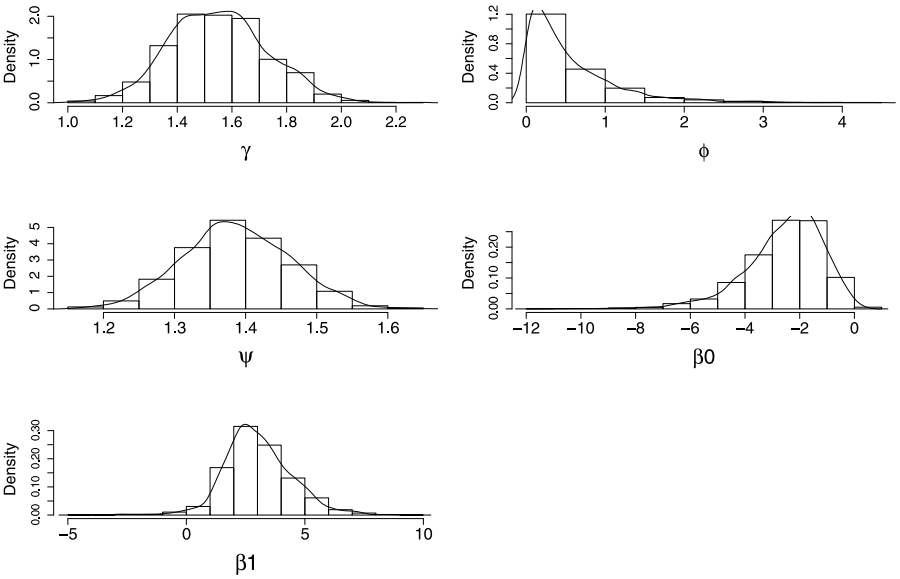

Fig. 2 Approximated posterior marginal densities

References

Berkson J, Gage RP (1952) Survival curve for cancer patients following treatment. J Am Stat Assoc 42:501–515

Boag J (1949) Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J R Stat Soc B 11:15–44

Chib S, Greenberg E (1995) Understanding the Metropolis–Hastings algorithm. Am Stat 49:327–335

Cobre J, Louzada-Neto F (2009) A sampling-based approach for a hybrid scale intensity model. J Stat, Adv Theory Appl 1:159–168

Cowless MK, Carlin BP (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. J Am Stat Assoc 91(434):883–904

Cox DR (1972) The statistical analysis of dependencies in point process. Stoch Point Process, Stat Anal Theory Appl, 55–66

Cox DR, Oakes D (1984) Analysis of Survival Data. Chapman and Hall, London

Farewell VT (1982) The use of mixture models for the analysis of survival data with long term survivors. Biometrics 38:1041–1046

Gail MH, Santner TJ, Brown CC (1980) An analysis of comparative carcinogenesis experiments based on multiple times to tumor. Biometrics 36:255–266

Gelfand AE, Smith AFM (1990) Sampling based approaches to calculating marginal densities. J Am Stat Assoc 85:398–409

Gelman A, Rubin BD (1992) Inference from iterative simulation using multiple sequences. Stat Sci 4:457–511

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Bayesian Stat 4:169–193

Ghitany M, Maller R (1992) Asymptotic results for exponential mixture models with long-term survivors. Statistics 23:321–336

Ghitany M, Maller R, Zhou S (1994) Exponential mixture models with long-term survivors and covariates. J Multivar Anal 49:218–241

Goldman AI (1984) Survivorship analysis when cure is a possibility: a Monte Carlo study. Stat Med 3:153–163

Ibrahim JG, Chen MH, Sinha D (2001) Bayesian survival analysis. Springer, New York

Kuk AYC, Chen C-H (1992) A mixture model combining logistic regression with proportional hazards regression. Biometrika 79:531–541

Lawless J, Thiagarajah K (1996) A point-process model incorporating renewals and time trends, with application to repairable systems. Technometrics 38:131–138

Lawless JF (2003) Statistical models and methods for lifetime data. Jonh Wiley and Sons, New York

Louzada-Neto F (2004) A hybrid scale intensity model for recurrent event dada. Commun Stat 33:119–133

Louzada-Neto F (2008) Intensity models for parametric analysis of recurrent events data. Braz J Probab Stat 22(1):23–33

Louzada-Neto F, Cobre J. (2010) A multiple time scale survival model. Adv Appl Stat 14:1–15

Maller R, Zhou X (1996) Survival analysis with long-term survivors. Wiley Series, Chichester

Massey WA, Parker GA, Whitt W (1996) Estimating the parameters of a nonhomogeneous Poisson process with linear rate. Telecommun Syst 5(2):361–388

McDonald JW, Rosina A (2001) Mixture modelling of recurrent events times with long-term survivors: Analysis of Hutterite birth intervals. Stat Methods Appl 10:257–272

McShane B, Adrian M, Bradlow E, Fader P (2008) Count models based on Weibull interarrival times. J Bus Econ Stat 26(3):369–378

Ng SK, McLachlan G, Yau K, Lee A (2004) Modelling the distribution of ischaemic stroke-specific survival time using an em-based mixture approach with random effects adjustment. Stat Med 23:2729–2744

Paulino CD, Turkman MAA, Murteira B (2003) Estatísitca Bayesiana Fundação Calouste, Gulbenkian

Peng Y, Dear KBG, Denham JW (1998) A generalized f mixture model for cure rate estimation. Stat Med 17:813–830

Prentice RL, Willians BJ, Peterson AV (1981) On the regression analysis of multivariate failure time data. Biometrika 68:373–379

R Development Core Team (2006) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Spiegelhalter DJ, Thomas A, Best NG, Gilks WR (1999) WinBugs: Bayesian Inference Using Gibbs Sampling. MRC Biostatistics Unit, Cambridge

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). J R Stat Soc B 64:583–639

Sy JP, Taylor JMG (2000) Estimation in a proportional hazards cure model. Biometrics 56:227–336

Yannaros N (1994) Weibull renewal process. Ann Inst Stat Math 46(4):641–648

Yu B (2008) A frailty mixture cure model with application to hospital readmission data. Biom J 50(3):386–394