Bayesian survival model induced by frailty for lifetime with long-term survivors

Vicente G. Cancho
Departamento de Matemática Aplicada e Estatística, ICMC,
Universidade de São Paulo, São Carlos, SP, Brazil.
Gladys D.C. Barriga
Departamento de Engenharia de Produção, FEB,
Universidade Estadual Paulista, Bauru, SP, Brazil.
Gauss M. Cordeiro
Departamento de Estatística, CCEN,
Universidade Federal de Pernambuco, Recife, PE, Brazil.
Edwin M. M. Ortega*
Departamento de Ciências Exatas, ESALQ,
Universidade de São Paulo, Piracicaba, SP, Brazil.
Adriano K. Suzuki
Departamento de Matemática Aplicada e Estatística, ICMC,
Universidade de São Paulo, São Carlos, SP, Brazil.

Abstract

It is introduced the proportional hazards frailty model to allow a discrete distribution for the frailty variable. Frailty zero can be interpreted as being immune or cured. It is defined a class of survival models induced by a discrete frailty having a mixed Poisson distribution, which can account for unobserved dispersion. Further, a new regression to evaluate the effects of covariates in the cure fraction is constructed. Several former cure survival models are special cases of the proposed modeling framework. The inferential approach is based on Bayesian methods. Some

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/stan.12236

simulation results are provided to assess the performance of the new regression. Its importance is illustrated by means of an application to colorectal cancer data.

Keywords: Bayesian procedure; colorectal cancer; cure rate models; frailty model; survival models.

*__Corresponding author__: Departamento de Ciências Exatas, USP Av. Pádua Dias 11 - Caixa Postal 9, 13418-900 Piracicaba - São Paulo - Brasil. e-mail: edwin@usp.br

1 Introduction

The analysis of survival data is usually based on the assumption that the population under study is homogeneous. In this context, conditional on the covariates, every individual has the same risk for experiencing an event such as death or disease recurrence. The event times of the individuals in the population, conditional on the observed covariates, are assumed independent. However, this hypothesis can not be used for all of them as many applications require heterogeneous observations, i.e., individuals with different risks and hazards. Though individuals may look identical in some aspects, they may differ in unmeasured ways. In applications of survival analysis, a few covariates such as age, sex, severity of disease or laboratory data, are usually known. But there are many other factors that can influence survival including health status, life style, smoking, occupation and genetic risk factors. In retrospective studies, these factors are unknown and can not be included in the analysis. If these factors are ignored, it can be considered that a population analysis describes an average person, not an individual. Vaupel *et al.* (1979) and Lancaster (1979) suggested random effects models to account for the heterogeneity due to unobserved covariates. Vaupel *et al.* (1979) pioneered the term *frailty* to deal with unobserved heterogeneity, random effects, and association in univariate survival models. They introduced the concept of frailty to biostatistics by applying it on population mortality data. Lancaster (1979) discussed a model in economics under the terminology of *mixed proportional hazards model*. Frailty models account for unobserved heterogeneity that occur because some observations are more prone to failure, and therefore more frail than others in the data under investigation.

Studies on frailty models generally consider a non-negative and continuous frailty random variable. The gamma (Vaupel *et al.*, 1979; Hougaard, 1984), Gaussian (Hougaard, 1984), log-normal and positive stable (Hougaard, 1986a) distributions are commonly used frailty models. As pointed out by

Wienke (2011) a generalized class of frailty models that includes gamma, inverse Gaussian and positive stable distributions is the *power variance function* (PVF) family of frailty distributions defined by Tweedie (1984) and derived independently by Hougaard (1986b). This is a three-parameter family with parameters $\mu > 0$, $\lambda > 0$ and $0 < \alpha < 1$, and density function

$$g(w) = \mathrm{e}^{\lambda(1-\alpha)(\frac{\mathrm{w}}{\mu}-\frac{1}{\alpha})}\frac{1}{\pi}\sum_{\mathrm{k}=1}^{\infty}(-1)^{\mathrm{k}+1}\frac{[\lambda(1-\alpha)]^{\mathrm{k}(1-\alpha)}\mu^{\mathrm{k}(1-\alpha)}\lambda(\mathrm{k}\alpha+1)}{\alpha^{\mathrm{k}}\mathrm{k}!}\mathrm{w}^{-\mathrm{k}\alpha-1}\sin(\mathrm{k}\alpha\pi),\ \mathrm{w}>0. \quad (1)$$

The Laplace transform (Aalen, 1992) of the PVF distribution is

$$\mathcal{L}_W(s) = \mathrm{e}^{\frac{\lambda(1-\alpha)}{\alpha}\left[1-\left(1+\frac{\mu s}{\lambda(1-\alpha)}\right)^{\alpha}\right]}.$$

The mean and variance of the PVF distribution follow from the Laplace transform as $E(W) = \mu$ and $Var(W) = \mu^2/\lambda$, respectively.

Wienke (2011) showed that the stable distribution is a special case of the PVF distribution under some assumptions. Also, if $\alpha \longrightarrow 0$ and $\alpha = 0.5$, this distribution reduces to the gamma and inverse Gaussian models, respectively. The compound Poisson (cP) model (Aalen, 1992) is a frailty distribution, which can be constructed as a sum of a Poisson-distributed number of independent and identically gamma random variables. Wienke (2011) parameterized the cP model and showed that its Laplace transform is the same as that one of the PVF except for the range of α, which can be negative in the cP model. Consequently, the density function coincides with the corresponding function in the PVF model.

These distributions are convenient to derive unconditional survival functions in closed-form, density functions and hazard rates using the Laplace transform in computational terms. However, a continuous frailty distribution does not allow the existence of zero risk. Frailty zero indicates that there is a subgroup of non-susceptible individuals, where the event of interest has not happened even after a long period of observation.

The event of interest in many survival studies or cancer-relapse trials can be the death of a patient or a tumor recurrence. However, due to recent advances in cancer treatment therapies, a high proportion of individuals are expected to be *cured*, i.e., remaining disease-free after prolonged follow-ups. For this reason, there is a vast literature on 'cure rate models' for survival data, also called 'survival models with a surviving fraction' or 'long-term survival models'. Most of these models were investigated in a competing risks scenario (Tsodikov *et al.*, 2003; Rodrigues *et al.*, 2009; Cancho *et al.*, 2011b), but

they can also be obtained from the proportional hazard models with discrete frailty distributions (De Angelis *et al.*, 1999; Dunson & Zhou, 2000; Caroni *et al.*, 2010; Wienke, 2011; Mazroui *et al.*, 2013; Economou & Stehlik, 2015; Leão *et al.*, 2017; de Souza *et al.*, 2017; Barriga *et al.*, 2018).

Wienke (2011) introduced the frailty model in a context of univariate survival data analysis to model the unobserved heterogeneity of individuals. On the other hand, Aalen *et al.* (2008) used the theory of counting processes to model the times between events of interest under the presence of heterogeneity not observed between individuals. Based on these two works, we also model the presence of unobserved dispersion. Thus, the frailty model considers a proportional hazard structure conditional on the random effect Z. This random effect is a non-negative frailty variable that indicates the individual level of risk. Let T be the time until the occurrence of an event of interest following the frailty model. In the proportional hazard scenario (Cox & Oakes, 1984), the conditional hazard rate function (hrf) of T given the frailty Z can be expressed as $h(t|Z) = Z\,h_0(t)$, where $h_0(t)$ is a baseline hrf. This latter expression is called the multiplicative frailty model. We obtain the baseline density $f_0(t)$ and baseline cumulative distribution function (cdf) $F_0(t)$ of T corresponding to $h_0(t)$ when Z is equal to a constant one, i.e., all individuals at risk have the same constant frailty.

The survival function of T conditional on $Z = z$ has the form

$$S(t|Z=z) = P(T > t|Z=z) = \exp\left[-z\,H_0(t)\right], \tag{2}$$

where $H_0(t) = \int_0^t h_0(u)du$ is the baseline cumulative hazard function (chf) of T.

Many studies on these models often consider that Z is a non-negative continuous random variable. As it was mentioned previously, in specific situations, it may be appropriate to consider discrete frailty distributions, for example, in survival data containing experimental units in which the event of interest has not happened even after a long period of observation. In this situation, these units have zero frailty and survival models induced by frailty with continuous distribution would not be appropriate (Hougaard, 1984).

Thus, assuming that the frailty variable Z in Equation (2) is a discrete random variable with support $\{0, 1, 2, \dots\}$ and probability mass function (pmf) $P(Z = z) = p_z$, the unconditional survival function $S(t)$ for the entire population is

$$S(t) = \sum_{z=0}^{\infty} S(t|z)\,p_z = E_Z\{\exp[-ZH_B(t)]\} = \Psi_Z[S_0(t)], \tag{3}$$

where $S_0(t) = \exp[-H_0(t)]$ is the baseline survival function and $\Psi_Z(\cdot)$ is the probability generating function (pgf) of Z. Note that $S(t)$ in Equation (3) is the survival function proposed by Tsodikov *et al.* (2003) and Rodrigues *et al.* (2009). Moreover, if $P(Z = 0) > 0$, the survival function (3) is an improper function, i.e., $\lim_{t\to\infty} S(t) = \Psi_Z(0) = P(Z = 0) > 0$, and then it can be adopted to describe survival models with a cure fraction. If $P(Z = 0) = 0$, the survival function is proper, i.e., $\lim_{t\to\infty} S(t) = \Psi_Z(0) = P(Z = 0) = 0$ and $\lim_{t\to 0} S(t) = \Psi_Z(1) = 1$. It is simple to derive models from (2) for modeling survival data with a cure rate. For example, if Z follows the Poisson distribution with mean $\theta > 0$ and pgf $\Psi_Z(w) = \exp[-\theta(1 - w)]$, the survival function (3) becomes

$$S(t) = \exp[-\theta F_0(t)]. \tag{4}$$

The proportion of individuals with zero risk follows from (4) as $\lim_{t\to\infty} S(t) = \mathrm{e}^{-\theta} > 0$. For the data set on a colon cancer clinical trial introduced by Moertel *et al.* (1990) and discussed in Section 6, we note that the Kaplan–Meier estimate of the survival function of the recurrence time in Figure 2 (left panel) stabilizes at a level above zero and indicates disease-free patients. Equation (4) refers to the *Bounded Cumulative Hazard* (BCH) model defined by Yakovlev & Tsodikov (1996). The BCH model considers that non-immune patients to the event of interest (recurrence) are exposed to Z unknown competitive risk factors or causes of the event of interest with $E(Z) = Var(Z)$. The overdispersion of these risk factors ($Var[Z] > E[Z]$) is a phenomenon which increases the risk of death of the patients after a long treatment motivating the formulation of flexible distributions. The negative-binomial distribution is one of the most popular to explain overdispersion. In this context, de Souza *et al.* (2017) studied survival models induced by a discrete frailty variable with hyper-Poisson distribution. Cancho *et al.* (2018) used an inflated zero power series distribution for the frailty term. Based on these works, we propose a class of survival models induced by a frailty discrete random variable following a mixed Poisson distribution. The mixture distribution is the PVF family (Wienke, 2011) which is a multiplier on the expected value of the frailty variable. The PVF family includes as special cases the gamma, inverse Gaussian and stable positive distributions. For this formulation, the heterogeneity of the random variables Z's is equivalent to the model introduced by Barriga *et al.* (2018), which defines a random component in the expected value of the latent causes in the BCH model by Yakovlev & Tsodikov (1996). It is different from the models proposed by de Souza *et al.* (2017) and Cancho *et al.* (2018) since it does not include a random component in the expected value of the number of latent

risks.

This class of models includes the BCH model and the model defined by Barriga *et al.* (2018), among other models. Further, a new regression model is constructed to evaluate the covariate effects on the disease-free proportion. Note that Barriga *et al.* (2018) used a classic analysis and the inferential part was carried out under the asymptotic distribution of the maximum likelihood estimators (MLEs), which in situations when the sample is small, may present difficult results to be justified. In this research, we develop the Bayesian inference procedure based on the Monte Carlo Markov Chain (MCMC), specifically the collapsed Gibbs of Liu (1994) for efficient sampling from posterior distribution of the model parameters under noninformative improper priors. For the Bayesian inference, it is not necessary to adopt likelihood inference to fit the model to a data set. First, a simulation study is provided to evaluate the performance of the Bayesian inference based on the MCMC for estimating parameters. Then, the model is applied to a colonrectal cancer data set.

The paper is organized as follows. In Section 2, we formulate a model called the Bounded Cumulative Hazard-Power Variance Function (BCH-PVF) cure rate model, a generalization of the model introduced by Barriga *et al.* (2018). In Section 3, we obtain some of its mathematical properties directly from those of the exponentiated distribution defined from the baseline F_0. In Section 4, we carry out a Bayesian inference for the BCH-PVF regression. A simulation study is addressed in Section 5 to examine some finite sample properties of the estimators. In Section 6, the proposed methodology is illustrated by means of a real data set. In Section 7, we offer some concluding remarks.

2 The BCH-PVF cure rate model

Let Z be a Poisson random variable with mean $\theta\, W$, where $\theta > 0$ is a real constant and W is a random variable (frailty) varying over the positive real line with cdf $G(w)$. We emphasize that the frailty W is an unobservable random variable which increases the individual risk if $W > 1$ or decreases if $W < 1$. Under this setup, the marginal distribution of Z is the class of mixed Poisson distributions with pmf

$$P(Z = z) = \frac{1}{z!} \int_0^\infty (w\,\theta)^z \, \mathrm{e}^{-\mathrm{w}\,\theta} \, \mathrm{dG(w)}. \tag{5}$$

The pgf of Z comes from Equation (5) as

$$\Psi_Z(s) = \int_0^\infty \exp[-w\theta(1-s)] dG(w) = \mathcal{L}_W(\theta[1-s]), \tag{6}$$

where $\mathcal{L}_W$ is the Laplace transform of the distribution of W. Consider that the random variable W has the PVF density function (1) with $E(W) = 1$ and $Var(W) = \gamma$. For this parameterization, the Laplace transform of W is

$$\mathcal{L}_W(s) = \exp\left\{\frac{1-\alpha}{\alpha\,\gamma}\left[1-\left(1+\frac{\gamma\,s}{1-\alpha}\right)^{\alpha}\right]\right\}. \tag{7}$$

By combining (6) and (7), the mean and variance of Z are $E(Z) = \theta$ and $Var(Z) = \theta + \gamma\theta^2$, respectively. Note that $Var(Z) > E(Z)$, and then γ is a kind of dispersion parameter.

A survival model can be defined by assuming that the frailty random variable Z in (3) (representing the number of latent risks) has a mixed Poisson distribution with pmf (5). Under this assumption, the unconditional population survival function of T in Equation (3) is

$$S(t) = \exp\left\{\frac{1-\alpha}{\alpha\,\gamma}\left[1-\left(1+\frac{\gamma\,\theta\,F_0(t)}{1-\alpha}\right)^{\alpha}\right]\right\},\ t>0, \tag{8}$$

where $0 < \alpha < 1$, $\gamma > 0$, $\theta > 0$ and $F_0(t)$ is the baseline cdf of T (i.e., if Z is constant equal to one).

Henceforth, we refer to Equation (8) as the survival function of the BCH-PVF cure rate model. The cure fraction is determined from (8) as

$$p_0 = \lim_{t\to\infty} S(t) = \exp\left\{\frac{1-\alpha}{\alpha\,\gamma}\left[1-\left(1+\frac{\gamma\,\theta}{1-\alpha}\right)^{\alpha}\right]\right\} > 0, \tag{9}$$

which implies that (8) is not a proper survival function. The cure fraction tends to zero when $\theta \longrightarrow \infty$ and it goes to one when $\theta \longrightarrow 0$.

The probability that an individual is immune or cured of a disease, given that has been followed up for time $t > 0$ after treatment, has the form

$$\pi(t) = Pr(Z = 0|T > t) = \exp\left\{\frac{1-\alpha}{\alpha\,\gamma}\left[\left(1+\frac{\gamma\,\theta\,F_0(t)}{1-\alpha}\right)^{\alpha}-\left(1+\frac{\gamma\,\theta}{1-\alpha}\right)^{\alpha}\right]\right\},$$

which is an increasing function in t. For $t = 0$, corresponding to no information regarding the immunity of an individual, other than overall probability being immune, the probability is equal to p_0, whereas as $t \longrightarrow \infty$ corresponding to certainly of immunity if the individual's lifetime is very large, the probability tends to one.

The unconditional population density of T corresponding to (8) is

$$f(t) = \theta f_0(t) S(t)\left[1+\frac{\gamma\,\theta\,F_0(t)}{1-\alpha}\right]^{\alpha-1},$$

and then the hrf of T becomes

$$h(t) = \theta f_0(t) \left[1 + \frac{\gamma\, \theta\, F_0(t)}{1-\alpha}\right]^{\alpha-1}.$$

We note that $h(t) \to 0$ at a fast rate when $t \to \infty$ and $\int_0^\infty h(t)dt < \infty$. Further, if $\alpha \longrightarrow 0$ and $\gamma \longrightarrow 0$, the model reduces to the BCH model (Yakovlev & Tsodikov, 1996), that is, $h(t) = \theta f_0(t)$, which has the proportional hazard structure when the covariates are modeled through θ. The cumulative hazard function (chf) of T is

$$H(t) = \frac{\alpha - 1}{\alpha\gamma}\left[1 - \left(1 + \frac{\gamma\, \theta\, F_0(t)}{1-\alpha}\right)^{\alpha}\right].$$

The chf of T tends to zero when $t \longrightarrow 0$, whereas it goes to $-\log(p_0)$ when $t \longrightarrow \infty$, which implies that it is bounded by $-\log(p_0)$, i.e., $H(t) \leq -\log(p_0)$. Figure 1 reveals the flexibility of this distribution in terms of the additional parameter $\alpha < 1$ and dispersion parameter $\gamma > 0$.

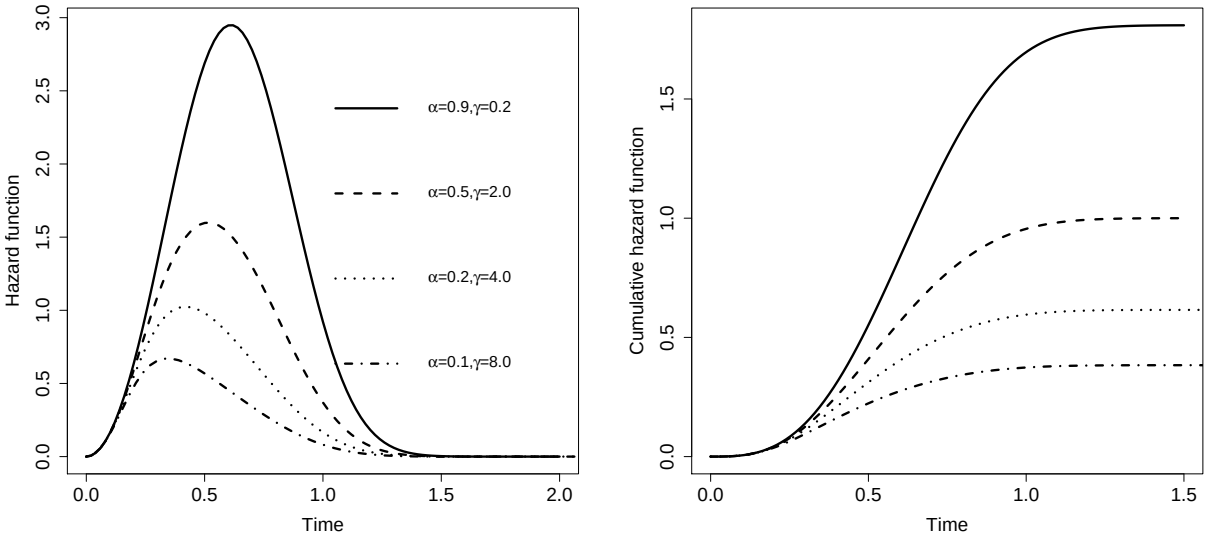

Figure 1: Hazard functions (left panel) and chf (right panel) of the BCH-PVF cure rate model with parameters $\alpha = 0.1, 0.2, 0.5, 0.9$, $\gamma = 0.2, 2.0, 4.0, 8.0$, $\theta = 2.0$ and baseline hrf $h_0(t) = 3t^2$.

It follows from this proposal that the BCH-Gamma model (Cancho *et al.*, 2011b) is a special case when $\alpha \longrightarrow 0$. Moreover, it gives the cure rate model, so-called the BCH-Inverse Gaussian model,

when $\alpha = 0.5$. Table 1 provides the survival function $S(t)$, the improper hazard function $h(t)$ and the cure fraction p_0 corresponding to three special cases of the BCH-PVF model.

Table 1: Long-term survival function ($S(t)$), hazard function ($h(t)$) and cured fraction (p_0).

Model	$S(t)$	$h(t)$	p_0
BCH ($\alpha \longrightarrow 1$)	$\mathrm{e}^{-\theta F_0(t)}$	$\theta f_0(t)$	$\mathrm{e}^{-\theta F_0(t)}$
BCH-Gamma ($\alpha \longrightarrow 0$)	$[1+\gamma\theta F_0(t)]^{-1/\gamma}$	$\frac{\theta f_0(t)}{1+\gamma\theta F_0(t)}$	$(1+\gamma\theta)^{-1/\gamma}$
BCH-Inverse Gaussian ($\alpha = 0.5$)	$\mathrm{e}^{\frac{1}{\gamma}\left(1-\sqrt{1+2\gamma\theta F_0(t)}\right)}$	$\frac{\theta f_0(t)}{\sqrt{1+2\gamma\theta F_0(t)}}$	$\mathrm{e}^{\frac{1}{\gamma}\left(1-\sqrt{1+2\gamma\theta}\right)}$

The (proper) survival function for the individuals under risk in the population, say $S_{\mathrm{R}}(t) = P(T > t \ | Z \geq 1)$, has the from

$$S_{\mathrm{R}}(t) = \frac{\exp\left\{\frac{1-\alpha}{\alpha\,\gamma}\left[1-\left(1+\frac{\gamma\,\theta\,F_0(t)}{1-\alpha}\right)^{\alpha}\right]\right\} - p_0}{1-p_0}. \tag{10}$$

Note that $S_{\mathrm{R}}(t)$ is a proper survival function since $S_{\mathrm{R}}(0) = 1$ and $S_{\mathrm{R}}(\infty) = 0$. So, the density function for the individuals under risk in the population is

$$f_R(t) = \frac{f(t)}{1-p_0}, \tag{11}$$

which is a proper density function. Then, the corresponding hrf (under risk) reduces to

$$h_R(t) = \frac{S(t)}{S(t)-p_0}\,h(t). \tag{12}$$

Equation (12) has a multiplicative factor $\frac{S(t)}{S(t)-p_0} > 1$ compared to the hazard function $h(t)$ of the entire population. Moreover, $h_R(t) \longrightarrow h_0(t)$ when $t \longrightarrow \infty$ and hence $h_R(t)$ converges to the baseline hazard function.

The relationship between the model (8) and the mixture cure rate model (Boag, 1949; Berkson & Gage, 1952) can be expressed as

$$S(t) = p_0 + (1\ -\ p_0)\ S_{\mathrm{R}}(t), \tag{13}$$

where $S_{\mathrm{R}}(t)$ comes from (10). Hence, $S(t)$ is a mixture cure rate model with cure fraction $p_0 = \exp\left\{\frac{1-\alpha}{\alpha\,\gamma}\left[1-\left(1+\frac{\gamma\,\theta}{1-\alpha}\right)^{\alpha}\right]\right\}$ and survival function $S_{\mathrm{R}}(t)$ for individuals under risk in the population.

This result implies that every mixture cure rate model corresponds to some model in (8) for some α, γ, θ and $F_0(t)$.

The following theorem gives the form of the class of cure rate models that includes as a special case the BCH model.

Theorem 2.1 *Let $h(t|Z) = h_0(t)\,Z$ be a basic frailty model, where $h_0(t)$ is the baseline hrf and Z is a discrete random variable having Poisson distribution with mean $\theta\eta$, $\theta > 0$ is a real constant and $\eta > 0$ is a random component (frailty). If the distribution of η has the Laplace transform $\mathcal{L}_\eta(s)$, then the unconditional survival function is*

$$S(t) = \mathcal{L}_\eta[\theta F_0(t)], \tag{14}$$

where $F_0(t) = 1 - \exp[-\int_0^\infty h_0(u)du]$ is the baseline cdf of T.

Proof. For a non-negative random variable T and under the conditions of Theorem 2.1, the conditional survival function of T is $S(t|Z) = [S_0(t)]^Z$. It is easily proved that the unconditional survival function of T can be expressed as

$$S(t) = E_\eta\{E_Z[S_0^Z(t)]\} = \int_0^\infty \left[\sum_{z=0}^\infty S_0(t)^z P(z = z|\theta, \eta)\right] g(\eta)d\eta, \tag{15}$$

where $S_0(t) = 1 - F_0(t)$ is the baseline survival function, $g(\cdot)$ is the density function of η, $P(Z = z|\theta, \eta)$ is the pmf of the Poisson distribution with mean $\theta\eta$ and

$$\sum_{z=0}^\infty S_0(t)^z P(z = z; \theta, \eta) = \mathrm{e}^{-\theta\eta} \sum_{n=1}^\infty \frac{[\eta\,\theta S_0(t)]^z}{z!} = \mathrm{e}^{-\theta\eta[1-S_t(t)]}. \tag{16}$$

Inserting (16) in Equation (15) gives

$$S(t) = \int_0^\infty \mathrm{e}^{-\theta\,\eta[1-S_0(t)]} g(\eta)d\eta = \mathcal{L}_\eta\{\theta[1 - S_0(t)]\}.$$

The last step follows from the definition of the Laplace transform. □

The proposed model (8) can be obtained from Theorem 2.1. In order to be more specific, the random variable η with PVF density (1) implies that the model (14) is equal to the BCH-PVF model. Further, assuming that η is a random variable with a positive stable distribution (Wienke, 2011), and unity scale parameter to avoid identifiability problems, the Laplace transformation of η is $\mathcal{L}_\eta(s) = \exp\left(-\eta^\phi\right)$, $\phi \in (0, 1]$. Hence, based on Theorem 2.1, we have a cure rate model with survival function $S(t) = \exp\{-\theta^\phi[1 - S_0(t)]^\phi\}$.

3 Mathematical properties

For the previous quantity in (10), the power series holds everywhere on t

$$\left[1-\left(1+\frac{\gamma\,\theta\,F_0(t)}{1-\alpha}\right)^{\alpha}\right]=\sum_{i=1}^{\infty}s_i\,F_0(t)^i,$$

where $s_1=-\alpha\,\gamma\,\theta/(1-\alpha)$, $s_2=\alpha\,(\gamma\,\theta)^2/[2(1-\alpha)]$, $s_3=-\alpha\,(\gamma\,\theta)^3\,(2-\alpha)/[6(1-\alpha)^2]$, etc.

Then, the survival function for the individuals under risk can be rewritten as

$$S_{\rm R}(t)=\frac{\exp\left\{\frac{1-\alpha}{\alpha\,\gamma}\,\sum_{i=1}^{\infty}s_i\,F_0(t)^i\right\}-p_0}{1-p_0}.$$

Consider the exponential partial Bell polynomials defined by

$$\exp\left(u\sum_{i\geq 1}x_i\,\frac{z^i}{i!}\right)=\sum_{n,k\geq 0}B_{n,k}\,u^k\,\frac{z^n}{n!},\tag{17}$$

where

$$B_{n,k}=B_{n,k}(x_1,x_2,\cdots,x_{n-k+1})=\sum\frac{n!}{c_1!\,c_2!\cdots(1!)^{c_1}(2!)^{c_2}\cdots}\,x_1^{c_1}\,x_2^{c_2}\cdots,$$

and the summation takes place over all integers $c_1,c_2,\cdots\geq 0$ that verify $c_1+c_2+c_3+\cdots=k$ and $c_1+2c_2+3c_3+\cdots=n$. These polynomials can be computed by BellY[n,k,{1,...,n-k+1}] in MATHEMATICA and IncompleteBellB(n, k, z[1],z[2],..., z[n = k+1]) in MAPLE. Further, $B_{0,0}=1$, $B_{n,0}=0$ (for $n\geq 1$) and $B_{0,k}=0$ (for $k\geq 1$).

Hence,

$$S_{\rm R}(t)=\frac{1}{(1-p_0)}\sum_{n=0}^{\infty}q_n\,\frac{F_0(t)^n}{n!}-\frac{p_0}{(1-p_0)},$$

where $q_n=\sum_{k=0}^{\infty}\left(\frac{1-\alpha}{\alpha\,\gamma}\right)^k B^{\star}_{n,k}$ and $B^{\star}_{n,k}=B_{n,k}(s_1,2!s_2,\cdots,(n-k+1)!s_{n-k+1})$ (for $n,k\geq 0$).

By differentiating $S_{\rm R}(t)$, the density function of T (for the individuals under risk) has the form

$$f_{\rm R}(t)=\sum_{n=0}^{\infty}w_n\,h_{n+1}(t),\tag{18}$$

where (for $n\geq 0$) $w_n=-q_{n+1}/[(1-p_0)(n+1)!]$ and $h_{n+1}(t)=(n+1)\,F_0(t)^n\,f_0(t)$ is the exponentiated density with baseline pdf $f_0(t)$ and power parameter $(n+1)$.

Equation (18) reveals that the density function for the individuals under risk is a linear combination of exponentiated densities under the same baseline density $f_0(t)$. Thus, some mathematical properties of the distribution of T can be determined from those of the exponentiated-F_0 (exp-F_0) distribution.

Let Y_{n+1} have the exp-F_0 distribution with power parameter $n+1$, i.e., $Y_{n+1} \sim \text{exp-}F_0(n+1)$. So, $Y_0 = T \sim F_0$. We define the (r,s)th probability weighted moment (PWM) of T by $\tau_{r,s} = \mathrm{E}[T^r\, F_R(T)^s]$ (for $r, s = 0, 1, \ldots$). The PWMs are usually evaluated numerically since they are available in closed-form for few distributions.

The rth ordinary moment of T can be expressed from (18) as

$$\mu'_r = \mathrm{E}(T^r) = \sum_{n=0}^{\infty} w_n\, \mathrm{E}(Y_{n+1}^r) = \sum_{n=0}^{\infty} (n+1)\, w_n\, \tau_{r,n}. \tag{19}$$

Equation (19) shows that the ordinary moments of T can be determined from an infinite weighted linear combination of PWMs of the baseline F_0 distribution.

In a similar manner, the moment generating function (mgf) of T is $M_R(v) = \mathrm{E}(\mathrm{e}^{v\,T})$. Let $M_{n+1}(v)$ be the mgf of Y_{n+1} and $Q_0(u) = F_0^{-1}(u)$ be the baseline quantile function (qf).

It follows from Equation (18)

$$M_R(v) = \sum_{n=0}^{\infty} w_n\, M_{n+1}(v) = \sum_{n=0}^{\infty} (n+1)\, w_n\, \rho_n(v),$$

where

$$\rho_n(v) = \int_0^1 \exp\left[v\, Q_0(u)\right]\, u^n du.$$

Hence, $M_R(v)$ can be determined for any baseline distribution F_0 by integrating its qf.

4 Bayesian Inference

Consider n individuals and let Z_i be the number of latent risks for the ith individual, $i = 1, \ldots, n$. We assume that the Z_i's are independent Poisson random variables with mean $\theta\, w_i$ and that the w_i's are i.i.d. random variables having the PVF distribution with unity mean and variance γ. Further, suppose that the distribution of T_i $(i = 1, \ldots, n)$ can be represented by a basic frailty model $h(t_i|Z_i) = h_0(t_i|\boldsymbol{\phi})\, Z_i$, where $h_0(t_i|\boldsymbol{\phi})$ is the baseline hrf for T_i with parameter vector $\boldsymbol{\phi}$. We consider that the T_i's are conditionally independent given Z_i. Let C_i denote the censoring time such that it is observed $y_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_{ki})$, where $\delta_i = 1$ if T_i is a lifetime and $\delta_i = 0$ if it is right censored.

Let $\boldsymbol{y}' = (y_1, \cdots, y_n)$, $\boldsymbol{\delta}' = (\delta_1, \cdots, \delta_n)$, $\boldsymbol{Z}' = (Z_1, \cdots, Z_n)$ and $\boldsymbol{w}' = (w_1, \cdots, w_n)$. The "complete data" is denoted by $\mathcal{D} = (n, \boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{Z}, \boldsymbol{w})$, and the observed data is $\mathcal{D}_{obs} = (n, \boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{Z}, \boldsymbol{w})$, where $\boldsymbol{Z}$ and

$\boldsymbol{w}$ are unobserved random vectors. The likelihood function for $(\theta, \boldsymbol{\phi})$ based on the complete data $\mathcal{D}$ can be expressed as

$$L(\theta, \boldsymbol{\phi}|\mathcal{D}) = \prod_{i=1}^{n} [S_0(y_i|\boldsymbol{\phi})]^{Z_i-\delta_i} [Z_i f_0(y_i|\boldsymbol{\phi})]^{\delta_i} \exp\left\{\sum_{i=1}^{n} [Z_i \log(w_i\theta) - \log(Z_i) - w_i\theta]\right\}. \quad (20)$$

The likelihood function of the observed data is obtained by integrating (20) in relation to $(\boldsymbol{Z}, \boldsymbol{w})$ and assuming a PVF distribution for each w_i with unit mean and variance γ. The likelihood function for $\boldsymbol{\vartheta} = (\alpha, \gamma, \boldsymbol{\phi}, \theta)$ based on the observed data, say $L(\boldsymbol{\vartheta}|\mathcal{D}_{\text{obs}})$, is

$$L(\boldsymbol{\vartheta}|\mathcal{D}_{\text{obs}}) = \prod_{i=1}^{n} [\theta f_0(y_i|, \boldsymbol{\phi})]^{\delta_i} \left[1 + \frac{\gamma\,\theta F_0(y_i;\boldsymbol{\phi})}{1-\alpha}\right]^{(\alpha-1)\delta_i} \exp\left\{\frac{1-\alpha}{\alpha\,\gamma} \sum_{i=1}^{n} \left[1 - \left(1 + \frac{\gamma\,\theta F_0(y_i;\boldsymbol{\phi})}{1-\alpha}\right)^{\alpha}\right]\right\}, \quad (21)$$

where $F_0(y_i;\boldsymbol{\phi})$ is the baseline cdf and $f_0(y_i;\boldsymbol{\phi}) = dF_0(y_i;\boldsymbol{\phi})/dy_i$. Consider the Weibull baseline density re-parameterized as

$$f_0(y_i;\boldsymbol{\phi}) = \phi_1\, y^{\phi_1-1} \exp[\phi_2 - \exp(\phi_2) y^{\phi_1}], \quad (22)$$

whose parameters $\phi_1 > 0$ and $\phi_2 > 0$ control the shape and scale. We incorporate covariates to the BCH-PVF model (8) through the mean latent risks θ. By including covariates, a different parameter θ_i follows for each individual $(i = 1, \ldots n)$. Let $\boldsymbol{x}_i' = (x_{i1}, \cdots, x_{ip})$ be the $p \times 1$ vector of covariates for the ith individual, and let $\boldsymbol{\beta}' = (\beta_1, \cdots, \beta_p)$ denote the corresponding vector of regression coefficients for the failure time T. Further, we relate θ to the covariates by the logarithmic link, say $\theta_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, such that the cure rate for the individual i who failed according to T is $\exp(-\theta_i) = \exp[-\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]$ $(i = 1, \ldots n)$. The likelihood function for the model parameters is

$$\begin{aligned} L(\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\phi}|\mathcal{D}_{\text{obs}}) = \exp\left(\sum_{i=1}^{n} \delta_i \boldsymbol{x}_i'\boldsymbol{\beta}\right) \prod_{i=1}^{n} [f_0(y_i|\phi_k)]^{\delta_i} \left[1 + \frac{\gamma \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) F_0(y_i;\boldsymbol{\phi})}{1-\alpha}\right]^{(\alpha-1)\delta_i} \\ \times \exp\left\{\frac{1-\alpha}{\alpha\gamma} \sum_{i=1}^{n} \left[1 - \left(1 + \frac{\gamma \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) F_0(y_i;\boldsymbol{\phi})}{1-\alpha}\right)^{\alpha}\right]\right\}, \end{aligned} \quad (23)$$

where $\mathcal{D}_{\text{obs}} = (n, \boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{X})$ and $\boldsymbol{X}$ is the $n \times p$ model matrix with rows $\boldsymbol{x}_i'$.

Further, some inferential tools are investigated under a Bayesian point of view. We consider the joint improper prior for $(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}, \gamma)$

$$\pi(\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\phi}) = \pi(\alpha)\pi(\boldsymbol{\beta})\pi(\gamma)\pi(\boldsymbol{\phi}) = \pi(\alpha)\pi(\phi_1, \phi_2)\pi(\gamma). \quad (24)$$

Equation (24) means that α, $\boldsymbol{\beta}$, γ and $\boldsymbol{\phi}$ are independent priors with an improper uniform prior for $\boldsymbol{\beta}$, i.e., $\pi(\boldsymbol{\beta}) \propto 1$, $\pi(\alpha)$ and $\pi(\gamma)$ (they can be any proper priors) and $\pi(\phi_1, \phi_2)$ are considered independents. Further, we assume that

$$\pi(\phi_1, \phi_2) = \pi(\phi_1|a_0, b_0)\, \pi(\phi_2),$$

where

$$\pi(\phi_1|a_0, b_0) \propto \alpha^{a_0-1}\mathrm{e}^{-b_0\phi_1}, \;\; \pi(\phi_2) \propto \mathrm{e}^{-c_0\phi_2^2},$$

and a_0, b_0 and c_0 are specified hyper-parameters. Combining the likelihood function (23) and the prior distribution in (24), the joint posterior distribution for α, $\boldsymbol{\beta}, \boldsymbol{\phi}$ and γ can be expressed as

$$\pi(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}, \gamma|\boldsymbol{\mathcal{D}}_{\text{obs}}) \propto L(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma|\boldsymbol{\mathcal{D}}_{\text{obs}})\pi(\alpha)\pi(\phi_1|a_0, b_0)\pi(\phi_2)\pi(\gamma). \tag{25}$$

The following theorem can provide the properties for the posterior distribution (25) under the improper prior distribution (24).

Theorem 4.1 *Let X^* be a $n \times p$ matrix with rows $\delta_i \boldsymbol{x}_i'$, $d = \sum_{i=1}^n \delta_i$ and $\vartheta = (\boldsymbol{\beta}, \boldsymbol{\phi}, \alpha, \gamma)$. Then, if (i) X^* is of full rank, (ii) $\pi(\phi_2)$ is proper, and (iii) $b_0 > 0$ and $a_0 > -d$, the posterior distribution (25) is proper.*

The proof of Theorem 4.1 is given in Appendix A.

The joint posterior density (25) is analytically intractable because the integration of the joint posterior density is not easy to perform. An alternative is to rely on Markov Chain Monte Carlo (MCMC) simulations carried out with the Gibbs sampler algorithm to make inference, which is now discussed. In this direction, we consider a modified version of the collapsed Gibbs technique of Liu (1994) to sample from the posterior distribution.

We introduce some auxiliary (latent) variables to facilitate the Gibbs sampler: $\boldsymbol{Z}' = (Z_1, \cdots, Z_n)$ and $\boldsymbol{w}' = (w_1, \cdots, w_n)$. The joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma, \boldsymbol{Z}, \boldsymbol{w})$ is

$$\begin{aligned}\pi(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}, \gamma, \boldsymbol{N}, \boldsymbol{w}|\boldsymbol{\mathcal{D}}_{\text{obs}}) = &\prod_{i=1}^n \left[S(y_i|\boldsymbol{\phi})\right]^{Z_i-\delta_i} \left[Z_i f(y_i|\boldsymbol{\phi})\right]^{\delta_i} g(w_i|\alpha, \gamma) \\ &\times \exp\left\{\sum_{i=1}^n \left[Z_i \log(w_i\theta_i) - \log(Z_i) - w_i\theta_i\right]\right\} \\ &\times \pi(\alpha)\pi(\phi_1|a_0, b_0)\pi(\phi_2)\pi(\gamma),\end{aligned} \tag{26}$$

where $\theta_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, $b_0 > 0$, $a_0 > -\sum \delta_i$, $c_0 > 0$ and $g(w_i|\alpha,\gamma)$ is the PVF density with mean zero and variance γ given in (1). We require to sample from the following conditional distributions $[\boldsymbol{\phi}|\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{Z},\boldsymbol{w},\boldsymbol{\mathcal{D}}_{\text{obs}}]$ and $[\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{Z},\boldsymbol{w}|\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$ to run the Gibbs sampler.

The posterior density for $[\boldsymbol{\phi}|\boldsymbol{\beta},\gamma,\boldsymbol{Z},\boldsymbol{w},\boldsymbol{\mathcal{D}}_{\text{obs}}]$ is

$$\pi(\boldsymbol{\phi}|\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{Z},\boldsymbol{w},\boldsymbol{\mathcal{D}}_{\text{obs}}) \propto \phi_1^{d+a_0-1}\exp\{-b_0\phi_1 + d\phi_2 + \sum_{i=1}^{n}[\delta_i\phi_1\log(y_i) - Z_i e^{\phi_2} y_i^{\phi_1} - c_0\phi_2^2]\}.$$

We can show that $\pi(\boldsymbol{\phi}|\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{Z},\boldsymbol{w},\boldsymbol{\mathcal{D}}_{\text{obs}})$ is log-concave in ϕ_1 and ϕ_2. Thus, the adaptive rejection algorithm can be used here to sample $\boldsymbol{\phi}$.

Sampling from $[\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{Z},\boldsymbol{w}|\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$ is the most complicated and expensive part of this algorithm. As pointed out by Chen *et al.* (1999, 2002), sampling the four complete conditional distributions may result in high correlation among α, $\boldsymbol{\beta},\gamma,\boldsymbol{Z}$ and $\boldsymbol{w}$ due to the high dimensions of the latent vectors. A way to remedy this potential problem would be to apply the collapsed procedure of Liu (1994). We can write

$$[\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{Z},\boldsymbol{w}|\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}] = [\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{w}|\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]\,[\boldsymbol{Z}|\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]. \tag{27}$$

We draw $(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{w})$ from (27) by collapsing $\boldsymbol{Z}$, which is crucial for achieving convergence in the MCMC algorithm. Considering $[\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{w}|\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$, we draw $\boldsymbol{\beta}$ from $[\boldsymbol{\beta}|\alpha,\gamma,\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$ and $(\alpha,\gamma,\boldsymbol{w})$ from $\pi(\alpha,\gamma,\boldsymbol{w}|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}})$.

The density of $[\boldsymbol{\beta}|\alpha,\gamma,\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$ can be expressed as

$$\pi(\boldsymbol{\beta}|\alpha,\gamma,\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}) \propto \exp\left[\sum_{i=1}^{n}\delta_i\boldsymbol{x}_i'\boldsymbol{\beta} - w_i F(y_i|\boldsymbol{\phi})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right]. \tag{28}$$

Note that $\pi(\boldsymbol{\beta}|\gamma,\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}})$ is log-concave in each component of $\boldsymbol{\beta}$ and therefore we can use the adaptive rejection algorithm to draw $\boldsymbol{\beta}$. We adopt the collapsed Gibbs procedure to $[\alpha,\gamma,\boldsymbol{w}|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$. That is, we draw (α,γ) from $[\alpha,\gamma|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$ by collapsing $\boldsymbol{w}$, and draw $\boldsymbol{w}$ from $[\boldsymbol{w}|\alpha,\gamma,\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$.

The conditional posterior density for $[\alpha,\gamma|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$ takes the form

$$\begin{aligned}\pi(\alpha,\gamma|\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}) \propto \exp\Bigg\{&\frac{1-\alpha}{\alpha\,\gamma}\sum_{i=1}^{n}\left[1-\left(1+\frac{\gamma\exp(\boldsymbol{x}_i'\boldsymbol{\beta})F_0(y_i;\boldsymbol{\phi})}{1-\alpha}\right)^{\alpha}\right]\\ &+(\alpha-1)\sum_{i=1}^{n}\delta_i\log\left[1+\frac{\gamma\exp(\boldsymbol{x}_i'\boldsymbol{\beta})F_0(y_i;\boldsymbol{\phi})}{1-\alpha}\right]\Bigg\}\pi(\alpha)\pi(\gamma).\end{aligned} \tag{29}$$

Generating α and γ from (29) is hard since $\pi(\alpha, \gamma|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}})$ is not log-concave. Thus, we consider the following Metropolis-Hasting (M-H) algorithm with a transformation to draw α and γ. Thus, since $0 < \alpha < 1$ and $\gamma > 0$, we set

$$\alpha = \frac{\mathrm{e}^{\eta_1}}{(1+\mathrm{e}^{\eta_1})}, \qquad \eta_1 \in \mathbb{R}. \tag{30}$$

and

$$\gamma = \mathrm{e}^{\eta_2}, \qquad \eta_2 \in \mathbb{R}. \tag{31}$$

Then,

$$\pi(\eta_1, \eta_2|\ \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}}) = \pi(\alpha, \gamma|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}})\ \frac{\mathrm{e}^{\eta_1+\eta_2}}{(1+\mathrm{e}^{\eta_1})}.$$

Instead of generating directly α and γ, we generate $\boldsymbol{\eta}' = (\eta_1, \eta_2)$ by considering a normal proposal with mean $\hat{\boldsymbol{\eta}}$ and variance-covariance matrix $\hat{\Sigma}_{\hat{\boldsymbol{\eta}}}$, where $\hat{\boldsymbol{\eta}}$ is the maximizer of the logarithm of $\pi(\boldsymbol{\eta}|\ \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}})$ and $\hat{\Sigma}_{\hat{\boldsymbol{\eta}}}$ is the inverse matrix of minus second derivatives of $\log \pi(\boldsymbol{\eta}|\ \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}})$ evaluated at $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}$, that is,

$$\widehat{\Sigma}_{\hat{\boldsymbol{\eta}}} = \frac{\partial^2 \log \pi(\boldsymbol{\eta}|\ \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}\Big|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}}.$$

The M-H algorithm to generate $\boldsymbol{\eta}$ can be stated as follows:

(i) Let $\boldsymbol{\eta}$ be the current value.

(ii) Generate a proposal value $\boldsymbol{\eta}^*$ from $N_2(\hat{\boldsymbol{\eta}}, \hat{\Sigma}_{\hat{\boldsymbol{\eta}}})$.

(iii) Move from $\boldsymbol{\eta}$ to $\boldsymbol{\eta}^*$ with probability

$$\min\left\{\frac{\pi(\boldsymbol{\eta}^*|\ \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}})\psi(\boldsymbol{\eta}; \hat{\boldsymbol{\eta}}; \widehat{\Sigma}_{\hat{\boldsymbol{\eta}}})}{\pi(\boldsymbol{\eta}|\ \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}})\psi(\boldsymbol{\eta}^*; \hat{\boldsymbol{\eta}}; \widehat{\Sigma}_{\hat{\boldsymbol{\eta}}})}, 1\right\},$$

where $\psi(\cdot)$ is the bivariate normal density function.

We obtain $\boldsymbol{\eta}$ and calculate α and γ from Equations (30) and (31), respectively.

We can demonstrate that the joint conditional density function of $[\boldsymbol{w}|\gamma, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}}]$ is

$$\pi(\boldsymbol{w}|\alpha, \gamma, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\mathcal{D}}_{\text{obs}}) \propto \prod_{i=1}^{n} w_i^{\delta_i} \exp\left[-w_i \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) F_0(y_i|\boldsymbol{\phi})\right] g(w_i|\alpha, \gamma), \tag{32}$$

where

$$g(w_i|\alpha, \gamma) = \mathrm{e}^{\frac{(1-\alpha)}{\gamma}(-\mathrm{w}-\frac{1}{\alpha})}\ \frac{1}{\pi} \sum_{\mathrm{k}=1}^{\infty} (-1)^{\mathrm{k}+1} \frac{[(1-\alpha)/\gamma]^{\mathrm{k}(1-\alpha)}\Gamma(\mathrm{k}\alpha+1)}{\gamma \alpha^{\mathrm{k}} \mathrm{k}!} \mathrm{w_i}^{-\mathrm{k}\alpha-1} \sin(\mathrm{k}\alpha\pi).$$

The rejection algorithm to generate w_i (for $i = 1, \ldots, n$) follows as:

(i) Draw $w_i \sim \text{Gamma}(\delta_i + 1, \exp(\boldsymbol{x}_i'\boldsymbol{\beta})F(y_i|\boldsymbol{\phi}))$ and $u \sim U(0,1)$.

(ii) Accept w_i if $u \leq \frac{g(w_i|\alpha,\gamma)}{\max\limits_{w_i} g(w_i|\alpha,\gamma)}$, otherwise go to (i).

Finally, we draw $\boldsymbol{Z}$ from $[\boldsymbol{Z}|\boldsymbol{\beta},\gamma,\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}}]$. Since

$$Z_i|\boldsymbol{\beta},\gamma,\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{\mathcal{D}}_{\text{obs}} \sim \text{Poisson}(w_i \exp(\boldsymbol{x}_i'\boldsymbol{\beta})S_0(y_i|\boldsymbol{\phi})) + \delta_i,\ i = 1,\ldots,n,$$

then sampling $\boldsymbol{Z}$ from its conditional posterior distribution is quite simple.

5 Simulation study

We evaluate some frequentist properties of the Bayes estimators under quadratic loss of the model parameters by carrying out a simulation study from the BCH-PVF model (8) with a Weibull baseline distribution and parameter vector $\boldsymbol{\phi} = (2, 0.6)$. The proposed model has parameters $\theta_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})$ $(i = 1,\ldots,n)$, where $\beta_0 = -0.5$, $\beta_1 = 0.3$, $\beta_2 = 1$, $\gamma = 2.0$ and $\alpha = 0.6$. The covariate x_{1i} is generated from a Bernoulli distribution with success probability 0.5 and x_{2i} is generated from the uniform $U(0,1)$ distribution. The censored times C_i are sampled from the uniform $U(0,\tau_i)$ distribution, where τ_i is defined to control the proportion of censored observations. We consider the proportion of censored observations around 55%. The observed times are generated as follows:

1. Generate $u_i \sim U(0,1)$;
2. If $u_i < p_{0i}$, then $t_i = \infty$, else $y_i = F^{-1}\left(\frac{1-\alpha}{\theta_i\gamma}\left[(1 - \frac{\alpha\gamma}{1-\alpha}\log u_i)^{1/\alpha} - 1\right]; \boldsymbol{\phi}\right)$, where $F^{-1}(\cdot;\boldsymbol{\phi})$ is the Weibull qf with parameter vector $\boldsymbol{\phi} = (2, 0.6)$;
3. Generate a censoring time C_i from $U(0,\tau_i)$. If $T_i \leq c_i$, then $t_i = T_i$ and $\delta_i = 1$, else $t_i = C_i$ and $\delta_i = 0$.

For the simulations, $n = 100, 200, 400$ and 800. For the Bayesian estimates, we use non-informative prior given by (24), $\pi(\boldsymbol{\beta}) \propto 1$, $\pi(\alpha) \propto \alpha(1-\alpha)^2$, $\pi(\gamma) \propto \text{e}^{-0.01\gamma}$, $\pi(\phi_1) \propto \text{e}^{-0.01\phi_1}$ and $\pi(\phi_2) \propto \text{e}^{-0.001\phi_2^2/2}$. For each configuration, we conduct 500 simulations to calculate the averages of the Bayes estimates, standard deviations (SDs), biases, roots of the mean squared errors (RMSEs) of the estimates and empirical coverage probabilities (CPs) corresponding to the 95% highest posterior density (HPD) intervals for the parameters.

For each simulated data set, the MCMC algorithm run is based on chains of $55,000$ iterations. The first $5,000$ iterations are discarded as a burn-in period to eliminate the effect of the initial values. To avoid correlation between the generated values, we take a spacing of size 5 leading to samples of size $10,000$. The convergence of the chains is monitored by using the methods of Cowles & Carlin (1996) as well as trace plots. The simulation results listed in Table 2 reveal that the Bayes estimates are close to the true parameters. Further, the RMSEs and SDs decrease and the empirical CPs are closer to the nominal coverage levels when the sample size increases, which are all expected if the underlying estimation scheme is working correctly to produce consistent estimates.

A misspecification study is also performed to verify if we can distinguish between the BCH-PVF model and its sub-models in the light of a data set based on the log-pseudo marginal likelihood (LPML) criteria discussed by Gelfand *et al.* (1992) and de Souza *et al.* (2017). We generate 500 samples from the BCH-PVF model with $\alpha = 0, 6, 0.95$ and the other parameters fixed at the same values given above. For these parameters, the average cure rate is approximately 50% for $\alpha = 0.6$ and 35% for $\alpha = 0.95$. Table 3 gives the average of the LPML criteria of the BCH-PVF model and its sub-models. For all cases, the simulations indicate that the data can be fitted by the BCH-PVF model.

6 Application

We provide an application of the BCH-PVF model to a data set on a colon cancer clinical trial reported by the North Central Cancer Treatment Group (Moertel *et al.*, 1990) and analyzed from the frequentist perspective by Barriga *et al.* (2018) in the context of survival cure rate models. The colorectal cancer is the third most common cancer in the United States in both men and women and it is also the third cause of death by cancer in both genders. The American Cancer Society estimates 53,200 deaths during 2020.

The data set comes from an assay for the efficiency of treatment by Fluorouracil (5-FU) along Levamisole in randomly selected patients with Stage C of colorectal cancer (clinical stage) and the efficiency by Levamisole alone in order to prevent recurrence. The study includes 929 patients observed after the full resection of the tumor surgically and subsequent follow-up of the patient for a median time of seven years. After deleting subjects with incomplete data and missing observation times, we have a subset of $n = 888$ patients with approximately 50% of censoring. The time until recurrence

Table 2: Results from the simulation study.

		Parameters						
Sample size		α	γ	ϕ_1	ϕ_2	β_0	β_1	β_2
$n = 100$	Mean	0.404	2.108	2.073	0.414	-0.472	0.353	1.160
	SD	0.411	0.332	0.271	0.269	0.593	0.433	0.826
	Bias	-0.196	0.108	0.073	-0.086	0.028	0.053	0.160
	RMSE	0.455	0.349	0.281	0.282	0.594	0.436	0.841
	CP	0.932	0.937	0.912	0.937	0.944	0.958	0.951
$n = 200$	Mean	0.454	2.108	2.049	0.459	-0.514	0.347	1.118
	SD	0.391	0.340	0.215	0.202	0.400	0.306	0.603
	Bias	-0.146	0.108	0.049	-0.041	-0.014	0.047	0.118
	RMSE	0.418	0.357	0.220	0.206	0.400	0.310	0.614
	CP	0.957	0.965	0.938	0.957	0.956	0.968	0.946
$n = 400$	Mean	0.513	2.089	2.014	0.469	-0.512	0.318	1.073
	SD	0.360	0.376	0.166	0.163	0.294	0.207	0.404
	Bias	-0.087	0.089	0.014	-0.031	-0.012	0.018	0.073
	RMSE	0.371	0.386	0.166	0.166	0.294	0.208	0.410
	CP	0.941	0.948	0.934	0.949	0.948	0.947	0.940
$n = 800$	Mean	0.534	2.075	2.007	0.483	-0.515	0.311	1.032
	SD	0.314	0.417	0.126	0.124	0.209	0.149	0.290
	Bias	-0.066	0.075	0.007	-0.017	-0.015	0.011	0.032
	RMSE	0.321	0.424	0.126	0.125	0.210	0.149	0.292
	CP	0.947	0.949	0.952	0.936	0.955	0.954	0.949

of the disease (in years) is the response variable. The following variables are collected from each patient: t: observed time (recurrence or censored) (in years); x_1: treatment (observation, Levamisole, Levamisole+5-FU), where 5-FU=Fluorouracil; x_2: extent of local spread (1=submucosa, 2=muscle, 3=serosa, 4=contiguous structures); x_3: time from surgery to registration (0=short, 1=long) and x_4: more than 4 positive lymph nodes (0=no, 1=yes).

Table 3: Average of LPML for the fitted models.

α	n	BCH	BCH-IG	BCH-Gamma	BCH-PVF
0.60	100	-81.43	-85.72	-79.67	-76.06
	200	-160.17	-164.38	-158.62	-151.96
	400	-318.53	-322.57	-317.09	-313.36
	800	-632.32	-636.02	-630.65	-626.85
0.95	100	-80.210	-84.376	-78.573	-75.085
	200	-157.889	-161.958	-156.411	-152.851
	400	-312.119	-316.074	-310.677	-306.044
	800	-619.714	-623.504	-618.192	-614.477

The Kaplan–Meier estimate of the survival function of the recurrence time in Figure 2 (left panel) indicates disease-free patients. Further, the Kaplan–Meier estimate of the chf in Figure 2 (right panel) is limited and concave and supports that a distribution with a monotone hazard function could be adequate for modeling these data. Thus, the Weibull density function (22) can be a suitable baseline density for these data.

We consider the BCH-PVF model with all covariates on the mean of latent risks θ to these data (for $i = 1, \ldots, 888$)

$$\theta_i = \exp\left[\beta_0 + \beta_{1_1} x_{i1_1} + \beta_{1_2} x_{i1_2} + \beta_{2_1} x_{i2_1} + \beta_{2_1} x_{i2_2} + \beta_{2_3} x_{i2_3} + \beta_3 x_{i3} + \beta_4 x_{i4}\right] = \exp\left[\boldsymbol{x}_i'\boldsymbol{\beta}\right], \quad (33)$$

where the categorical covariates with more then two levels are taken as dummy variables. For example, for the covariate treatment (x_{1i}), the variables are defined as:

$$x_{i1_1} = \begin{cases} 1, & \text{if, levamisole;} \\ 0, & \text{if, otherwise} \end{cases} \quad \text{and} \quad x_{i1_2} = \begin{cases} 1, & \text{if, Levamisole+5-FU;} \\ 0, & \text{if, otherwise.} \end{cases}$$

For the Bayesian analysis, we adopt non-informative prior in (24) as described in Section 5. Then, the MCMC algorithms are implemented in **R** language. In this example, 55,000 MCMC iterations are performed in all computations after a burn-in of 5,000 iterations and thinning to every fifth. Posterior results are then based on 10,000 iterations of the Markov chain. The convergence of the chains is monitored by the methods of Cowles & Carlin (1996). The trace plots for the parameters of

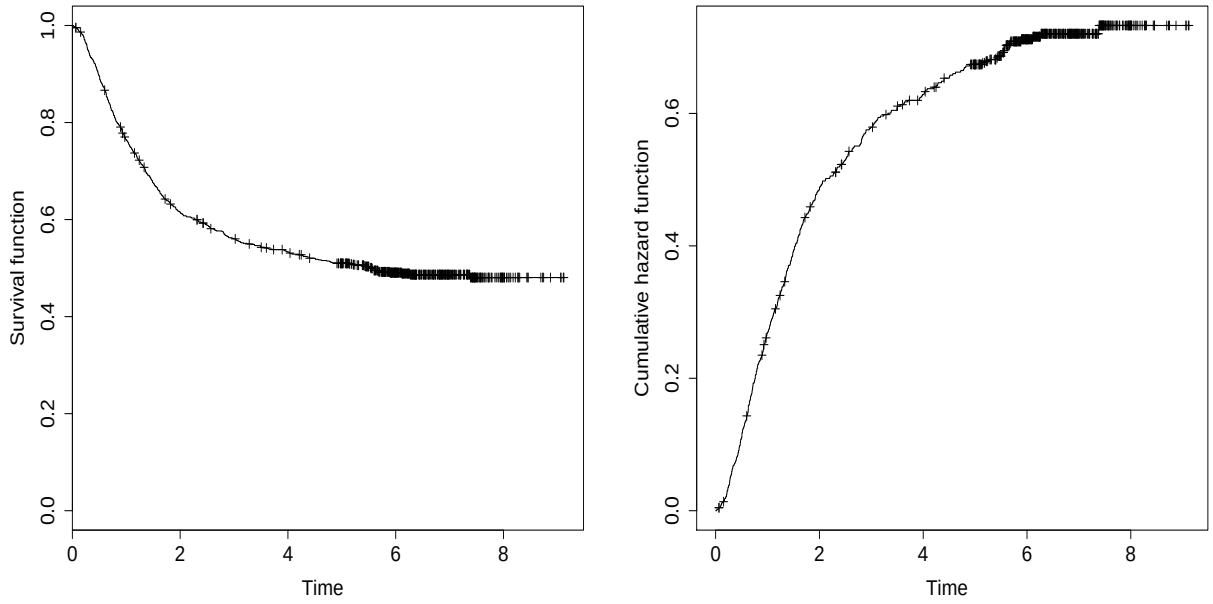

Figure 2: The Kaplan-Meier estimate of survival functions (left panel) and chfs (right panel) for the recurrence time of patients treated of colorectal cancer.

the BCH-PVF model are displayed in Figure 6 (see Appendix B) which indicate convergence of the chains.

The Bayesian estimates under quadratic loss of parameters of the BCH-PVF model and sub-models jointly with the 95% highest HPD intervals are reported in Table 4. All covariates are statistically significant on mean of latent risks at 5% level for all models and the estimate of the shape parameter (ϕ_1) furnishes an evidence against the exponential distribution ($\phi_1 = 1$) as baseline for these data. Figure 3 provides boxplots of the posterior means of the cure rates for the BCH–PVF model and sub-models. By including the covariates in the model, each patient has an individual cure rate. We note From Figure 3 that the median cure rate for all models are approximately the same and equal to 0.50. The ranges of the whiskers of the BCH, BCH-Gamma, BCH-IG and BCH–PVF models are 0.796, 0.773, 0.708 and 0.654, respectively, thus indicating that the last model is the most homogeneous.

The plot of the marginal posterior density of the parameter α for the BCH-PVF model is given in Figure 4 (left panel). This plot reveals that the posterior distribution of α has right skewness

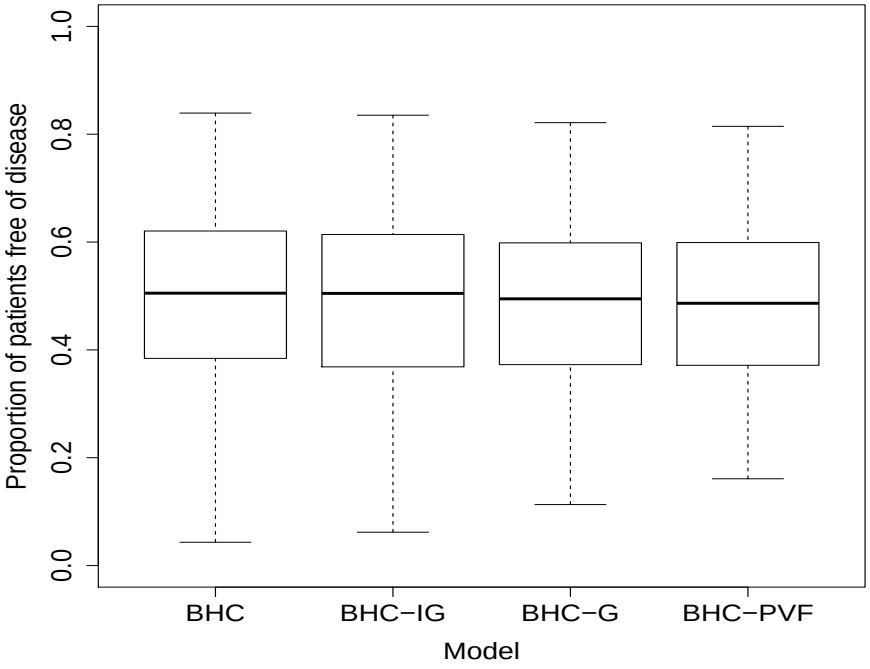

Figure 3: Boxplots of the posterior means of cure rate for all patients.

with mode at 0.205. Also, from Table 4, the posterior mean of α is 0.232 with a 95% HPD interval of (0.067,0.447). This interval does not contain *zero*, 0.5 nor *one*, thus indicating that the proposed BCH-PVF model corresponds to the best fit to the current data.

In Table 5, we apply the selection criteria on the competitive models. According to the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) and log-pseudo marginal likelihood (LPML) (Gelfand *et al.*, 1992; de Souza *et al.*, 2017) criteria, the BCH-PVF model stands out as the best model. Also, to evaluate the adequacy of the fit of the BCH-PVF model, we construct the posterior normalized randomized quantile residuals (Dunn & Smyth, 1996; Rigby & Stasinopoulos, 2005) and the QQ plot of these residuals in Figure 4 (right panel). These plots reveal that the proposed model yields an acceptable fit. Taking into account the values of the DIC and LPLM criteria in Table 5 and the QQ plot in Figure 4, we select the BCH-PVF model as our working model.

For the BHC–PVF model in Table 4, we can note the covariates that significantly influence the

Table 4: Bayesian estimates and 95% HPD intervals of the parameters for the BCH, BCH-G, BCH-IG and BCH–PVF models fitted to the cancer colon data.

	BCH		BCH–IG		BCH–G		BCH–PVF	
Parameter	Mean	HPD (95%) Interval	Mean	HPD (95%) Interval	Mean	HPD (95%) Interval	Mean	HPD (95%) Interval
β_0	-1.339	(-2.456,-0.446)	-0.768	(-2.269,0.641)	-0.525	(-2.088,1.253)	-0.360	(-1.942,1.334)
β_{1_1}	-0.021	(-0.237,0.199)	-0.038	(-0.367, 0.284)	-0.062	(-0.494,0.367)	-0.047	(-0.470 ,0.366)
β_{1_2}	-0.505	(-0.747,-0.271)	-0.743	(-1.125,-0.385)	-0.886	(-1.376,-0.441)	-0.943	(-1.557,-0.461)
β_{2_1}	0.353	(-0.608,1.515)	0.377	(-0.933,1.852)	0.199	(-1.341,1.788)	0.335	(-1.166,1.914)
β_{2_2}	0.956	(0.055,2.088)	1.250	(-0.033,2.662)	1.273	(-0.159,2.802)	1.390	(-0.045, 3.029)
β_{2_3}	1.490	(0.492,2.696)	2.051	(0.699,3.538	2.259	(0.658,3.983)	2.372	(0.741,4.239)
β_3	0.237	(0.035,0.439)	0.351	(0.029,0.667)	0.399	(0.031,0.801)	0.410	(0.008,0.830)
β_4	0.844	(0.646,1.041)	1.352	(1.013,1.708)	1.659	(1.208,2.168)	1.725	(1.189,2.415)
ϕ_1	1.275	(1.178,1.371)	1.562	(1.366,1.763)	1.673	(1.452,1.946)	1.757	(1.471,2.152)
ϕ_2	-0.984	(-1.120,-0.856)	-1.623	(-1.980,-1.278)	-2.019	(-2.751,-1.493)	-2.117	(-3.004, -1.504)
γ	–	–	4.326	(1.273,10.137)	2.560	(1.367,4.197)	4.555	(1.721,9.869)
α	–	–	–	–	–	–	0.232	(0.067,0.447)

Table 5: Criteria DIC and LPML for the fitted models.

	Model			
Criteria	BCH	BCH-IG	BCH-Gamma	BCH-PVF
DIC	2345.92	2334.441	2329.443	2328.79
LPLM	-1173.579	-1167.675	-1165.082	-1164.865

mean of the factor latent risks. In fact, to assess which covariate levels provide different means, we obtain Bayes estimates and 95% HPD interval of the mean risk ratio of the individuals j and k, that is, $\theta_j/\theta_k = \exp[(\boldsymbol{x}_j - \boldsymbol{x}_k)'\boldsymbol{\beta}]$. Table 6 provides the means and HPD intervals of the mean ratio. The mean risk of patients treated with Levamisole+5-FU jointly is 0.422 and 0.403 times the average risks of patients treated with Levamisole and treated with standard treatment (observation), which imply that the cure rate is higher for patients treated with Levamisole+5-FU than patients treated with Levamisole. The HPD 95% interval for the risk ratio of patients treated between Levamisole and standard treatment includes the value *one*, thus indicating that there is no difference in the proportion of cure for patients treated with Levamisole and standard treatment.

We also fit the BCH–PVF cure rate model by each covariate. In Figure 5, we plot the empirical

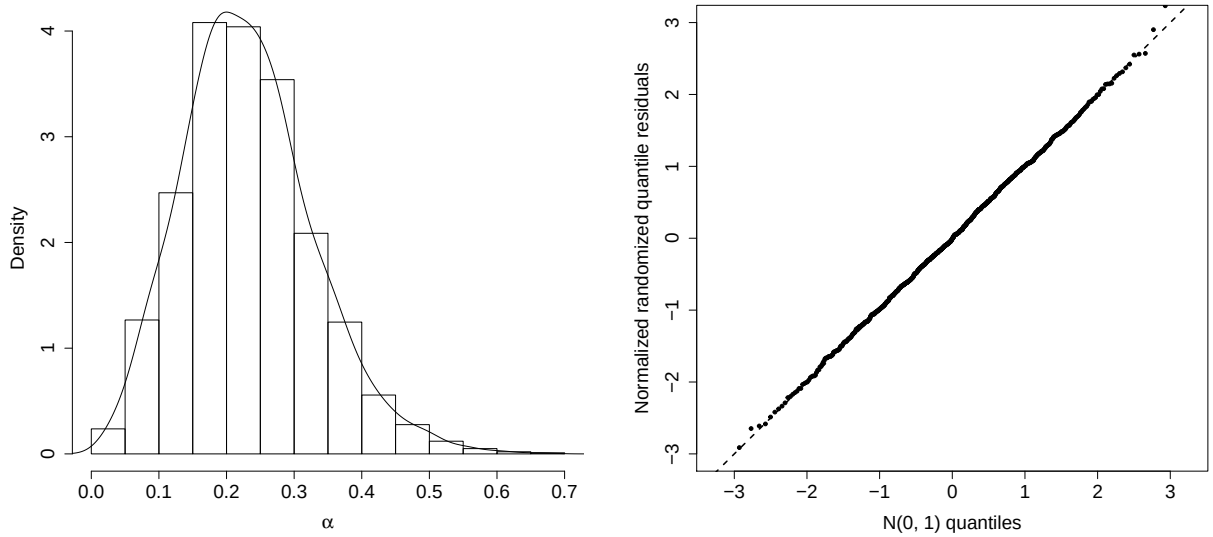

Figure 4: Marginal posterior density for α of the BCH-PVF model (left panel) and QQ plot of the posterior normalized randomized quantile residuals with identity line for the BCH-PVF model (right panel).

survival function and the posterior survival function for each covariate. We conclude that the BCH–PVF cure rate model provides a good fit to these data.

- Regarding the treatment, there is a significant difference among Levamisole + 5-FU levels and observation in relation to the cured proportion.
- Regarding the covariable extent of the spread of the site, there is only a significant difference between the contiguous levels of structures and submucosa. For other levels of this covariate, there is no significant difference.
- There is also a significant difference between the small and large levels of the time of surgery recorded in relation to the cured proportion.
- Finally, there is a significant difference in relation to the existence or not of more than four positive lymph nodes in relation to the cured proportion.

Table 6: Bayesian comparison between the means of latent risks.

Covariate	Comparison	Mean	HPD(95%)
Treatment	Levamisole vs Observation	0.977	(0.625, 1.442)
	Levamisole+5-FU vs Observation	0.403	0.211, 0.630)
	Levamisole+5-FU vs Levamisole	0.422	(0.236, 0.647)
Extent of local spread	Submucosa vs Muscle	0.976	(0.148, 3.209)
	Submucosa vs Serosa	0.332	(0.048, 1.046)
	Submucosa vs Contiguous	0.135	(0.014, 0.477)
	Muscle vs Serosa	0.367	(0.178, 0.647)
	Muscle vs Contiguous	0.150	(0.040, 0.345)
	Serosa vs Contiguous	0.410	(0.146, 0.835)
Time from surgery	Short vs Long	0.678	(0.436, 0.992)
$\geq$4 positive lymph nodes	No vs Yes	0.186	(0.089, 0.305)

The Bayes estimate of the probability of the individuals disease-free (proportion being cured) after follow-up $t > 0$ is

$$\pi(t) = \exp\left\{\frac{1-\alpha}{\alpha\,\gamma}\left[\left(1+\frac{\gamma\,\theta(1-\exp[-(t/\phi_2)^{\phi_1}])}{1-\alpha}\right)^{\alpha}-\left(1+\frac{\gamma\,\theta}{1-\alpha}\right)^{\alpha}\right]\right\}, \tag{34}$$

where θ is given by (33).

We estimate the proportions of patients disease-free from (34) for eight hypothetical patients A, B, C, D, E, F, G and H, who had surgically resection of the tumor and after treatment with selected values for the covariates. Table 7 gives the Bayes estimates under loss quadratic and 95% HPD intervals for the proportion of patients disease-free (cure rate) after treatment ($t = 0$). For example, for the Patient A with extent of local spread in *submucosa*, time from surgery to registration *long*, more than four positive lymph nodes and treatment *Levamisole+5-FU*, and for the Patient E with extent of local spread in *submucosa*, time from surgery to registration *long* and treatment *Levamisole*, we obtain different cured proportions of 0.774 for Patient A and 0.653 for Patient E. The probabilities of these patients being disease-free after four years are 0.956 and 0.936, respectively. The plots reveal that patients who are treated with Levamisole+5-FU have mean time to recurrence disease larger when compared to patients treated with Levamisole, thus indicating the effectiveness of the Levamisole+5-

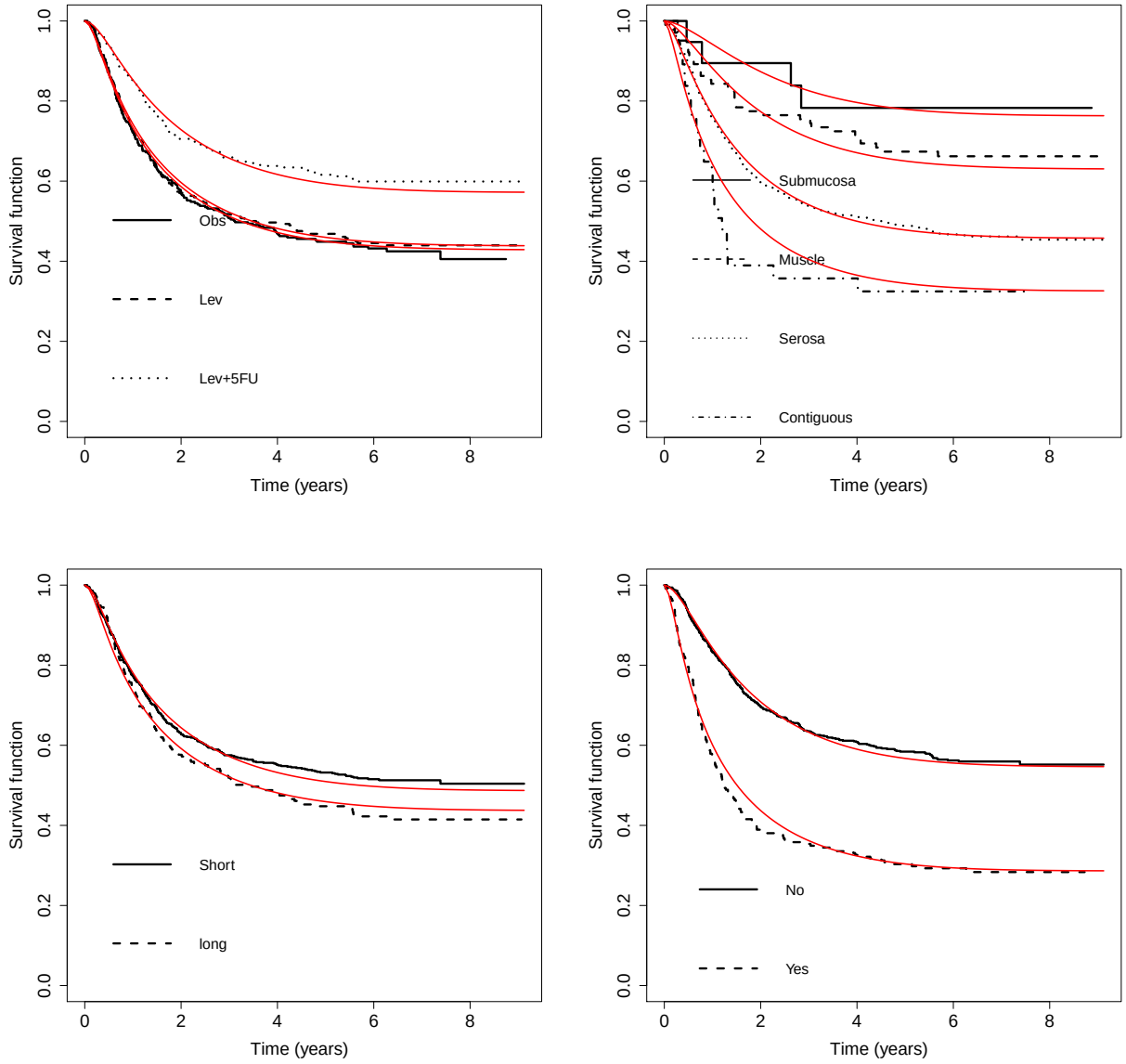

Figure 5: Kaplan-Meier curves stratified by explanatory variable and posterior survival functions to the colorectal carcinoma data, (left panel): extent of local spread and (right panel): more than 4 positive lymph nodes.

FU treatment.

Table 7: Bayes estimate cured proportions and 95% HPD intervals for eight hypothetical patients with colorectal cancer for end follow-up and after four years of the follow-up period.

Patient	Treatment*	Local Spread	Time-surgery	nodes > 4	p_0		$\pi(4)$	
					Mean	HPD(95%) Interval	Mean	HPD(95%) Interval
A	L+5-FU	submucosa	long	no	0.774	(0.566,0.932)	0.956	(0.917,0.986)
B	L+5-FU	muscle	long	no	0.739	(0.636,0.831)	0.949	(0.919,0.972)
C	L+5-FU	serosa	long	no	0.580	(0.508,0.656)	0.925	(0.892,0.951)
D	L+5-FU	contiguous	long	no	0.423	(0.280,0.573)	0.901	(0.852,0.938)
E	L	submucosa	long	no	0.653	(0.425,0.868)	0.936	(0.890,0.975)
F	L	muscle	long	no	0.606	(0.491,0.718)	0.929	(0.893,0.957)
G	L	serosa	long	no	0.437	(0.366,0.508)	0.903	(0.863,0.935)
H	L	contiguous	long	no	0.294	(0.167,0.427)	0.876	(0.812,0.922)

* L: Levamisole.

7 Concluding remarks

In this paper, we define a class of survival models for modeling time-to-event with long-term and obtain some of its structural properties. The class is constructed from a discrete frailty model based on a mixed Poisson distribution, where the distribution of the mixture is the *power variance function* (PVF) family. The model extends the *Bounded Cumulative Hazard* (BCH) model introduced by Yakovlev & Tsodikov (1996) and includes as special cases some models defined by Cancho *et al.* (2011a) and Barriga *et al.* (2018). Moreover, we extend the model in order to evaluate the effects of covariates in the cure fraction. We develop Bayesian inference based on Markov Chain Monte Carlo (MCMC) methods. Specifically, we implement a modified version of the collapsed Gibbs technique Liu (1994) to sample from the posterior distribution. This development will lead to an efficient Gibbs sampling procedure, which would otherwise be extremely difficult. We characterize the propriety of the joint posterior distribution of the parameters using non informative improper priors. We perform a simulation study to examine the parameter estimates. We prove empirically the usefulness of the BCH-PVF model by means of an application with colorectal cancer data. In a future work, we intend to extend this model to a multivariate survival model for jointly modeling any type of failure time data with a surviving

fraction.

Acknowledgment: We gratefully acknowledge financial support from CNPq, CAPES and FAPESP and Brazil.

Appendix A

To prove Theorem 4.1 we have to show that

$$\int L(\boldsymbol{\beta}, \boldsymbol{\phi}, \alpha, \gamma | \boldsymbol{\mathcal{D}}_{\text{obs}}) \pi(\phi_1 | a_0, b_0) \pi(\phi_2) \pi(\alpha) \pi(\gamma) d\boldsymbol{\vartheta} < \infty, \tag{35}$$

for which is suffices to verify that

$$\int \int L(\boldsymbol{\vartheta} | w, \boldsymbol{\mathcal{D}}_{\text{obs}}) \left[\prod_{i=1}^{n} g(w_i | \alpha, \gamma) \right] \pi(\phi_1 | a_0, b_0) \, \pi(\phi_2) \, \pi(\alpha) \, \pi(\gamma) \, d\boldsymbol{\vartheta} \, dw < \infty,$$

where $\boldsymbol{\vartheta}' = (\alpha, \gamma, \boldsymbol{\phi}, \boldsymbol{\beta})$, $w' = (w_1, \cdots, w_n)$ and $L(\boldsymbol{\vartheta} | w, \boldsymbol{\mathcal{D}}_{\text{obs}})$ is given by (21) with θ being replaced by $\theta_i = \exp(\boldsymbol{x}_i' \boldsymbol{\beta})$.

For doing this, we use each step of the proof presented by Chen *et al.* (2002). First, we prove that

$$[w_i \, \theta_i \, f_0(y_i | \boldsymbol{\phi})]^{\delta_i} \exp\{-w_i \, \theta_i [1 - S_0(y_i | \boldsymbol{\phi})]\} \leq M \phi_1^{\delta_i},$$

where $M \geq 1$ is a constant.

For $\delta_i = 0$, Equation (7) is immediate, since $\exp\{-w_i \theta_i [1 - S_0(y_i | \phi)]\} \leq 1$. For $\delta_i = 1$, the left side of (7) can be rewritten as

$$y_i^{-1} \frac{\phi_1 \, y_i^{\phi_1} \mathrm{e}^{\phi_2} \exp(-\mathrm{e}^{\phi_2} y_i^{\phi_1})}{1 - \exp(-\mathrm{e}^{\phi_2} y_i^{\phi_1})} \times \{[1 - S_0(y_i | \boldsymbol{\phi})] w_i \, \theta_i \exp[-w_i \, \theta_i (1 - S_0(y_i | \boldsymbol{\phi}))]\}. \tag{36}$$

Next, we define

$$\tau_1(z) = \frac{z \, \mathrm{e}^{-z}}{1 - e^{-z}} \qquad \text{and} \qquad \tau_2(z) = z \, \mathrm{e}^{-z}, \quad \text{for} \quad z > 0.$$

As pointed out by Chen *et al.* (2002), it can be shown that exist a common constant $1 \leq \tau_0 < \infty$ satisfying

$$\tau_1(z) \leq \tau_0 \quad \text{and} \quad \tau_2(z) \leq \tau_0, \quad \text{for all } z > 0. \tag{37}$$

By expression (37), we have that Equation (36) is less than or equal to $y_i^{-1}\phi_1\tau_0^2$. Therefore, setting $M^* = \tau_0^2 \max_{i:\delta_i=1}\{y_i^{-1}\}$ and $M = \max\{1, M^*\}$, we obtain (7).

Since $\mathbf{X}^*$ has full rank, we can assert the existence of p linearly independent row vectors $\boldsymbol{x}'_{i_1}, \cdots, \boldsymbol{x}'_{i_p}$ such that $\delta_{i_1} = \cdots = \delta_{i_p} = 1$. Then, the left side of (7) is less than or equal to

$$\int_0^\infty \int_0^1 \int_{\mathbb{R}^{+n}} \int_{-\infty}^\infty \int_0^\infty M^{n-p}\phi_1^{d-p} \int_{\mathbb{R}^p} \prod_{j=1}^p f_0\left(y_{i_j}|\boldsymbol{\phi}\right) w_{i_j} \exp\left\{\boldsymbol{x}'_{i_j}\boldsymbol{\beta} - \left[1 - S_0\left(y_{i_j}|\boldsymbol{\phi}\right)\right] w_{i_j} \exp\left(\boldsymbol{x}'_{i_j}\boldsymbol{\beta}\right)\right\}$$
$$\times \left[\prod_{i=1}^n g(w_i|\alpha,\gamma)\right] \pi\left(\phi_1|a_0,b_0\right)\pi\left(\phi_2\right)\pi\left(\alpha\right)\pi\left(\gamma\right) d\boldsymbol{\beta}\, d\phi_1\, d\phi_2\, dw\, d\alpha\, d\gamma.$$

Setting the transformation $u_j = \boldsymbol{x}'_{i_j}\boldsymbol{\beta} + \log(w_{i_j})$ for $j = 1, \ldots, p$ (a one-to-one linear transformation from $\boldsymbol{\beta}$ to $\boldsymbol{u}' = (u_1, \cdots, u_p)$), the previous expression is proportional to

$$\int_0^\infty \int_0^1 \int_{\mathbb{R}^{+n}} \int_{-\infty}^\infty \int_0^\infty \left\{\phi_1^{d-p} \int_{\mathbb{R}^p} \prod_{j=1}^p f_0\left(y_{i_j}|\boldsymbol{\phi}\right) \exp\left\{u_j - \left(1 - S_0\left(y_{i_j}|\boldsymbol{\phi}\right)\right)\exp\left(u_j\right)\right\} du\right\} \left[\prod_{i=1}^n g(w_i|\alpha,\gamma)\right]$$
$$\times \pi\left(\phi_1|a_0,b_0\right)\pi\left(\phi_2\right)\pi\left(\alpha\right)\pi\left(\gamma\right) d\phi_1\, d\phi_2\, d\boldsymbol{\beta}\, dw\, d\alpha\, d\gamma.$$

Integration out $\boldsymbol{u}$ leads to

$$\int_0^\infty \int_0^1 \int_{\mathbb{R}^{+n}} \int_{-\infty}^\infty \int_0^\infty \left\{\phi_1^{d-p} \prod_{j=1}^p \frac{f_0\left(y_i|\boldsymbol{\phi}\right)}{1 - S_0\left(y_i|\boldsymbol{\phi}\right)}\right\} \left[\prod_{i=1}^n g(w_i|\alpha,\gamma)\right] \pi\left(\phi_1|a_0,b_0\right)\pi\left(\phi_2\right)\pi\left(\gamma\right) d\phi_1\, d\phi_2\, dw\, d\alpha\, d\gamma.$$

Using (37) in the last expression gives

$$\frac{f_0\left(y_i|\boldsymbol{\phi}\right)}{1 - S_0\left(y_i|\boldsymbol{\phi}\right)} \leq M\phi_1.$$

Therefore, by conditions (ii) and (iii) of Theorem 4.1 and the fact that $g\left(w_i|\alpha,\gamma\right)$ is a proper density and $\pi(\gamma)$ is a proper prior, Equation (7) is less than or equal to

$$M^p \left[\int_0^\infty \phi_1^d \pi\left(\phi_1|a_0,b_0\right) d\phi_1 \int_{-\infty}^\infty \pi\left(\phi_2\right) d\phi_2\right] \int_0^\infty \int_0^1 \left[\prod_{i=1}^n \int_0^\infty g\left(w_i|\alpha,\gamma\right) dw_i\right] \pi\left(\alpha\right)\pi\left(\gamma\right) d\alpha\, d\gamma < \infty.$$

This fact completes the proof of theorem.

Appendix B

The trace plots for the parameters of the BCH-PVF model for the colorectal data are displayed in Figure 6.

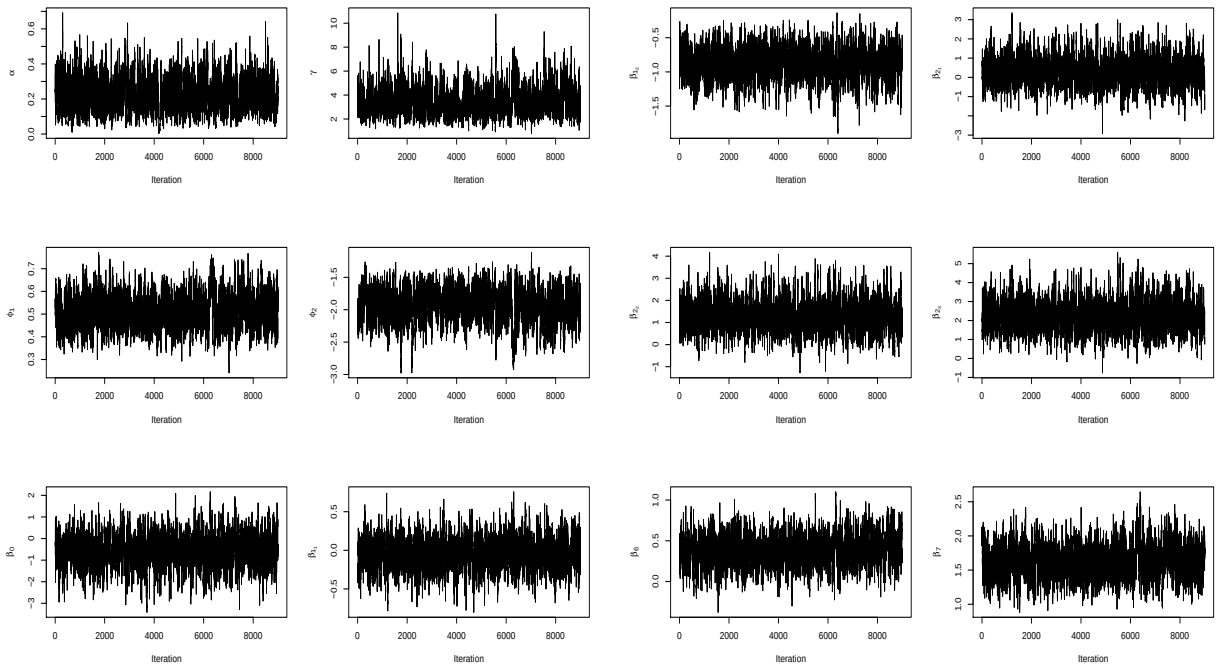

Figure 6: Trace plots for the parameters of the BCH-PVF model.

References

Aalen, O., Borgan, O. & Gjessing, H. (2008). *Survival and Event History Analysis*. Springer, New York.

Aalen, O. O. (1992). Modelling heterogeneity in survival analysis by the compound poisson distribution. *The Annals of Applied Probability*, pages 951–972.

Barriga, G. D., Cancho, V. G., Garibay, D. V., Cordeiro, G. M. & Ortega, E. M. (2018). A new survival model with surviving fraction: An application to colorectal cancer data. *Statistical Methods in Medical Research*, **28**(9), 1–15.

Berkson, J. & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**(2), 501–515.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society Series B (Methodological)*, **11**(1), 15–53.

Cancho, V., Dey, D., Lachos, V. & Andrade, M. (2011a). Bayesian nonlinear regression models with scale

mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Computational Statistics & Data Analysis*, **55**(1), 588–602.

Cancho, V. G., Rodrigues, J. & de Castro, M. (2011b). A flexible model for survival data with a cure rate: a bayesian approach. *Journal of Applied Statistics*, **38**(1), 57–70.

Cancho, V. G., Zavaleta, K., Macera, M., Suzuki, A. K. & Louzada, F. (2018). A bayesian cure rate model with dispersion induced by discrete frailty. *Communications for Statistical Applications and Methods*, **25**(1), 471–488.

Caroni, C., Crowder, M. & Kimber, A. (2010). Proportional hazards models with discrete frailty. *Lifetime Data Analysis*, **16**(3), 374–384.

Chen, M.-H., Ibrahim, J. G. & Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**(447), 909–919.

Chen, M.-H., Ibrahim, J. G. & Sinha, D. (2002). Bayesian inference for multivariate survival data with a cure fraction. *Journal of Multivariate Analysis*, **80**(1), 101–126.

Cowles, M. K. & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, **91**(434), 883–904.

Cox, D. R. & Oakes, D. (1984). *Analysis of survival data*. Chapman and Hall, London.

De Angelis, R., Capocaccia, R., Hakulinen, T., Soderman, B. & Verdecchia, A. (1999). Mixture models for cancer survival analysis: Application to population-based data with covariates. *Statistics in Medicine*, **18**(4), 441–454.

de Souza, D., Cancho, V. G., Rodrigues, J. & Balakrishnan, N. (2017). Bayesian cure rate models induced by frailty in survival analysis. *Statistical Methods in Medical Research*, **26**(5), 2011–2028.

Dunn, P. K. & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**(3), 236–244.

Dunson, D. B. & Zhou, H. (2000). A Bayesian model for fecundability and sterility. *Journal of the American Statistical Association*, **95**(452), 1054–1062.

Economou, P. & Stehlik, M. (2015). On small samples testing for frailty through homogeneity test. *Communications in Statistics- Simulation and Computation*, **44**(1), 40–65.

Gelfand, A. E., Dey, D. K. & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian statistics: proceedings of the Fourth Valencia International Meeting, April 15-20, 1991*, volume 4, pages 147–167. Oxford University Press, USA.

Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, **71**(1), 75–83.

Hougaard, P. (1986a). A class of multivanate failure time distributions. *Biometrika*, **73**(3), 671–678.

Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**(2), 387–396.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, **47**(4), 939–956.

Leão, J., Leiva, V., Saulo, H. & Tomazella, V. (2017). Birnbaum-saunders frailty regression models: Diagnostics and application to medical data. *Biometrical Journal*, **59**(2), 291–314.

Liu, J. S. (1994). The collapsed gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, **89**(427), 958–966.

Mazroui, Y., Mathoulin-Pelissier, S., MacGrogan, G., Brouste, V. & Rondeau, G. (2013). Multivariate frailty models for two types of recurrent events with a dependent terminal event: application to breast cancer data. *Biometrical Journal*, **55**(6), 866–884.

Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H. et al. (1990). Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, **322**(6), 352–358.

Rigby, R. A. & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*, **54**(3), 507–554.

Rodrigues, J., de Castro, M., Cancho, V. G. & Balakrishnan, N. (2009). COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, **139**(10), 3605–3611.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **64**(4), 583–639.

Tsodikov, A. D., Ibrahim, J. G. & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*, **98**(464), 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*, page 579604.

Vaupel, J. W., Manton, K. G. & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**(3), 439–454.

Wienke, A. (2011). *Frailty models in survival analysis*. Chapman and Hall, London.

Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, New Jersey.