CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open access journal indexed by

Volume 12 Number 1
April 2021
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Chilean Journal of Statistics

Volume 12, Number 1
April 2021

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Contents

Chilean Journal of Statistics
Vol. 12, No. 1, April 2021, 53–70

Bayesian Survival Analysis
Research Paper

Bayesian long-term survival model including a frailty term: Application to melanoma data

Agatha S. Rodrigues[1,*], Vinicius F. Calsavara[2,3], Eduardo Bertolli[4], Stela V. Peres[5], and Vera L.D. Tomazella[6]

[1]Department of Statistics, Federal University of Espírito Santo, Vitória-ES, Brazil
[2]Department of Epidemiology and Statistics, A.C.Camargo Cancer Center, São Paulo-SP, Brazil
[3]Biostatistics and Bioinformatics Research Center, Cedars-Sinai Medical Center, Los Angeles-CA, USA
[4]Skin Cancer Department, A.C.Camargo Cancer Center, São Paulo-SP, Brazil
[5]Department of Information and Epidemiology, Fundação Oncocentro de São Paulo, São Paulo-SP, Brazil
[6]Department of Statistics, Federal University of São Carlos, São Carlos-SP, Brazil

(Received: 04 August 2020 · Accepted in final form: 28 January 2021)

Abstract

In this paper, we propose a flexible cure rate model including a frailty term, which was obtained by incorporating a random effect in the risk function of latent competing causes. The number of competing causes of the event of interest follows a negative binomial distribution, and the frailty variable follows a power variance function distribution, which includes other frailty models such as gamma, positive stable, and inverse Gaussian frailty models as special cases. The proposed model takes into account the presence of covariates and right-censored data, which are suitable for populations with a long-term survivors. Besides, it allows quantification of the degree of unobserved heterogeneity induced by unobservable risk factors, which is important to explain the lifetime. Once the posterior density function is not expressed in the closed form, Markov chain Monte Carlo algorithms are performed for the estimation procedure. Simulation studies were considered in order to evaluate the proposed model performance, and its practical relevance was illustrated in a real medical dataset from a population-based study of incident cases of melanoma diagnosed in the state of São Paulo, Brazil.

Keywords: Competing causes · Frailty models · Markov chain Monte Carlo · Negative binomial distribution · Power variance function

Mathematics Subject Classification: Primary 62N01 · Secondary 62P10.

1. Introduction

Clinical outcomes in oncology are fundamental for all healthcare providers. Information such as overall survival, disease-free survival, and cancer-specific survival can be obtained based on the cancer type and patient features, such as the clinical stage, sex, age, education level, type of treatment, and other information that is often available in medical records. The incidence of a tumor is not always related to its severity. For instance, carcinomas of the

*Corresponding author. Email: agatha.rodrigues@ufes.br

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

skin are very common worldwide, but their clinical outcomes are among the best in oncology. Melanoma is not the most common skin malignancy; however, it is one of the most dangerous ones due to its potential of metastatic dissemination. According to the Brazilian National Institute of Cancer (INCA), approximately 6,000 new cases were expected in 2018 (INCA, 2018); whereas, according to the International Agency for Research on Cancer (IARC), approximately 7,000 new cases were reported (IARC, 2021). The number of deaths in Brazil due to melanoma is estimated to be approximately 2,000 cases per year (INCA, 2018).

The staging system proposed by the American Joint Committee on Cancer (AJCC) is commonly used worldwide for melanoma. According to the latest edition (Gershenwald et al., 2017), clinical stage IV corresponds to metastatic disease, which carries the worst prognosis. Even though several new modalities of treatment have been reported recently, treating these patients is still challenging (Ascierto et al., 2018). Clinical stage III corresponds to the nodal spreading of the melanoma; in this scenario, surgery is routinely associated with radiotherapy and/or some modality of systemic treatment such as immunotherapy or targeted therapy (Eggermont and Dummer, 2017). Clinical stages I and II correspond to the melanoma being limited to the skin, which is associated with a better prognosis. These patients are normally treated with surgery, and the great majority will be alive after 10 years of follow-up (Gershenwald et al., 2017).

In the traditional survival analysis approach, it is assumed that all units under study are susceptible to the event of interest. However, such an assumption is violated in several situations, such as in melanoma cancer studies, when the event of interest is death by disease. In the literature, it is known that clinical stages I and II have a better prognosis, meaning that a proportion of patients will not die from the disease; these patients are termed as having "immune" elements, "cured", or long-term survivors. Thus, a class of models, referred to as cure rate models consider this type of situation and have been studied by several authors. The Berkson-Gage model (Berkson and Gage, 1952) was probably the first model to propose the cured fraction, which is based on the assumption that only one cause is responsible for the occurrence of an event of interest (Cooner et al., 2007).

For melanoma, a patient death can be attributed to latent competing causes as the presence of cancer cells. These causes are based on the fact that each surviving carcinogenic cell can be characterized by an unknown time during which the cell could become a definitive tumor (Tsodikov et al., 2003). The books by Maller and Zhou (1996) and Ibrahim et al. (2001) as well as the articles by Tsodikov et al. (2003), Chen et al. (1999), Yin and Ibrahim (2005) and Rodrigues et al. (2009a) are key references.

Different distributions have been considered for the number of competing causes related to the occurrence of an event of interest. Chen et al. (1999) used Poisson distribution under a Bayesian approach, Rodrigues et al. (2009a) considered the negative binomial and geometric distributions, Rodrigues et al. (2009b) utilize the COM-Poisson distribution, Cancho et al. (2013) employed the power series distribution, Gallardo et al. (2017) considered the Yale-Simon distribution, Leão et al. (2018) assumed the Birnbaum-Saunders distribution, and Leão et al. (2020) used the zero-modified geometric distribution.

The promotion times are usually assumed to be independent and identically distributed, that is, the lifetimes of the carcinogenic cells follow a common distribution function, with the most common being exponential, piecewise exponential, and Weibull, among others (Calsavara et al., 2017). Besides, the long-term survival models implicitly assume a homogeneous population for the susceptible units. Although covariates can be included in the model in order to explain some observable heterogeneity, there is an unobserved heterogeneity induced by unobservable risk factors that are not commonly considered in the model (Wienke, 2011).

The models that take into account the unobservable heterogeneity are known as frailty models (Vaupel et al., 1979). These models are characterized by the inclusion of a random effect, that is, an unobservable random variable that represents the information that cannot

be observed, such as unobservable risk factors. If an important covariate is not included in the model, this will increase the unobservable heterogeneity, thus affecting the inferences about the parameters in the model. Therefore, the inclusion of a frailty term can help to relieve this problem (Hougaard, 1991).

The frailty term can be included in an additive form in the model. However, a multiplicative effect on the baseline hazard function is often used. Multiplicative frailty models represent a generalization of the proportional hazards model introduced by Cox (1972), in which the frailty term acts multiplicatively on the baseline hazard function. This approach has been studied by several authors, notably Hougaard (1995), Sinha and Dey (1997) and Balakrishnan and Peng (2006). Other authors, such as Calsavara et al. (2013), Calsavara et al. (2017), Scudilio et al. (2019) and Calsavara et al. (2020) considered the frailty models in the presence of long-term survivors.

We propose a class of survival models including a frailty term in the risk function of latent competing causes (Cancho et al., 2011), where the distribution of the frailty is the power variance function (PVF) family suggested by Tweedie (1984) and derived independently by Hougaard (1986). This approach allows that the competitive causes (cancer cells) have different frailties and that the frailest will fail earlier than those that are less frail. In addition, we consider that the number of competing causes related to the occurrence of an event of interest is modeled by the negative binomial distribution. This class of models allows some well-known models, depending on the parameter values, to be used. Herein, we illustrate the applicability of the proposed model in a real medical dataset from a population-based study of incident cases of melanoma diagnosed in the state of São Paulo, Brazil.

The rest of the paper is organized as follows. In Section 2, we present cure rate models under latent competing causes and the frailty model following a PVF distribution for the random effect, and the proposed model. Bayesian inference and simulation studies are described in Section 3. The proposed methodology is illustrated with real melanoma data also in this section. Finally, some final remarks are considered in Section 4.

2. Background and proposed model

In this section, we provide preliminary notions of long-term survival models under the biologic perspective, considering a negative binomial distribution for latent causes. Also, notions of the frailty model with their respective unconditional survival and density functions, as well as the proposed model, are provided here.

2.1 Cure rate models and frailty models

The time for the jth competing cause to produce the promotion time is denoted by Z_j, $j = 1, \ldots, N$, where N represents the number of cancer cells. The variable N is unobservable with the probability mass function (PMF) $p_n = P(N = n|\boldsymbol{\Theta})$ for $n = 0, 1, \ldots$. We assume that, conditional on N and on the parameter vector $\boldsymbol{\varphi}$, Z_js are independent and identically distributed with the cumulative distribution function $F(t|\boldsymbol{\varphi})$ and the survival function $S(t|\boldsymbol{\varphi}) = 1 - F(t|\boldsymbol{\varphi})$. Also, we assume that $Z_1, Z_2, \ldots$ are independent from N. The observable time of the occurrence of the event of interest is defined as $T = \min\{Z_0, Z_1, \ldots, Z_N\}$, where $P(Z_0 = \infty) = 1$, which leads to long-term survivors p_0 of the population not susceptible to the event occurrence. According to Rodrigues et al. (2009a), the survival function of the random variable T, conditional to parameter vector $\boldsymbol{\vartheta}$, is given by

$$S_{\text{pop}}(t|\boldsymbol{\vartheta}) = P(T \geq t|\boldsymbol{\vartheta}) = \sum_{n=0}^{\infty} P(N = n|\boldsymbol{\Theta})[S(t|\boldsymbol{\varphi})]^n = A_N[S(t|\boldsymbol{\varphi})],$$

where A_N is the probability generating function (PGF) of the random variable N, which converges when $s = S(t|\boldsymbol{\varphi}) \in [0, 1]$.

We suppose that the number of cancer cells (N), conditional to $\boldsymbol{\Theta} = (\eta, \theta)^\top$, follows a negative binomial distribution (Saha and Paul, 2005) with the PMF and PGF stated, respectively, as

$$p_n = P(N = n|\boldsymbol{\Theta}) = \frac{\Gamma(n + \eta^{-1})}{n!\Gamma(\eta^{-1})} \left(\frac{\eta\theta}{1 + \eta\theta}\right)^n (1 + \eta\theta)^{-1/\eta}$$

and

$$A_N(s) = \sum_{n=0}^{\infty} p_n s^n = [1 + \eta\theta(1 - s)]^{-1/\eta}, \quad 0 \le s \le 1,$$

for $n = 0, 1, \ldots$, $\theta > 0$, $\eta \ge 0$ and $1 + \eta\theta > 0$, so that $\mathrm{E(N|\boldsymbol{\Theta})} = \theta$ and $\mathrm{Var(N|\boldsymbol{\Theta})} = \theta + \eta\theta^2$. As discussed by Tournoud and Ecochard (2008), the parameters of the negative binomial distribution have biological interpretations in which the mean number of competing causes is represented by θ, whereas η is the dispersion parameter.

Under this setup, the population survival is given by

$$S_{\mathrm{pop}}(t|\boldsymbol{\vartheta}) = \{1 + \eta\theta[1 - S(t|\boldsymbol{\varphi})]\}^{-1/\eta}. \tag{1}$$

The long-term survivors is determined from Equation (1) as $p_0 = \lim_{t\to\infty} S_{\mathrm{pop}}(t|\boldsymbol{\vartheta}) = (1 + \eta\theta)^{-1/\eta} > 0$.

Amico and Van Keilegom (2018) reviewed the literature on long-term survival models and it is a recommended reference about the subject.

The frailty model considers a proportional hazard structure conditional on the random effect V. The random effect, called frailty, is a nonnegative variable that indicates the fragility of the unit. According to proportional hazard approach described by Cox and Oakes (1984), the conditional hazard is expressed as $h(t|V) = Vh_0(t)$, where h_0 is the baseline hazard function.

The survival function of T conditional to $V = v$ is given by

$$S(t|V, \boldsymbol{\varphi}) = S_0(t|\boldsymbol{\varphi})^V, \tag{2}$$

where S_0 denotes the baseline survival function.

In this paper, we suppose that the frailty variable V in Equation (2) follows the family of PVF distributions with parameters μ, ψ, and γ, suggested by Tweedie (1984) and derived independently by Hougaard (1986).

Let V be a random variable following a PVF distribution with parameters μ, ψ, and γ so that the density function can be written as (Wienke, 2011)

$$f_v(v; \mu, \psi, \gamma) = \exp\left[-\psi(1 - \gamma)\left(\frac{v}{\mu} - \frac{1}{\gamma}\right)\right] \frac{1}{\pi} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{[\psi(1 - \gamma)]^{k(1-\gamma)} \mu^{k\gamma} \Gamma(k\gamma + 1)}{\gamma^k k!} v^{-k\gamma-1} \times \sin(k\gamma\pi),$$

where $\mu > 0$, $\psi > 0$ and $0 < \gamma \le 1$.

Following the historical definition of frailty originally introduced in the field of demography (Vaupel et al., 1979) and to make sure that the model is identifiable (Wienke, 2011), we consider the restriction $\mathrm{E(V}|\mu, \psi, \gamma) = \mu = 1$. Consequently the $\mathrm{Var(V}|\mu, \psi, \gamma) = \mu^2/\psi = \sigma^2$, where σ^2 is interpreted as the measure of unobserved heterogeneity.

In order to eliminate the unobserved quantities, the random effect can be integrated out. Thus, the marginal survival function is given by

$$S(t|\boldsymbol{\varphi}^*) = \mathrm{E}_V[S(t|v_j, \boldsymbol{\varphi})] = \int_0^\infty \exp\left[-H_0(t|, \boldsymbol{\varphi})v_j\right] f_v(v_j|\gamma, \sigma^2)dv_j = L_v[H_0(t|\boldsymbol{\varphi})],$$

where $\boldsymbol{\varphi}^* = (\boldsymbol{\varphi}, \gamma, \sigma^2)^\top$ is the parameter vector, f_v is the density function of V conditional to γ and σ^2, H_0 is the cumulative baseline hazard function and L_v denotes the Laplace transform of the frailty distribution.

The unconditional survival and density functions in the PVF frailty model are expressed, respectively, by

$$S(t|\boldsymbol{\varphi}^*) = \exp\left\{\frac{1-\gamma}{\gamma\sigma^2}\left[1-\left(1+\frac{\sigma^2 H_0(t|\boldsymbol{\varphi})}{1-\gamma}\right)^\gamma\right]\right\} \tag{3}$$

and

$$f(t|\boldsymbol{\varphi}^*) = h_0(t|\boldsymbol{\varphi})\left(1+\frac{\sigma^2 H_0(t|\boldsymbol{\varphi})}{1-\gamma}\right)^{\gamma-1}\exp\left\{\frac{1-\gamma}{\gamma\sigma^2}\left[1-\left(1+\frac{\sigma^2 H_0(t|\boldsymbol{\varphi})}{1-\gamma}\right)^\gamma\right]\right\}. \tag{4}$$

Besides providing an algebraic treatment of the closed form for the marginal survival, the PVF family is a flexible model in the sense that it includes many other frailty models as special cases. For instance, the gamma frailty model is obtained if $\gamma = 0$; and, in the case of $\gamma = 0.5$, the inverse Gaussian distribution is derived. The positive stable is a special case of the PVF distribution; however, to show this fact, some asymptotic considerations are necessary.

2.2 The frailty long-term survival model

Thus, as an alternative to the usual cure rate models given in Equation (1), we propose a new model that incorporates a frailty term for each competing cause and consider that, conditional on $N = n$ and on $\boldsymbol{\varphi}^*$, the latent times follow a survival function as in Equation (3). As the number of competing causes follows a negative binomial distribution, the population survival function with the PVF frailty is given by

$$S_{\mathrm{pop}}(t|\boldsymbol{\vartheta}) = \left[1+\eta\theta\left(1-\exp\left\{\frac{1-\gamma}{\gamma\sigma^2}\left[1-\left(1+\frac{\sigma^2 H_0(t|\boldsymbol{\varphi})}{1-\gamma}\right)^\gamma\right]\right\}\right)\right]^{-1/\eta}, \tag{5}$$

where $\boldsymbol{\vartheta} = (\boldsymbol{\varphi}^*, \boldsymbol{\Theta})^\top$.

Usually, the most common choices for the promotion time distribution that specify the function $S(t|\boldsymbol{\varphi})$ have been exponential, piecewise exponential, or Weibull, among others. To capture the unobservable characteristics of each competing cause, we propose to incorporate a random effect (frailty term) on the baseline hazard function that acts multiplicatively in the promotion time. This approach allows that the competitive causes have different frailties and that the frailest will fail earlier than those that are less frail (Wienke, 2011).

We assume a Weibull distribution for the cumulative baseline hazard function, given by $H_0(t|\boldsymbol{\varphi}) = \exp(\alpha)t^\lambda$, where $\alpha \in \mathrm{R}$, $\lambda > 0$ and $\boldsymbol{\varphi} = (\alpha, \lambda)^\top$. Henceforward, we will refer to the model in which the survival function is as shown in Equation (5), by the PVF frailty cure rate model or simply the PVF cure rate model (PVFCR). Note that the usual cure rate model given in Equation (1) is obtained as $\sigma^2 \to 0$.

3. Bayesian inference and simulation study

In this section, we provide the Bayesian inference and simulation studies in order to evaluate the performance of the Bayesian estimators of the proposed model under different sample sizes and degree of heterogeneity in the sample. Also, we provide here the real data application.

3.1 Bayesian inference

Let us consider the situation when the time to the event is not completely observed and is subject to right censoring. For a given sample of size m, the observed time for the ith unit is $W_i = \min\{T_i, C_i\}$, with $T_i = \min\{Z_{i0}, Z_{i1}, \ldots, Z_{iN_i}\}$ and C_i is the censoring time, for $i = 1, \ldots, m$. Let δ_i be an indicator variable, in which $\delta_i = 1$ if $W_i = T_i$ and $\delta_i = 0$ otherwise. We include the covariate through the expected number of competing causes by $\mathrm{E}(\mathrm{N_i}|\boldsymbol{\Theta}) = \theta_\mathrm{i} = \exp(\mathbf{x}_\mathrm{i}^\top \boldsymbol{\beta})$, $i = 1, \ldots, m$, where $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients. The observed data are represented by $\mathbf{D} = (m, \boldsymbol{w}, \boldsymbol{\delta}, \mathbf{X})$, $\boldsymbol{w} = (w_1, \ldots, w_m)^\top$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)^\top$, and $\mathbf{X}$ is an $m \times k$ matrix containing the covariates.

The likelihood function of parameter $\boldsymbol{\vartheta} = (\boldsymbol{\varphi}^*, \boldsymbol{\Theta})^\top = (\alpha, \lambda, \gamma, \sigma^2, \eta, \boldsymbol{\beta})^\top$ under noninformative censoring can be written as

$$
\begin{aligned}
L(\boldsymbol{\vartheta}|\mathbf{D}) &\propto \prod_{i=1}^{m} [f_{\mathrm{pop}}(w_i|\boldsymbol{\vartheta})]^{\delta_i} [S_{\mathrm{pop}}(w_i|\boldsymbol{\vartheta})]^{1-\delta_i} \\
&\propto \prod_{i=1}^{m} \Big[\exp(\mathbf{x}_i^\top \boldsymbol{\beta}) f(w_i|\boldsymbol{\varphi}^*) \Big]^{\delta_i} \Big\{ 1 + \eta \exp(\mathbf{x}_i^\top \boldsymbol{\beta})[1 - S(w_i|\boldsymbol{\varphi}^*)] \Big\}^{-\frac{1}{\eta} - \delta_i},
\end{aligned}
$$

where $S(w_i|\boldsymbol{\varphi}^*)$ and $f(w_i|\boldsymbol{\varphi}^*)$ are given in Equations (3) and (4), respectively.

The posterior distribution of $\boldsymbol{\vartheta}$ comes out to be

$$
\begin{aligned}
\pi(\boldsymbol{\vartheta}|\mathbf{D}) \propto \pi(\boldsymbol{\vartheta}) \lambda^r \exp\left[\sum_{i=1}^{m} \delta_i x_i^\top \boldsymbol{\beta} + r \left(\alpha + \frac{1-\gamma}{\gamma \sigma^2} \right) \right] \prod_{i=1}^{m} \left[w_i^{\lambda-1} \left(1 + \frac{\sigma^2 \exp(\alpha) w_i^\lambda}{1-\gamma} \right)^{\gamma-1} \right]^{\delta_i} \\
\times \prod_{i=1}^{m} \left[1 + \eta \exp(x_i^\top \boldsymbol{\beta}) \left(1 - \exp\left\{ \frac{1-\gamma}{\gamma \sigma^2} \left[1 - \left(1 + \frac{\sigma^2 \exp(\alpha) w_i^\lambda}{1-\gamma} \right)^{\gamma} \right] \right\} \right) \right]^{-1/\eta - \delta_i} \\
\times \prod_{i=1}^{m} \exp\left[- \left(\frac{1-\gamma}{\gamma \sigma^2} \right) \left(1 + \frac{\sigma^2 \exp(\alpha) w_i^\lambda}{1-\gamma} \right)^{\gamma} \right]^{\delta_i}, \qquad (6)
\end{aligned}
$$

where $r = \sum_{i=1}^{m} \delta_i$ and $\pi(\boldsymbol{\vartheta})$ is the prior distribution of $\boldsymbol{\vartheta}$.

We consider independent prior distributions by defining them as $\boldsymbol{\beta} \sim \mathrm{Normal}_{k+1}(\mathbf{0}, 100\mathbf{I})$, with $\mathbf{I}$ being a $(k+1) \times (k+1)$ identity matrix, $\alpha \sim \mathrm{Normal}(0, 100)$, $\gamma \sim \mathrm{Uniform}(0, 1)$, and η, λ and σ^2 following a gamma distribution with mean value of 1 for all and variances of 1, 100 and 1, respectively. In this paper, no prior information about the parameters is available, which is the reason for the choice of non-informative prior distributions, besides of the assumption that the parameters are independent a prior. It is possible that the prior distributions can be postulated by expert knowledge and past experiences in situations they are available.

The posterior density of $\boldsymbol{\vartheta}$ in Equation (6) is analytically intractable because the integration of the joint density is not easy to perform. An alternative is to rely on Markov chain Monte Carlo (MCMC) simulations. Here, we consider the adaptive Metropolis-Hastings algorithm with a multivariate distribution as the proposed distribution (Haario et al., 2005)

implemented in the statistical package *LaplacesDemon* (Hall et al., 2020), which provides a friendly environment for Bayesian inference within the R program (R Core Team, 2020).

As a result, a sample of size n_p from the joint posterior distribution of $\boldsymbol{\vartheta}$ is obtained (eliminating burn-in and jump samples). The sample from the posterior can be expressed as $(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \ldots, \boldsymbol{\vartheta}_{n_p})$. The posterior mean of $\boldsymbol{\vartheta}$ can be approximated by

$$\widehat{\boldsymbol{\vartheta}} = \frac{1}{n_p} \sum_{k=1}^{n_p} \boldsymbol{\vartheta}_k, \tag{7}$$

and the posterior mean of the long-term survivors is approximated by

$$\widehat{p}_0 = \frac{1}{n_p} \sum_{k=1}^{n_p} (1 + \eta_k \theta_k)^{-1/\eta_k}. \tag{8}$$

Considering the function $Y_k(t) = S_{\text{pop}}(t|\boldsymbol{\vartheta}_k)$, where $S_{\text{pop}}(t|\boldsymbol{\vartheta}_k)$ is presented in Equation (5), conditional to $\boldsymbol{\vartheta}_k$, the posterior mean of the improper survival function is approximated by

$$\widehat{S_{\text{pop}}}(t|\boldsymbol{\vartheta}) = \frac{1}{n_p} \sum_{k=1}^{n_p} Y_k(t), \quad \text{for each } t > 0. \tag{9}$$

3.2 Simulation study

For data generation in this simulation study, we consider the model in given in Equation (5) with the Weibull distribution for the cumulative baseline hazard function with $\alpha = 0$, $\lambda = 1$ (exponential distribution with a rate of $\exp(\alpha)$), and one binary covariate X drawn from a Bernoulli distribution with the parameter 0.5. The PVF frailty distribution parameters are $\gamma = 0.5$ and $\sigma^2 \in \{0.5,\ 1,\ 1.5,\ 2\}$. The data of failure times were simulated with $\eta = 0.5$, $\theta_l = \exp(\beta_0 + l\beta_1)$, and $l = 0, 1$, where $\beta_0 = -0.5$ and $\beta_1 = 0.7$. The attribution of the parameters' values is motivated by the estimates obtained from real dataset application in Section 3.3 when fitted the model with only sex as a covariate.

In this way, $p_{0l} = (1 + \eta\theta_l)^{-1/\eta}$, so that the long-term survivors for the two levels of X are $p_{00} = 0.59$ and $p_{01} = 0.39$. The censoring times were sampled from the exponential distribution with the parameter τ (rate), where τ was set in order to control the proportion of censored observations. The algorithm to generate the observed times and censoring indicators is presented in the Algorithm *1*.

Algorithm *1* Data generation algorithm.

1: Draw $X_i \sim$ Bernoulli(0.5) and $u_i \sim$ Uniform$(0, 1)$.
2: Let $X_i = l$. If $u_i < p_{0l}$, $t_i = \infty$, otherwise,

$$t_i = \frac{(1-\gamma)}{\sigma^2 \exp(\alpha)} \left(\left\{ 1 - \frac{\gamma\sigma^2}{1-\gamma} \log \left[1 - \left(\frac{u^{-\eta} - 1}{\eta \exp(\beta_0 + \beta_1 x_i)} \right) \right] \right\}^{1/\gamma} - 1 \right).$$

3: Draw $c_i \sim$ Exponential (τ), which controls the proportion of censored observations.
4: Let $w_i = \min\{t_i, c_i\}$.
5: If $t_i < c_i$, set $\delta_i = 1$, otherwise, $\delta_i = 0$, for $i = 1, \ldots, m$.

We consider four sample sizes, $m = 100$, 300, 500 and 1000. For each combination of parameter values and sample sizes, we simulated $B = 1000$ random samples.

As mentioned previously, the Bayesian estimation procedures were performed using the adaptive Metropolis-Hastings algorithm such that the estimation of the covariance matrix is updated every 100 iterations. We generated 40,000 values for each parameter, disregarding the first 10,000 iterations to eliminate the effect of the initial values. In addition, jumps of size 30 were chosen to reduce the correlation effects between the samples. As a result, the final sample size of the parameters generated from the posterior distributions was $n_p = 1{,}000$. For good convergence results to be obtained, the convergence of the chains was monitored in all simulation scenarios, through monitoring graphics similar to what we did in the application (Section 3.3) and made available in the Appendix.

For each random sample, the estimates of $\boldsymbol{\vartheta}$ and the long-term survivors are obtained by Equation (7) and (8), respectively. We computed the average of B estimates of $\boldsymbol{\vartheta}$ (AE) and the root of the mean squared error (RMSE) of the estimators obtained from the PVFCR model. The results are summarized in Table 1.

According to the results, the average estimates of p_{00} and p_{01} were not affected by the increase of σ^2 value. Even for small sample sizes, the average estimates were close to the fixed values. The RMSE values appear reasonably close to zero as the sample size increases, except for the parameter σ^2, which needs a large sample size close to zero. For a fixed sample size, the RMSE of the σ^2 estimation increases as the σ^2 also increases.

To discuss the computational time, we simulated 100 datasets of each configuration and summarize these times (in seconds) in Table 2. The computational time increases as the sample size increases. For example, when $m = 100$ we take about 20 seconds, on average, to fit the proposed model, while we need about 80 seconds on average when $m = 1000$, regardless σ^2 value. This simulation study was conducted in a computer with the following configuration: Intel(R) Core(TM) i7-core 1.80GHz[Cores 4] processor (logical processors 8), 8 GB RAM, and Microsoft Windows 10 Home Single Language operating.

3.3 Application

The melanoma dataset used in this study is part of a retrospective cohort of patients diagnosed with melanoma in the state of São Paulo, Brazil, between 2000 and 2014, with follow-up conducted until 2018. The records were provided by the Fundação Oncocentro de São Paulo (FOSP), which is responsible for coordinating the Hospital Cancer Registry of the State of São Paulo, and it can be downloaded in http://www.fosp.saude.sp.gov.br. The FOSP is a public institution connected to the State Health Secretariat, which assists the study and implementation of public policies in the field of Oncology.

The time to death due to cancer was defined as the period between the dates of melanoma diagnosis and death. Those patients who did not die due to melanoma during the follow-up period were characterized as right-censored observations. The sample size was $m = 5358$ patients and the percentage of censored observations was 71%. The explanatory variables measured at baseline were as follows: sex (male or female), age (≤ 45 years or > 45 years), education level (no formal education, primary school, high school, or college), and cancer clinical stage (I, II, III or IV).

This datas were studied by Calsavara et al. (2020), where they evaluated only the effect of surgery in lifetime considering a non-proportional hazards model with a frailty term. Here, we also consider other relevant information available in the registry, such as gender, age at diagnosed, education level, and the clinical stage, as previously mentioned.

In the observations, 49.38% were male, and 79% were younger than 45 years old. For the education level, 58.3% had a primary school degree, 19.3% completed high school, 15% had a college degree and the remaining (7.4%) with no formal education. A total of 42.83% of the melanoma cases were classified as clinical stage I; II: 23.12%; III: 18.23%; and IV: 15.82%.

Table 1. The RMSE and the AE values for simulated data from the PVFCR model when $p_{00} = 0.59$, $p_{01} = 0.39$, $\beta_0 = -0.5$, $\beta_1 = 0.7$, $\alpha = 0$, $\lambda = 1$, $\eta = 0.5$, and $\gamma = 0.5$.

		Sample size (m)							
		100		300		500		1000	
σ^2	Parameter	RMSE	AE	RMSE	AE	RMSE	AE	RMSE	AE
0.5	p_{00}	0.071	0.587	0.039	0.588	0.032	0.587	0.022	0.588
	p_{01}	0.069	0.403	0.038	0.395	0.031	0.392	0.022	0.389
	β_0	0.545	-0.192	0.338	-0.299	0.294	-0.334	0.198	-0.393
	β_1	0.465	0.857	0.257	0.810	0.203	0.793	0.150	0.765
	η	1.055	1.394	0.825	1.116	0.727	1.008	0.531	0.835
	α	0.405	-0.210	0.289	-0.155	0.247	-0.133	0.188	-0.092
	λ	0.205	1.116	0.116	1.053	0.095	1.039	0.065	1.019
	γ	0.085	0.471	0.103	0.481	0.110	0.476	0.112	0.467
	σ^2	0.545	0.988	0.445	0.891	0.407	0.844	0.381	0.788
1	p_{00}	0.073	0.588	0.041	0.588	0.033	0.588	0.023	0.587
	p_{01}	0.072	0.404	0.038	0.396	0.031	0.394	0.021	0.390
	β_0	0.560	-0.182	0.392	-0.262	0.318	-0.307	0.232	-0.367
	β_1	0.481	0.871	0.288	0.837	0.218	0.810	0.159	0.775
	η	1.097	1.444	0.948	1.227	0.805	1.095	0.618	0.905
	α	0.538	-0.375	0.419	-0.310	0.360	-0.267	0.290	-0.192
	λ	0.168	1.064	0.108	1.015	0.086	1.001	0.069	0.996
	γ	0.094	0.455	0.121	0.444	0.124	0.436	0.127	0.422
	σ^2	0.293	1.056	0.278	0.990	0.273	0.956	0.325	0.960
1.5	p_{00}	0.068	0.585	0.042	0.585	0.033	0.586	0.023	0.586
	p_{01}	0.068	0.399	0.039	0.395	0.031	0.394	0.022	0.389
	β_0	0.584	-0.135	0.438	-0.220	0.333	-0.289	0.254	-0.344
	β_1	0.505	0.892	0.298	0.845	0.215	0.808	0.163	0.785
	η	1.167	1.516	1.048	1.309	0.836	1.130	0.673	0.961
	α	0.667	-0.534	0.536	-0.434	0.442	-0.359	0.369	-0.293
	λ	0.164	1.027	0.108	0.983	0.089	0.973	0.073	0.970
	γ	0.108	0.437	0.124	0.431	0.133	0.416	0.137	0.404
	σ^2	0.496	1.118	0.538	1.064	0.533	1.080	0.540	1.095
2	p_{00}	0.072	0.582	0.042	0.585	0.034	0.583	0.024	0.585
	p_{01}	0.066	0.395	0.038	0.393	0.030	0.392	0.021	0.390
	β_0	0.589	-0.128	0.463	-0.216	0.391	-0.245	0.269	-0.334
	β_1	0.483	0.894	0.289	0.850	0.239	0.823	0.162	0.784
	η	1.140	1.502	1.056	1.307	0.961	1.214	0.696	0.979
	α	0.751	-0.632	0.621	-0.530	0.572	-0.491	0.441	-0.370
	λ	0.154	0.988	0.110	0.957	0.099	0.950	0.083	0.950
	γ	0.105	0.438	0.130	0.419	0.133	0.415	0.132	0.407
	σ^2	0.920	1.147	0.948	1.112	0.946	1.117	0.890	1.211

Table 2. Minimum (Min.), first quartile (1qt), median, mean, third quartile (3qt), maximum (Max.) and standard deviation (SD) of the computational times (in seconds) to fit the proposed model for 100 simulated datasets when $p_{00} = 0.59$, $p_{01} = 0.39$, $\beta_0 = -0.5$, $\beta_1 = 0.7$, $\alpha = 0$, $\lambda = 1$, $\eta = 0.5$, and $\gamma = 0.5$.

σ^2	m	Min.	1qt	Median	Mean	3qt	Max.	SD
0.5	100	17.552	19.857	20.235	20.611	21.047	25.153	1.435
	300	32.851	33.305	33.680	34.034	34.615	37.338	1.000
	500	45.156	46.119	46.997	47.120	47.862	50.318	1.179
	1000	73.332	78.920	80.223	80.319	81.589	86.031	2.232
1	100	17.802	19.802	20.130	20.578	21.017	25.440	1.462
	300	32.669	33.108	33.414	33.756	34.104	36.512	0.908
	500	45.354	46.114	46.726	47.165	47.981	53.343	1.399
	1000	72.485	78.950	80.214	80.110	81.444	87.051	2.438
1.5	100	17.459	19.774	20.140	20.482	21.093	24.777	1.392
	300	32.601	33.218	33.464	33.786	34.084	36.288	0.894
	500	45.301	46.055	46.951	47.226	48.014	53.091	1.445
	1000	72.669	78.528	79.973	79.941	81.424	87.039	2.451
2	100	17.780	19.871	20.300	20.646	21.181	25.277	1.413
	300	32.703	33.261	33.598	34.031	34.622	37.533	1.062
	500	45.368	46.351	46.907	47.216	48.038	51.567	1.161
	1000	71.902	78.524	79.632	79.794	81.015	84.453	2.504

Figure 1 presents the Kaplan-Meier estimates for each explanatory variable. Of note, there was strong evidence that a fraction of the population had been long-term survivors. Among all of the variables considered in our study, those with clinical stage I melanoma had a better prognosis.

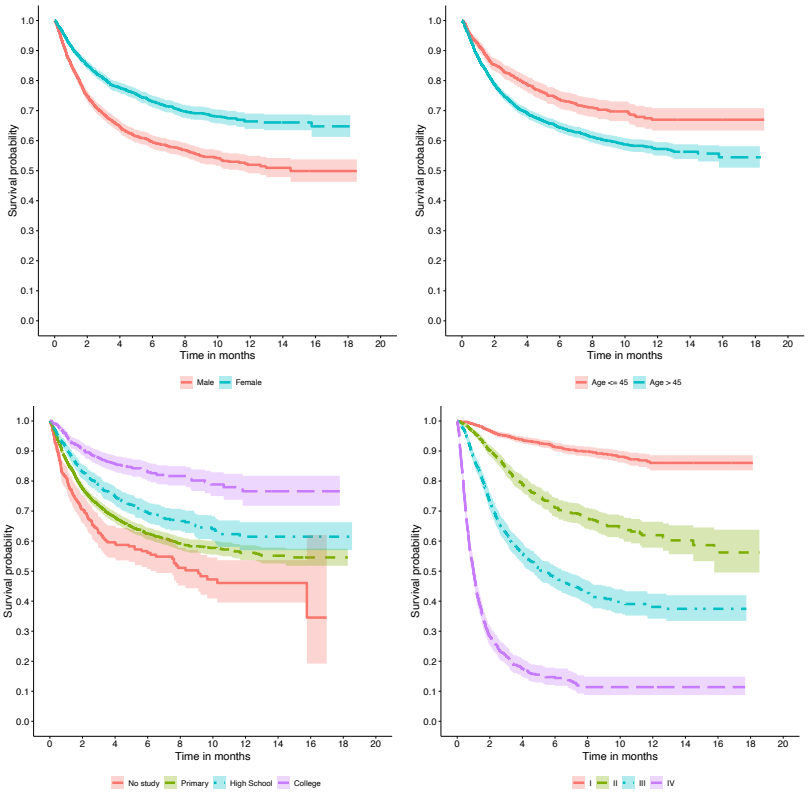

Figure 1. Kaplan-Meier estimates for the melanoma dataset grouped by sex, age, education level, and clinical stage, respectively.

To evaluate the effects of sex, age, education level, and clinical stage, the PVFCR model was fitted to the dataset. The adaptive Metropolis-Hastings algorithm was run, discarding the first 20,000 iterations as burn-in samples and using a jump of size 150 to avoid correlation problems, with a sample size of $n_p = 1{,}000$. The convergence of the chain was evaluated by multiple runs of the algorithm from different starting values and was monitored through graphical analysis. Good convergence results were obtained (see Appendix). The estimates of $\boldsymbol{\vartheta}$ and the long-term survivors were obtained by Equation (7) and (8), respectively, and the posterior mean of the improper survival function was given by Equation (9).

Table 3 lists the posterior mean, posterior standard deviation and 95% highest posterior density (95% HPD) intervals for all parameters from the PVFCR model. None of the parameters related to the explanatory variables have a 95% HPD value of zero.

The PVFCR model allows us to capture and to quantify the degree of unobservable heterogeneity, represented by σ^2, obtaining a posterior mean of 1.159 (95% HPD: 0.018; 2.687), which indicates a reasonable degree of unobserved heterogeneity in the sample. It is of great importance in clinical practice, once important covariates were not observed and not available in the dataset, such as Breslow thickness, ulceration and Mitotic rate.

Breslow thickness is the single most important prognostic factor for clinically localized primary melanoma. It is measured from the top of the granular layer of the epidermis (or, if the surface is ulcerated, from the base of the ulcer) to the deepest invasive cell across the broad base of the tumor (dermal/subcutaneous). Ulceration is an integral component of the AJCC staging system and an independent predictor of outcome in patients with clinically localized primary cutaneous melanoma. Multiple studies indicate that mitotic count is an important prognostic factor for localized primary melanoma since it represents tumor cells division (Bertolli et al., 2019; Fonseca et al., 2020).

Table 3. The posterior mean, standard deviation (SD) and 95% HPD of the fitted PVFCR model parameters.

			95% HPD	
Parameter	Mean	SD	Lower	Upper
λ	1.585	0.060	1.465	1.705
α	-3.045	0.220	-3.500	-2.659
η	1.477	0.175	1.126	1.818
β_0	-2.516	0.249	-2.984	-2.010
β_{sex} (male)	0.572	0.078	0.402	0.715
β_{age} (>45 years)	0.311	0.097	0.115	0.492
$\beta_{\text{education}}$ (no formal study)	1.094	0.174	0.782	1.415
$\beta_{\text{education}}$ (primary school)	0.595	0.129	0.339	0.832
$\beta_{\text{education}}$ (high school)	0.432	0.149	0.156	0.738
β_{stage} (II)	1.338	0.117	1.123	1.565
β_{stage} (III)	2.492	0.132	2.259	2.760
β_{stage} (IV)	4.697	0.183	4.354	5.060
γ	0.380	0.246	0.001	0.813
σ^2	1.159	0.763	0.018	2.687

College is the baseline for education level and stage I is the baseline for the melanoma clinical stage.

All of the findings of this study are consistent with those observed in routine clinical practice. Sex and age have already been reported as prognostic factors, suggesting that young patients and women have a better prognosis (Sabel et al., 2005; Balch et al., 2014). The education level is very likely to be related to knowledge about diseases and the necessity of medical evaluation for an early diagnosis. Clinical staging is also used for prognosis stratification, and the curves shown in this paper are very similar to those presented in the three latest updates of the AJCC staging system for melanoma (Balch et al., 2001, 2009; Gershenwald et al., 2017). The long-term survivors' estimates and survival estimates for a specific patient can be seen in Figures 2 and 3, respectively. As expected, the patients with clinical stage IV melanoma had a worse prognosis, regardless of their sex and age. On the other hand, the patients in clinical stage I melanoma as well as females and those younger than 45 years old had a better prognosis.

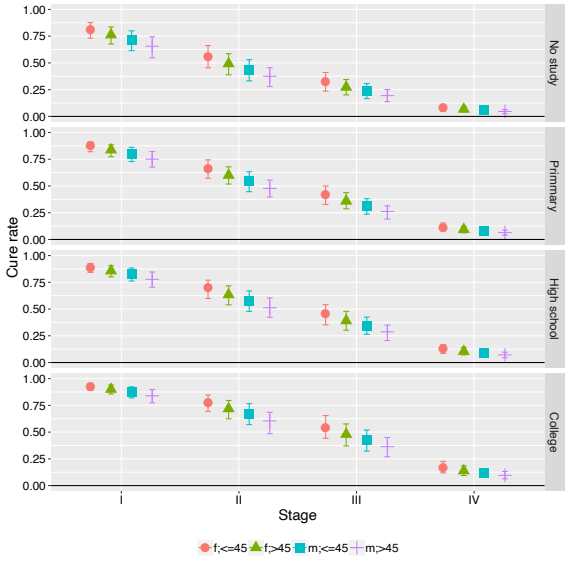

Figure 2. Long-term survivors' estimates (symbol) and 95% HPD intervals (bars) according to the fitted PVFCR model by considering sex (f, female and m, male), age ($\leq$ 45 years and $>$ 45 years), clinical stage (I, II, III and IV), and education level (no formal study, primary school, high school, and college).

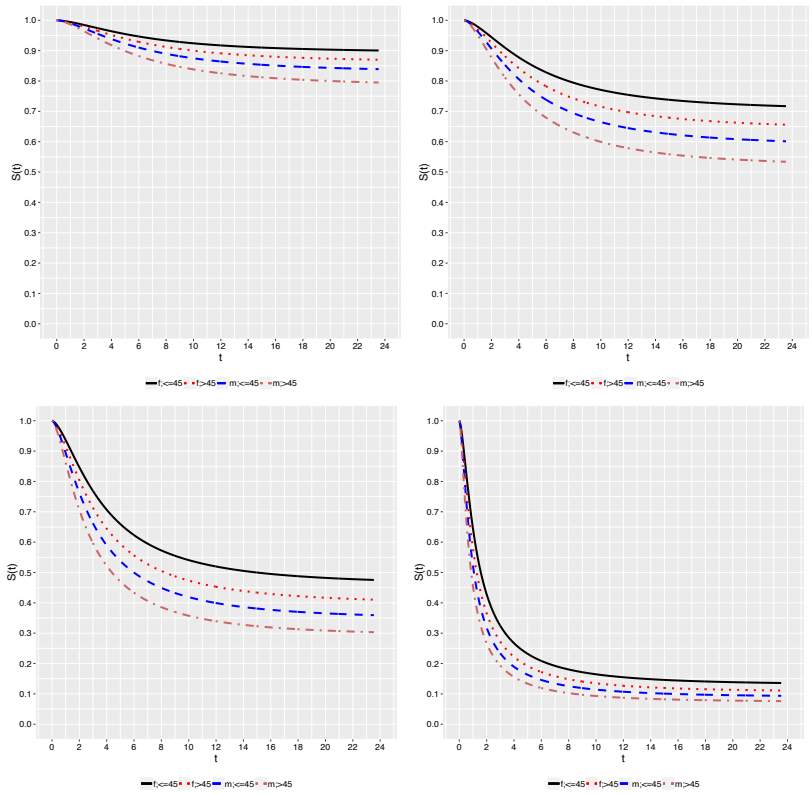

Figure 3. Survival functions estimated by the PVFCR model considering sex (f, female and m, male), age ($\leq$ 45 years and $>$ 45 years), and clinical stage I, II, III, and IV, respectively, for a fixed education level (high school category).

4. Final remarks

In this paper, we studied the cure rate model formulated by Cancho et al. (2011) in a different way, that is, we considered a random unobservable effect in promotion time of each competing cause, which allowed the unobserved heterogeneity to be quantified. The PVF frailty model was considered for the latent variables, and it included many other frailty models as special cases. A simulation study was conducted to illustrate the reliable performance of the Bayesian estimators of the proposed model, as the RMSE was reasonably close to zero as the sample size increased.

A point of attention is the fact that for large values of the parameter σ^2, one needs a large sample size for RMSE goes close to zero. However, it is worth to note that we obtained satisfactory values of RMSE and average of the estimates when $\sigma^2 = 1$ that is the close value of the estimated σ^2 in the application to the real dataset.

The applicability of the proposed model was demonstrated with a real melanoma dataset, explaining the model fit results and discussing its relevance in the real world. We hope that this model can be generalized to wider applications in survival analysis.

Acknowledgements

The authors thank the Fundação Oncocentro de São Paulo for providing the melanoma dataset. The authors would like to thank the reviewers and the editors for their valuable contributions.

Appendix

MCMC convergence monitoring for PVFCR model in Melanoma dataset

A jump of size 150 was considered to reduce correlation effects between the samples, as one can see in the autocorrelation graphs in figures 4, 5 and 6. Thus, final samples are considered with a lag of 150. After burn-in (20000) and jump samples, a sample of 1000 size from the posterior distribution of the parameters is generated. The time series graphs in Figures 7, 8 and 6 were built from the final posterior distribution sample, in which a type of blur is observed in a small variability of sampled values.

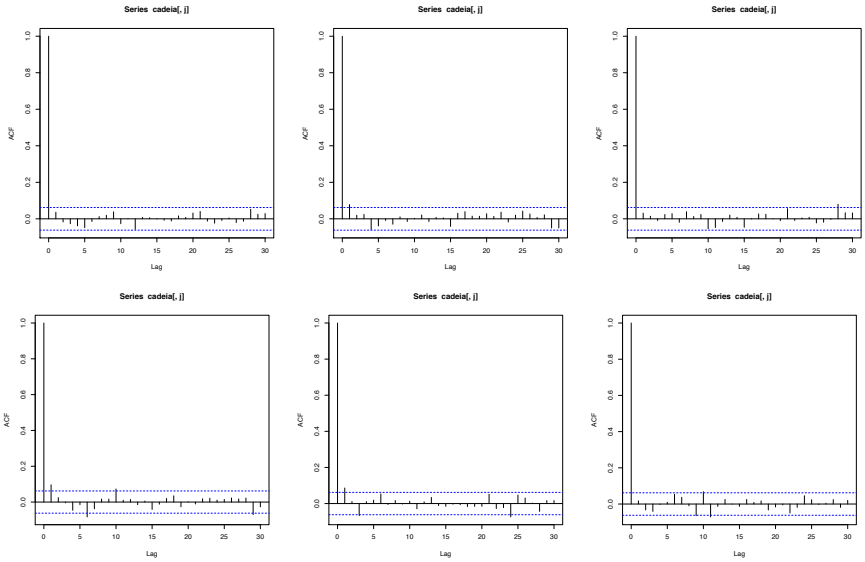

Figure 4. Autocorrelation graphs for λ, α, η, β_0, β_{sex} and β_{age} parameters.

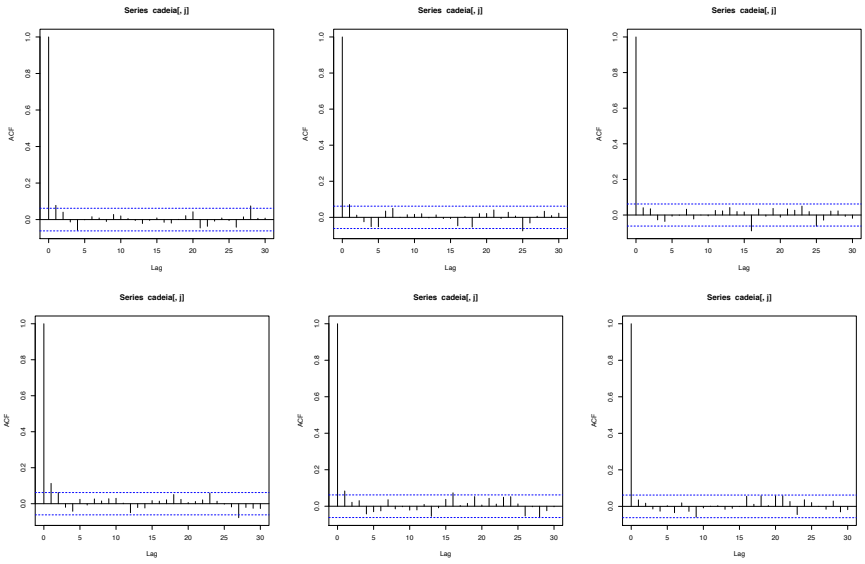

Figure 5. Autocorrelation graphs for β_{school} and β_{stage} parameters.

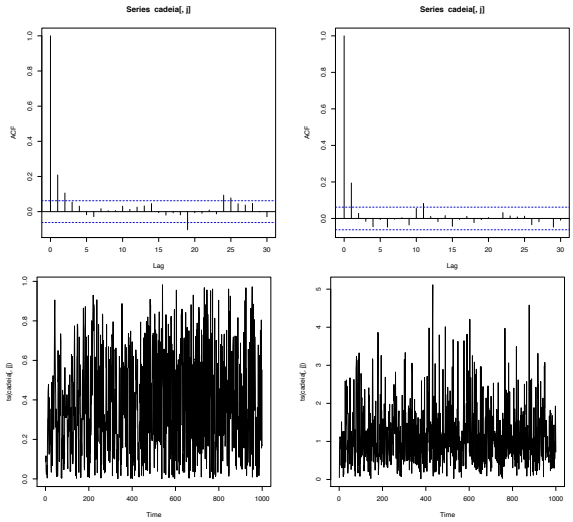

Figure 6. Autocorrelation (first panel) and time series (second panel) graphs for γ and σ^2 parameters.

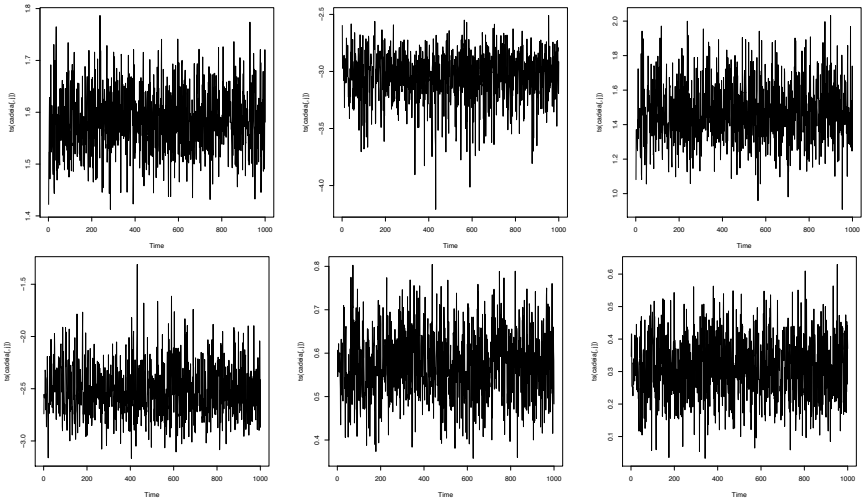

Figure 7. Time series graphs for λ, α, η, β_0, β_{sex} and β_{age} parameters.

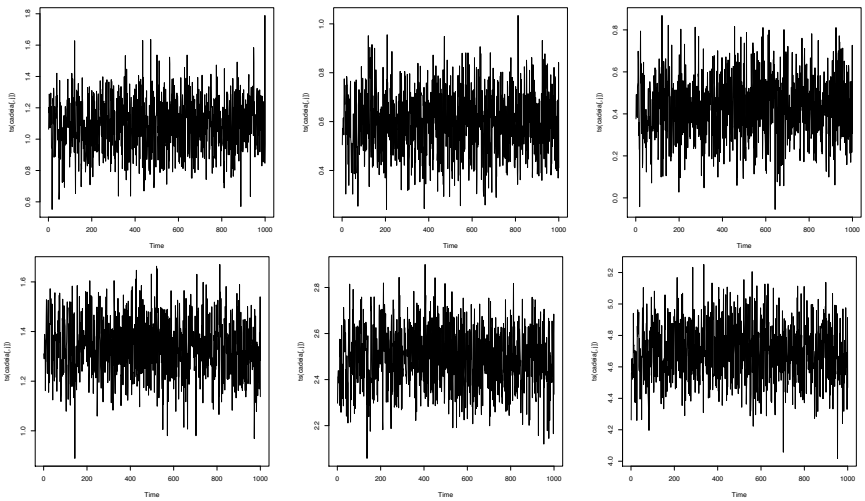

Figure 8. Time series graphs for β_{school} and β_{stage} parameters.

References

Amico, M. and Van Keilegom, I. 2018. Cure models in survival analysis. Annual Review of Statistics and Its Application, 5, 311–342.

Ascierto, P.A., Flaherty, K., and Goff, S., 2018. Emerging strategies in systemic therapy for the treatment of melanoma. American Society of Clinical Oncology Educational Book, 38, 751–758.

Balakrishnan, N. and Peng, Y., 2006. Generalized gamma frailty model. Statistics in Medicine, 25, 2797–2816.

Balch, C.M., Buzaid, A.C., Soong, S.J., et al., 2001. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. Journal of Clinical Oncology, 19, 3635–3648.

Balch, C.M., Gershenwald, J.E., Soong, S.J. et al., 2009. Final version of 2009 AJCC melanoma staging and classification. Journal of Clinical Oncology, 27, 6199.

Balch, C.M., Thompson, J.F., Gershenwald, et al., 2014. Age as a predictor of sentinel node metastasis among patients with localized melanoma: an inverse correlation of melanoma mortality and incidence of sentinel node metastasis among young and old patients. Annals of Surgical Oncology, 21, 1075–1081.

Berkson, J. and Gage, R.P., 1952. Survival curve for cancer patients following treatment. Journal of the American Statistical Association, 47, 501–515.

Bertolli, E., Franke, V., Calsavara, V.F., de Macedo, M.P., Pinto, C.A.L., van Houdt, W.J., Wouters, M.W., Neto, J.P.D., and van Akkooi, A.C., 2019. Validation of a nomogram for non-sentinel node positivity in melanoma patients, and its clinical implications: a Brazilian–Dutch study. Annals of Surgical Oncology, 26, 395–405.

Calsavara, V., Rodrigues, A., Tomazella, V., and de Castro, M., 2017. Frailty models power variance function with cure fraction and latent risk factors negative binomial. Communications in Statistics: Theory and Methods, 46, 9763–9776.

Calsavara, V.F., Milani, E.A., Bertolli, E., and Tomazella, V., 2020. Long-term frailty modeling using a non-proportional hazards model: Application with a melanoma dataset. Statistical Methods in Medical Research, 29, 2100–2118.

Calsavara, V.F., Tomazella, V.L.D., and Fogo, J.C., 2013. The effect of frailty term in the standard mixture model. Chilean Journal of Statistics, 4, 95–109.

Cancho, V.G., Louzada, F., and Ortega, E.M., 2013. The power series cure rate model: an application to a cutaneous melanoma data. Communications in Statistics: Simulation and Computation, 42, 586–602.

Cancho, V.G., Rodrigues, J., and de Castro, M., 2011. A flexible model for survival data with a cure rate: a Bayesian approach. Journal of Applied Statistics, 38, 57–70.

Chen, M.H., Ibrahim, J.G., and Sinha, D., 1999. A new Bayesian model for survival data with a surviving fraction. Journal of the American Statistical Association, 94, 909–919.

Cooner, F., Banerjee, S., Carlin, B.P., and Sinha, D., 2007. Flexible cure rate modeling under latent activation schemes. Journal of the American Statistical Association, 102, 560–572.

Cox, D.R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society B, 34, 187–220.

Cox, D.R. and Oakes, D., 1984. Analysis of Survival Data. Chapman and Hall, London.

Eggermont, A.M. and Dummer, R., 2017. The 2017 complete overhaul of adjuvant therapies for high-risk melanoma and its consequences for staging and management of melanoma patients. European Journal of Cancer, 86, 101–105.

Fonseca, I.B., Lindote, M.V.N., Monteiro, M.R., Doria Filho, E., Pinto, C. A.L., Jafelicci, A.S., de Melo Lôbo, M., Calsavara, V.F., Bertolli, E., and Neto, J.P.D., 2020. Sentinel node status is the most important prognostic information for clinical stage IIB and IIC melanoma patients. Annals of Surgical Oncology, 27, 4133–4140.

Gallardo, D.I., Gómez, H.W., and Bolfarine, H., 2017. A new cure rate model based on the Yule–Simon distribution with application to a melanoma data set. Journal of Applied Statistics, 44, 1153–1164.

Gershenwald, J.E., Scolyer, R.A., Hess, K.R., et al., 2017. Melanoma staging: evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA: A Cancer Journal for Clinicians, 67, 472–492.

Haario, H., Saksman, E., and Tamminen, J., 2005. Componentwise adaptation for high dimensional MCMC. Computational Statistics, 20, 265–274.

Hall, B., Hall, M., Statisticat, L.L.C., et al., 2020. Package 'LaplacesDemon'. R package version 16.1.4.

Hougaard, P., 1986. Survival models for heterogeneous populations derived from stable distributions. Biometrika, 73, 387–396.

Hougaard, P., 1991. Modelling heterogeneity in survival data. Journal of Applied Probability, 28, 695–701.

Hougaard, P., 1995. Frailty models for survival data. Lifetime Data Analysis, 1, 255–273.

IARC., 2021. Cancer Today. WHO. Available at https://gco.iarc.fr/today/home (accessed on 26-04-2021)

Ibrahim, J.G., Chen, M., and Sinha, D., 2001. Bayesian Survival Analysis. Springer, New York.

INCA, E., 2018. Incidência de câncer no Brasil [internet]. Instituto Nacional de Câncer José Alencar Gomes da Silva.

Leão, J., Bourguignon, M., Gallardo, D. I., Rocha, R., and Tomazella, V., 2020. A new cure rate model with flexible competing causes with applications to melanoma and transplantation data. Statistics in Medicine, 39, 3272–3284.

Leão, J., Leiva, V., Saulo, H., and Tomazella, V., 2018. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. Statistics in Medicine, 37, 4421-4440.

Maller, R.A. and Zhou, X., 1996. Survival Analysis with Long-term Survivors. Wiley New York.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodrigues, J., Cancho, V.G., de Castro, M., and Louzada-Neto, F., 2009a. On the unification of the long-term models. Statistics and Probability Letters, 79, 753–759.

Rodrigues, J., de Castro, M., Cancho, V.G., and Balakrishnan, N., 2009b. Com–Poisson cure rate survival models and an application to a cutaneous melanoma data. Journal of Statistical Planning and Inference, 139, 3605–3611.

Sabel, M.S., Griffith, K., Sondak, V.K., Lowe, L., Schwartz, J.L., Cimmino, V.M., Chang, A.E., Rees, R.S., Bradford, C.R., and Johnson, T.M., 2005. Predictors of nonsentinel lymph node positivity in patients with a positive sentinel node for melanoma. Journal of the American College of Surgeons, 201, 37–47.

Saha, K. and Paul, S., 2005. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. Biometrics, 61, 179–185.

Scudilio, J., Calsavara, V., Rocha, R., Louzada, F., Tomazella, V., and Rodrigues, A., 2019. Defective models induced by gamma frailty term for survival data with cured fraction. Journal of Applied Statistics, 46, 487–507.

Sinha, D. and Dey, D., 1997. Semiparametric Bayesian analysis of survival data. Journal of the American Statistical Association, 92, 1195–1212.

Tournoud, M. and Ecochard, R., 2008. Promotion time models with time-changing exposure and heterogeneity: Application to infectious diseases. Biometrical Journal, 50, 395–407.

Tsodikov, A., Ibrahim, J., and Yakovlev, A., 2003. Estimating cure rates from survival data: an alternative to two-component mixture models. Journal of the American Statistical Association, 98, 1063–1078.

Tweedie, M.C.K., 1984. An index which distinguishes between some important exponential families. In Statistics: Applications and New Directions: Proc. Indian Statistical Institute Golden Jubilee International Conference, pp. 579–604.

Vaupel, J.W., Manton, K.G., and Stallard, E., 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography, 16, 439–454.

Wienke, A., 2011. Frailty Models in Survival Analysis. Chapman an Hall, Boca Raton, FL, US.

Yin, G. and Ibrahim, J., 2005. Cure rate models: A unified approach. Canadian Journal of Statistics, 33, 559–570.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://soche.cl/chjs/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright