Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–16

DOI: 10.1177/0962280218786053
journals.sagepub.com/home/smm
SAGE

A new survival model with surviving fraction: An application to colorectal cancer data

Gladys DC Barriga,[1] Vicente G Cancho,[2] Daniel V Garibay,[3] Gauss M Cordeiro[4] and Edwin MM Ortega[5]

Abstract

We propose a new survival model for lifetime data in the presence of surviving fraction and obtain some of its properties. Its genesis is based on extensions of the promotion time cure model, where an extra parameter controls the heterogeneity or dependence of an unobserved number of lifetimes. We construct a regression model to evaluate the effects of covariates in the cured fraction. We discuss inference aspects for the proposed model in a classical approach, where some maximum likelihood tools are explored. Further, an expectation maximization algorithm is developed to calculate the maximum likelihood estimates of the model parameters. We also perform an empirical study of the likelihood ratio test in order to compare the promotion time cure and the proposed models. We illustrate the usefulness of the new model by means of a colorectal cancer data set.

Keywords

Colorectal cancer, cured fraction, cure rate model, EM algorithm, likelihood function, survival model

1 Introduction

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 estimates from the National Cancer Institute (NCI) for 2018 are 140,000 new cases in the USA, 8% of the total number of new cancer cases for this year. NCI also reported that the estimate number of deaths for 2018 is about 66,000 (8.4% of all cancer deaths) corresponding to 22 deaths per 100,000 people per year.

The improvement in technology to detect the cancer, the advance in techniques of treatment, and the decrease of exposure to risk factors (such as smoking and rued meat consumption), all contributed for a reduction in the mortality by colorectal cancer from seven decades ago when it was the first cause of cancer death in the USA.[1] The ideal treatment for this cancer is the full resection of the tumor surgically and subsequent follow-up of the patient for the possibility of recurrence of the disease. In certain stages of the cancer, the radiotherapy is recommended after the surgery to reduce the chance of recurrence. Additionally, the treatment by chemotherapy with the 5-fluorouracil (5-FU) drug as an adjuvant to the surgery leads to a reduction in the recurrence rates and an increase in survival rates, since it is directed to eradicate the undetected sites of metastasis.[2]

The staging of the colorectal cancer is determined by the number of layers of the intestine wall that the tumor invades and the presence of metastasis. Dukeýs classification[3] describes: Stage A, the tumor is limited to the bowel mucosa cells and the muscularis mucosa layer; Stage B, the tumor invades all intestine layers; Stage C, the tumor

[1]Departamento de Engenharia de Produção, FEB, Universidade Estadual Paulista, Bauru, Brazil
[2]Departamento de Matemática Aplicada e Estatística, ICMC, Universidade de São Paulo, São Carlos, Brazil
[3]Falcudade de Medicina de Riberão Preto, Universidade de São Paulo, Riberão Preto, Brazil
[4]Departamento de Estatística, Universidade Federal de Pernambuco, PE, Brazil
[5]Departamento de Ciências Exatas, ESALQ, Universidade de São Paulo, Piracicaba, Brazil

Corresponding author:
Edwin MM Ortega, Department of Exact Sciences, University of São Paulo, Piracicaba 13418-900, SP, Brazil.
Email: edwin@usp.br

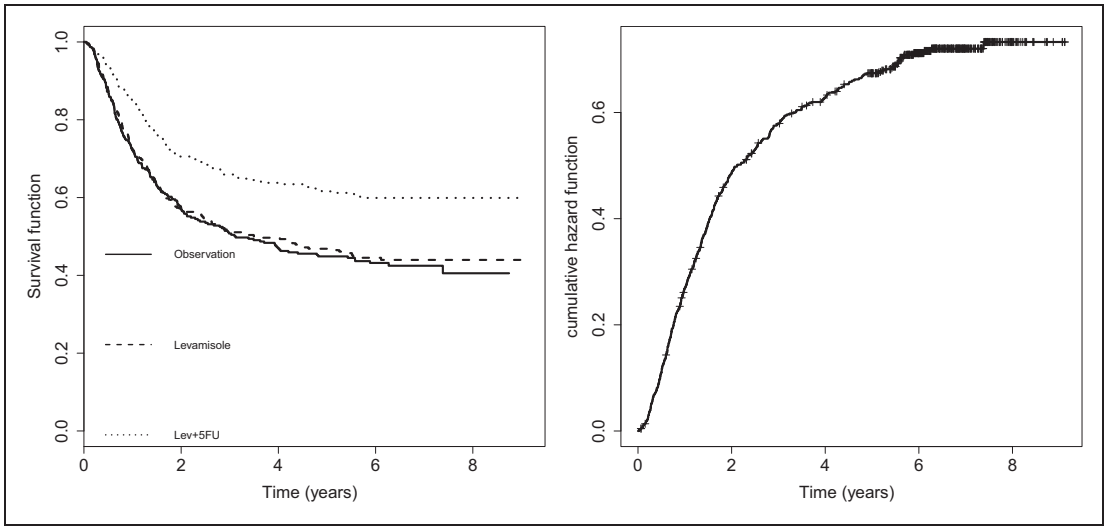

Figure 1. The Kaplan–Meier estimate of survival functions stratified by treatment (left panel) and cumulative hazard function (right panel) for recurrence time of patients treated with colorectal cancer.

metastasizes to the regional lymph nodes; and Stage D, the metastasis implants at the peritoneum, omentum and other distant regions.

An anthelmintic drug, *Levamisole*, has a presumed immunomodulatory activity, for that it was studied as a combination with 5-FU in the treatment of colorectal cancer. An experimental study by Moertel et al.[3] presented a comparison between the efficiency of treatment by 5-FU along Levamisole in randomly selected patients with Stage C of colorectal cancer and the efficiency by Levamisole alone, where 929 patients were classified with Stage C for which 50% of the patients experience disease recurrence. The Kaplan–Meier estimates of the survival function to a data set on the colorectal cancer recurrence stratified by treatment are plotted in Figure 1 (left panel) and the cumulative hazard functions (right panel), where the estimated survival functions tend to positive levels below to an upper horizonal limit. This fact, according to Yakovlev and Tsodikov,[4] may be thought of as an indication of the presence of a proportion of patients disease-free, and these patients can be considered cured. This behavior indicates that, at least in principle, models that ignore the possibility of cure will not be suitable for these data.

Models for survival data with a surviving fraction (also known as cure rate models or long-term survival models) have deserved a great deal of interest in the literature under the headings of reliability and survival analysis. Cure rate models cover the situations where there are sampling units insusceptible to the occurrence of the event of interest. The proportion of such units is termed as the cured fraction. In clinical studies, the event of interest may be the death of a patient or a tumor recurrence (which can be attributed to metastasis-component tumor cells left active after an initial treatment). Two formulations of cure rate models stand out in the literature as prevailing approaches to be developed. Here, we point out a distinguishing feature between them. In the standard mixture cure model,[5,6] the number of causes of the event of interest is a binary random variable on $\{0, 1\}$, whereas in the promotion time cure model[4] this number follows a Poisson distribution. Several other works have contributed to this area such as the studies presented in Tsodikov et al.,[7] Ibrahim et al.,[8] Rodrigues et al.,[9] Ortega et al.[10] and Cordeiro et al.[11]

In this paper,

- We propose a new survival model with cured fraction for modeling time of recurrence disease of patients with colorectal cancer.
- Also, we introduce a new model with a random component (that can measure a possible heterogeneity, dispersion or correlation in the number of carcinogenic cells) that extends the promotion time model discussed by several authors in the literature.
- In addition, we construct a regression model to evaluate the effects of covariates on the disease-free cured rate.

The rest of the paper proceeds as follows. In Section 2, we formulate the model. Some of its structural properties are investigated in Section 3. Inference based on maximum likelihood is developed in Section 4.

A simulation study is presented in Section 5 in order to study some finite sample properties of the estimates. In Section 6, our methodology is illustrated on a real dataset. Finally, Section 7 presents some concluding remarks.

2 The model

For an individual in the population, let N be the number of carcinogenic cells that would cause the detectable tumor. We consider that N has a Poisson distribution with mean $\eta\theta$, where $\theta > 0$ is a constant and $\eta > 0$ is a random parameter having an inverse Gaussian distribution with unity mean and variance $\gamma > 0$, which takes into account an unobserved heterogeneity or dependence from individual to individual. Thus, the probability generating function (pgf) of the random variable N given the random parameter (η) is $G_{N|\eta}(w) = \exp[-\eta\theta(1-w)]$. Then, integrating in η, we obtain the pgf of N as

$$G_N(w) = \exp\left\{\gamma^{-1}\left(1 - \sqrt{1 + 2\gamma\theta w}\right)\right\}, \quad |w| \leq 1 \tag{1}$$

We can demonstrate from equation (1) that the mean and variance of N are given by $E(N) = \theta$ and $Var(N) = \theta + \gamma\theta^2$, respectively. Hence, γ is a kind of dispersion parameter that can measure a possible heterogeneity, dispersion or correlation in the number of carcinogenic cells.

Given $N = n$, let Z_j $(j = 1, \ldots, n)$ be the time until the jth clonogenic cell produces a detectable tumor in the individual. We assume that the Z_j' s are conditionally independent of N with common cumulative distribution function (cdf) $F(t) = 1 - S(t)$. Here, we emphasize that the random variables Z_j' s and N are unobserved. Thus, the period of time a growing tumor cell would need to proliferate before being detectable for an individual can be defined by the random variable $T = \min(Z_1, \ldots, Z_N)$ with $P(T > t|N = 0) = 1$. Under these assumptions, the population survival function[7,9] reduces to

$$S_p(t) = \exp\left\{\gamma^{-1}\left(1 - \sqrt{1 + 2\gamma\theta F(t)}\right)\right\} \tag{2}$$

Thus, $\lim_{t\to\infty} S_p(t) = P(N = 0) = p_0$, which implies that $N = 0$ corresponds to the "cured" population.

From equation (2), we can write

$$p_0 = \exp\left\{\gamma^{-1}\left(1 - \sqrt{1 + 2\gamma\theta}\right)\right\} > 0$$

which implies that equation (2) is not a proper survival function. The cured fraction tends to zero when $\theta \to \infty$, whereas the cured fraction tends to one when $\theta \to 0$. Hereafter, the model (2) is called the *promotion time cure rate model with dispersion*.

Consider the function

$$\begin{aligned} \pi(t) &= Pr(N = 0|T > t) \\ &= \exp\left\{\gamma^{-1}\left(\sqrt{1 + 2\gamma\theta F(t)} - \sqrt{1 + 2\gamma\theta}\right)\right\} \end{aligned} \tag{3}$$

denoting the probability that an individual is immune or cured of the disease given that survives for time $t > 0$ after treatment. The probability (3) 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 $\pi(0) = p_0$. The probability tends to one when $t \to \infty$ (corresponding to certainly of immunity if the individual's lifetime is very long).

The density function associated with equation (2) is given by

$$f_p(t) = \frac{\theta f(t) S_p(t)}{\sqrt{1 + 2\gamma\theta F(t)}}$$

where $f(t) = -dS(t)/dt$. The associated hazard rate function (hrf) reduces to

$$h_p(t) = \frac{\theta f(t)}{\sqrt{1 + 2\gamma\theta F(t)}}$$

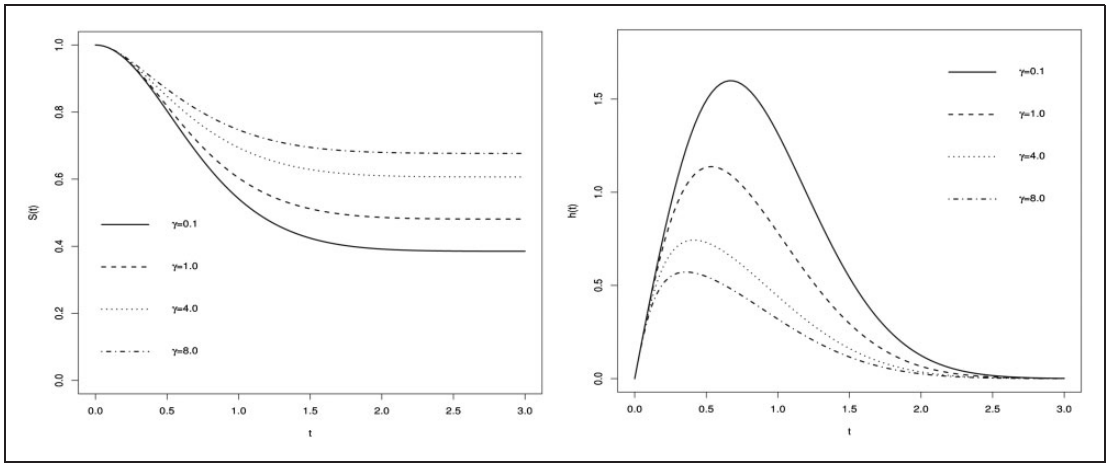

Figure 2. Survival functions (left panel) and hazard functions (right panel) with common survival function $S(t) = \exp(-t^2)$ and $\theta = 2$.

We note that $h_p(t) \to 0$ at a fast rate when $t \to \infty$ and $\int_0^\infty h_p(t)dt < \theta < \infty$. The plots in Figure 2 reveal the flexibility of the survival and hazard functions in terms of the additional dispersion parameter γ introduced in the model. Furthermore, in the special case when $\gamma \to 0$, the model is identical to the promotion time cure model investigated by Yakovlev and Tsodikov.[4]

The (proper) survival function for the individuals under risk, denoted by $S_R(t)$, is obtained as $S_R(t) = P(T > t \mid N \geq 1)$ (for $t > 0$) and then it reduces to

$$S_R(t) = \frac{\exp\left\{\gamma^{-1}\left(1 - \sqrt{1 + 2\gamma\theta F(t)}\right)\right\} - p_0}{(1 - p_0)} \tag{4}$$

We have $S_R(0) = 1$ and $S_R(\infty) = 0$, which implies that it is a proper survival function. The density function for the individuals under risk in the population (a proper density function) is given by

$$f_R(t) = \frac{f_p(t)}{(1 - p_0)}, \ t > 0 \tag{5}$$

and the hrf for the individuals under risk in the population becomes

$$h_R(t) = \left[\frac{S_p(t)}{S_p(t) - p_0}\right] h_p(t), \ t > 0 \tag{6}$$

Thus, equation (6) has a multiplier factor $\frac{S_p(t)}{S_p(t) - p_0} > 1$ compared to the hazard function $h_p(t)$ of the entire population. Also, it can be proved that $h_R(t) \to \frac{f(t)}{S(t)}$ when $t \to \infty$ and hence $h_R(t)$ converges to the hrf of the latent random variable Z.

The relationship between the model (2) and the mixture cure rate model[5,6] is given by

$$S_p(t) = p_0 + (1 - p_0) S_R(t)$$

where $S_R(t)$ is given by equation (4). Thus, $S_p(t)$ is a mixture cure rate model with cured fraction equal to p_0 and survival function $S_R(t)$ for individuals under risk in the population. This result implies that every mixture cure rate model corresponds to some model of the form (2) for some γ, θ and $S_R(\cdot)$.

There is a relationship between the cure rate model (2) and the frailty model. In fact, the frailty model assumes a proportional hazards structure conditional on the random effect (frailty). The hazard function of an individual depends on an unobservable time Z independent of the random variable N, and acts multiplicatively on the baseline hrf, $h_0(t)$, i.e. $h(t|N) = Nh_0(t)$. If N is a discrete random variable with support $\{0, 1, \ldots, \}$ and pgf

$G_N(\cdot)$, then the unconditional survival function is given by

$$S(t) = G_N(S_0(t)) \tag{7}$$

where $S_0(t) = \exp\{-\int_0^t h_0(u)\mathrm{d}u\}$ is a baseline survival function. If N is a random variable with pgf (1), the model (7) is identical to the proposed model (2). This relationship is useful in practice because it allows the choice of the baseline distribution family from the observed data.

Proposition. The model defined by the survival function (2) is identifiable.

Proof. Let $\vartheta_{\mathbf{1}} = (\gamma_1, \theta_1, \boldsymbol{\varphi}_{\mathbf{1}})$ and $\vartheta_{\mathbf{2}} = (\gamma_2, \theta_2, \boldsymbol{\varphi}_{\mathbf{2}})$ such that $\vartheta_1 \neq \vartheta_{\mathbf{2}}$, where $\boldsymbol{\varphi}_{\mathbf{1}}$ and $\boldsymbol{\varphi}_{\mathbf{2}}$ are the parameter vectors in $S(\cdot)$. Suppose that $S_{\mathrm{p}}(t; \vartheta_{\mathbf{1}}) = S_{\mathrm{p}}(t; \vartheta_{\mathbf{2}})$, for all $t > 0$, which from equation (2) implies that

$$\frac{\gamma_2}{\gamma_1} = \frac{1 - \sqrt{1 + 2\gamma_2\theta_2 F(t; \boldsymbol{\varphi}_2)}}{1 - \sqrt{1 + 2\gamma_1\theta_1 F(t; \boldsymbol{\varphi}_1)}}, \quad t > 0 \tag{8}$$

We know that $1 - \sqrt{1 + 2\theta\gamma F(t : \boldsymbol{\varphi})}$ is monotone decreasing in γ, θ and $\boldsymbol{\varphi}$. Without loss of generality, we assume that $\gamma_1 < \gamma_2$. Then, $\sqrt{1 + 2\gamma_2\theta_2 F(t; \boldsymbol{\varphi}_2)} < \sqrt{1 + 2\gamma_1\theta_1 F(t; \boldsymbol{\varphi}_1)}$. For $\theta_1 \neq \theta_2$ and $\boldsymbol{\varphi}_1 \neq \boldsymbol{\varphi}_2$, there exists a t_0 such that $\sqrt{1 + 2\gamma_2\theta_2 F(t_0; \boldsymbol{\varphi}_2)} > \sqrt{1 + 2\gamma_1\theta_1 F(t_0; \boldsymbol{\varphi}_1)}$. Therefore, the equality (8) cannot be satisfied, which completes the proof.

3 Properties

In this section, we obtain some mathematical properties of the random variable T_R representing the non-cured population, i.e. the individuals at risk at time t having the density function $f_R(t)$. The quantile function (qf) of T_R is determined by inverting $F_{\mathrm{R}}(t) = 1 - S_{\mathrm{R}}(t) = u$ in equation (4). The qf of T_R can be expressed in terms of the qf corresponding to $F(t)$ (see Section 2) defined by $Q_Z(u) = F^{-1}(u)$. For a given probability u, the percentage 100 $\%u$ of individuals of the population will be at risk at the time $t_u = Q_{\mathrm{R}}(u)$. Then, we have

$$t_u = F^{-1}\left(\{[1 - \gamma\log\{v(u)\}]^2 - 1\}/[2\gamma\theta]\right) \tag{9}$$

where $v(u) = 1 - (1 - p_0)u$.

By combining (1) and (2), we can write $S_p(t) = G_N(F(t)) = \sum_{n=0}^{\infty}\delta_n F(t)^n$, which holds for the hypothesized cdf of any clonogenic cell capable of producing a detectable tumor, where $\delta_n = \partial^n G_N(w)/\partial w^n|_{w=0}$. From $f_R(t) = f_p(t)/(1 - p_0)$, for $t > 0$, we can obtain the linear representation by simple differentiation of equation (4)

$$f_{\mathrm{R}}(t) = \sum_{k=0}^{\infty}\rho_k\pi_{k+1}(t) \tag{10}$$

where $\pi_{k+1}(t) = (k+1)F(t)^k f(t)$ denotes the *exponentiated-F* (exp-F) density function with power parameter $k+1$ of the random variable W_{k+1}, say $W_{k+1} \sim \text{exp-F}(k+1)$ (for $k \geq 0$). Here, the coefficients ρ_k's are easily determined in **Mathematica** as $\rho_0 = \theta/(1 - p_0)$, $\rho_1 = -2(1+\gamma)\theta^2/(1 - p_0)$, $\rho_2 = 3(3\gamma^2 + 3\gamma + 1)\theta^3/(1 - p_0)$, $\rho_3 = -4(15\gamma^3 + 15\gamma^2 + 6\gamma + 1)\theta^4/(1 - p_0)$, etc.

Equation (10) reveals that the density function for the non-cured population (T_R) is given by a linear combination of exp-F densities, whose coefficients depend only on γ and θ. So, several mathematical quantities (such as ordinary and incomplete moments, mean deviations, and generating function) of T_R can be determined from this equation by knowing those properties of the exp-F distribution. Several properties of the exponentiated distributions have been studied by many authors in recent years. See, for example, Mudholkar et al.,[12] Gupta and Kundu[13] and Nadarajah and Kotz,[14] among others.

3.1 Moments and applications

The nth moment of T_R, say $E(T_{\mathrm{R}}^n)$, can be expressed from equation (10) as

$$\mu'_n = \sum_{k=0}^{\infty}\rho_k E(W_{k+1}^n) = \sum_{k=0}^{\infty}(k+1)\rho_k\tau_{n,k} \tag{11}$$

where $\tau_{n,k}$ is given in terms of $Q_Z(u) = F^{-1}(u)$, namely

$$\tau_{n,k} = \int_0^\infty t^n F(t)^k f(t)\mathrm{d}t = \int_0^1 Q_Z(u)^n u^k \mathrm{d}u$$

Expressions for moments of several exponentiated distributions given by Nadarajah and Kotz (2006) can be used in equation (11) to produce $E(T_{\mathrm{R}}^n)$. As a simple example, if $F(t) = 1 - \exp[-(\lambda t)^\beta]$ $(t > 0)$ is the Weibull cdf with scale parameter $\lambda > 0$ and shape parameter $\beta > 0$, we can obtain the moments of W_{k+1} using the generalized binomial expansion. Hence, μ'_n follows from equation (11) as

$$\mu'_n = \frac{\Gamma(n/\beta + 1)}{\lambda^n} \sum_{k=0}^{\infty} (k+1)\rho_k \sum_{j=0}^{k} \frac{(-1)^j}{(j+1)^{n/\beta+1}} \binom{k}{j}$$

Further, the cumulants (κ_r) of T_R (for $r \geq 1$) can be determined from the ordinary moments as

$$\kappa_r = \mu'_r - \sum_{k=1}^{r-1} \binom{r-1}{k-1} \kappa_k \mu'_{r-k}$$

where $\kappa_1 = \mu'_1$. The skewness $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and kurtosis $\gamma_2 = \kappa_4/\kappa_2^2$ of T_R are obtained from the second, third and fourth cumulants.

For empirical purposes, the shape of many distributions can be usefully described by what we call the incomplete moments. These types of moments play an important role for measuring inequality, for example, income quantiles and Lorenz and Bonferroni curves, which depend upon the incomplete moments of a distribution. The nth incomplete moment of T_R, say $m_n(t) = E\big(T_{\mathrm{R}}^n | 0 < T_{\mathrm{R}} < t\big)$, can be expressed as

$$m_n t) = \sum_{k=0}^{\infty} (k+1)\rho_k \int_0^{F(t)} Q_Z(u)^n u^k \mathrm{d}u \tag{12}$$

where the integral can be computed for most selected distributions for F.

A general equation for the first incomplete moment $m_1(t)$ of T_{R} can be derived from equations (10) and (12) as

$$m_1(t) = \sum_{k=0}^{\infty} \rho_k J_k(t) = \sum_{k=0}^{\infty} (k+1)\rho_k \int_0^{F(t)} Q_Z(u) u^k \mathrm{d}u \tag{13}$$

where $J_k(t) = \int_0^t w h_{k+1}(w)\mathrm{d}w$.

A simple application of equation (13) is now addressed for the Weibull cdf of T discussed before. We can obtain after some algebra

$$m_1(t) = \sum_{k,r=0}^{\infty} \frac{(-1)^r (k+1)\binom{k}{r}\rho_k}{(r+1)^{1+1/\beta}\beta} \gamma\big(1 + \beta^{-1}, (r+1)(\lambda t)^\beta\big)$$

where $\gamma(a, x) = \int_0^x w^{a-1} \mathrm{e}^{-w}\mathrm{d}w$ is the incomplete gamma function.

The first incomplete moment can be applied to obtain Bonferroni and Lorenz curves defined (for a given probability π) by $B(\pi) = m_1(q)/(\pi\mu'_1)$ and $L(\pi) = \pi B(\pi)$, respectively, where $q = Q_{\mathrm{R}}(\pi)$ is the qf of T_{R} at π.

The mean deviations about the mean $(\chi_1 = E(|T_{\mathrm{R}} - \mu'_1|))$ and about the median $(\chi_2 = E(|T_{\mathrm{R}} - M|))$ of T_{R} can be expressed as

$$\chi_1 = 2\mu'_1 F_R\big(\mu'_1\big) - 2m_1\big(\mu'_1\big) \quad \text{and} \quad \chi_2 = \mu'_1 - 2m_1(M) \tag{14}$$

respectively, where $\mu'_1 = E(T_{\mathrm{R}})$, $M = Q_R(0.5)$ is the median given by equation (9), $F_{\mathrm{R}}(\mu'_1) = 1 - S_{\mathrm{R}}(\mu'_1)$ is easily calculated from equation (4) and $m_1(\cdot)$ is given by equation (13).

Other kinds of moments can also be obtained in closed form, but we consider only the previous moments for reasons of space.

3.2 Generating function

Here, we provide a formula the moment generating function (mgf) $M_R(s) = E[\exp(sT_R)]$ of T_R. The function $M_R(s)$ can be obtained from equation (10) as

$$M_R(s) = \sum_{k=0}^{\infty} \rho_k M_{k+1}(t) = \sum_{k=0}^{\infty} (k+1)\rho_k \xi_k(s) \tag{15}$$

where $M_{k+1}(t)$ is the mgf of W_{k+1} and

$$\xi_k(s) = \int_0^{\infty} e^{st} F(t)^k f(t) dt = \int_0^1 \exp[sQ_Z(u)] u^k du$$

Hence, $M_R(s)$ can be immediately determined from the generating function of the exp-F distribution. As a single example, if T has exponential cdf with rate parameter λ, we obtain

$$M_R(s) = \sum_{k=0}^{\infty} (k+1) B(k+1, 1-\lambda s)\rho_k$$

where $B(\cdot,\cdot)$ is the complete beta function.

4 Inference

Consider the lifetimes $T_1, \ldots, T_n$ of n individuals. Suppose the lifetime is not completely observed and it is subject to right censoring, where C_i denotes the censoring time. We then observe $t_i = \min\{T_i, C_i\}$ and $\delta_i = I(T_i \leq C_i)$, where $\delta_i = 1$ if T_i is a lifetime and $\delta_i = 0$ if it is right censored, for $i = 1, \ldots, n$. We incorporate the covariates through θ. For each individual i, let $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip})^\top$ denote the vector of covariates, and let $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ denote the corresponding vector of regression coefficients. We relate θ to the covariates by $g(\theta_i) = \eta(\mathbf{x}_i; \beta) = \mathbf{x}_i^\top \beta$, $i = 1, \ldots, n$, where the link function $g(\cdot)$ is a monotonic twice differentiable function. The logarithmic link function given by $g(\theta_i) = \log(\theta_i)$ can be adopted for our model.

From n independent individuals, the log-likelihood function under non-informative censoring is given by

$$\begin{aligned} \ell(\vartheta) = {} & n\gamma^{-1} + \sum_{i=1}^{n} \delta_i \log(\theta_i) + \sum_{i=1}^{n} \delta_i \log[f(t_i; \boldsymbol{\varphi})] \\ & - \sum_{i=1}^{n} \delta_i \log\left[\sqrt{1 + 2\gamma\theta_i F(t_i; \boldsymbol{\varphi})}\right] \\ & - \gamma^{-1} \sum_{i=1}^{n} \sqrt{1 + 2\gamma\theta_i F(t_i; \boldsymbol{\varphi})} \end{aligned} \tag{16}$$

where $\vartheta = (\gamma, \boldsymbol{\varphi}^\top, \beta^\top)^\top$, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_q)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^\top$ and $\mathbf{x} = (x_1, \ldots, x_n)$. Since the baseline distribution is specified, the maximum likelihood estimator (MLE) of ϑ can be obtained by direct maximization of the log-likelihood function (16) using, for example, a optimization procedure such as the BFGS (Broyden–Fletcher–Goldfarb–Shanno) method. This optimization is based on the Newton–Raphson method.[15] We implement the EM algorithm[16] to calculate the MLE of ϑ. The EM algorithm is efficient (produces robust estimates), flexible (can be applied in several situations) and has stable convergence stronger than the Newton–Raphson method as discussed by McLachlan and Krishnan.[17]

The ith element of the set of censored observations can be derived from two different groups, individual under risk (susceptible) or cured. Suppose we define a latent variable Δ that indicates this event. Let Δ_i be the ith latent variable given by

$$\Delta_i = \begin{cases} 1, & \text{if susceptible,} \\ 0, & \text{if cured} \end{cases}$$

Then, the complete likelihood function is given by

$$\begin{aligned} L_c(\vartheta) = & \prod_{i\in\bar{C}}[1-p_0(\mathbf{x_i})]\prod_{i\in\bar{C}} f_R(t_i;\mathbf{x}_i,\vartheta) \\ & \times \prod_{i\in C}[p_0(\mathbf{x_i})]^{1-\Delta_i} \\ & \times \prod_{i\in C}[(1-p_0(\mathbf{x_i}))S_R(t_i;\mathbf{x}_i,\vartheta)]^{\Delta_i} \end{aligned}$$

where $S_R(\cdot)$ and $f_R(\cdot)$ are survival and density functions for the individual under risk given in equations (4) and (5), respectively

$$p_0(\mathbf{x_i}) = \exp\left\{\gamma^{-1}\left(1-\sqrt{1+2\gamma\exp\{\mathbf{x}_i^\top\}}\right)\right\}$$

and $C = \{i \in \{1,2,\ldots,n\} : \delta_i = 0\}$ and $\bar{C} = \{i \in \{1,2,\ldots,n\} : \delta_i = 1\}$ are the censored and uncensored sets of observations, respectively.

For the E-Step of the EM-algorithm, we compute the expectation of the complete data log-likelihood function with respect to the distribution of the unobserved Δ_i, given the current values of the parameters and the observed data $\boldsymbol{O}$. Note that the Δ_i's are Bernoulli random variables in the complete likelihood and we have to compute $\pi_i^{(k)} = E[\Delta_i|\vartheta^{(k)},\mathbf{O}]$, $i = 1\ldots,n$, where $\vartheta^{(k)}$ denotes the current parameter value at the kth iteration step. Next, for $i \in C$, we have

$$\begin{aligned} \pi_i^{(k)} &= E[\Delta_i|\vartheta^{*(k)},\mathbf{O}] = P[\Delta_i|T_i > y_i] \\ &= \frac{[1-p_0(\mathbf{x_i})]S_R(y_i;\mathbf{x}_i,\vartheta)}{S_p(y_i;\mathbf{x}_i,\vartheta)}|_{\vartheta=\vartheta^{(k)}} \end{aligned}$$

and the conditional expectation of the complete log-likelihood function is given by

$$\begin{aligned} Q(\vartheta,\pi^{(k)}) = & \sum_{i\in\bar{C}}\log[1-p_0(\mathbf{x_i})] + \sum_{i\in\bar{C}}\log f_R(y_i;\mathbf{x}_i,\vartheta) \\ & + \sum_{i\in C}(1-\pi_i^{(k)})\log[p_0(\mathbf{x_i})] \\ & + \sum_{i\in C}\pi_i^{(k)}\log[(1-p_0(\mathbf{x_i}))] \\ & + \sum_{i\in C}\pi_i^{(k)}\log[S_R(y_i;\mathbf{x}_i,\vartheta)] \end{aligned}$$

M-Step: In this step, we maximize the function $Q(\vartheta,\pi^{(k)})$ with respect to ϑ over the corresponding parameter space Θ, given $\pi^{(k)}$. So, we obtain an improved estimate of ϑ given by

$$\vartheta^{(k+1)} = \arg\max_{\vartheta\in\Theta} Q(\vartheta,\pi^{(k)})$$

The E-step and the M-step are then continued iteratively until convergence to find the MLE of the parameter ϑ^*. In this paper, as the MLEs of β, γ and $\boldsymbol{\varphi}$ do not have explicit expressions, the maximization step is carried out using the EM gradient algorithm,[18] which is a one-step Newton–Raphson method and qualifies as a special case of the generalized EM algorithm.[16] The standard errors of the MLEs are obtained by inverting the observed information matrix. The computational program is available from the authors upon request. Under suitable regularity conditions(see Maller and Zhou,[19] Chapter 7), it can be proved that the asymptotic distribution of the MLE $\widehat{\vartheta}$ is multivariate normal with mean vector ϑ and covariance matrix $(\widehat{\vartheta})$, which can be estimated by

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\vartheta}) = \left\{-\frac{\partial^2\ell(\vartheta)}{\partial\vartheta\partial\vartheta^\top}\right\}^{-1} = \{-J(\vartheta)\}^{-1}$$

evaluated at $\vartheta = \widehat{\vartheta}$. The required second derivatives of the matrix $J(\vartheta)$ are given in Appendix 1.

An asymptotic confidence interval with significance level α for each parameter ϑ_r is given by

$$(\hat{\vartheta}_r - z_{\alpha/2}\sqrt{\widehat{\Sigma}^{r,r}}, \hat{\vartheta}_r + z_{\alpha/2}\sqrt{\widehat{\Sigma}^{r,r}})$$

where $\widehat{\Sigma}^{r,r}$ is the rth diagonal element of $\widehat{\Sigma}(\widehat{\vartheta})$ estimated at $\widehat{\vartheta}$, for $r = 1, \ldots, p + \dim(\boldsymbol{\varphi}) + 1$, where $\dim(\cdot)$ is the dimension of the parametric space, and $z_{\alpha/2}$ is the quantile $1 - \alpha/2$ of the standard normal distribution.

Another important issue, besides estimation, refers to the hypothesis tests. Let ϑ_1 and ϑ_2 be proper disjoint subsets of ϑ. We aim to test $H_0 : \vartheta_1 = \vartheta_{01}$ against $H_1 : \vartheta_1 \neq \vartheta_{01}$ (ϑ_2 unspecified). Let $\widehat{\vartheta}_0$ be the estimate that maximize $\ell(\vartheta)$ constrained to H_0. The log-likelihood ratio (LR) statistic is given by

$$\Lambda_n = 2\Big[\ell(\widehat{\vartheta}) - \ell(\widehat{\vartheta}_0)\Big] \tag{17}$$

where $\ell(\cdot)$ is the log-likelihood. Under H_0 and some regularity conditions, Λ_n converges to the chi-square distribution with $\nu = \dim(\vartheta_1)$ degrees of freedom. Another hypothesis of interest is related to the adequacy of the promotion time cure model ($H_0 : \gamma = 0$) against the non-suitability ($H_1 : \gamma > 0$). The null distribution of the LR test under H_0 is non-standard,[20] and it has been found that the distribution can be approximated by a 50–50 mixture of the chi-square distribution with one degree of freedom (χ^2_1) and a degenerated distribution at zero. That is, the statistic Λ_n converges to the $0.5 + 0.5P\big[\chi^2_1 \leq x\big]$ distribution. In the next section, we investigate the asymptotic properties of the distributions of the MLEs and LR statistics by means of a simulation study.

5 Simulation study

We evaluate the performance of the MLEs of the parameters of the promotion time model with dispersion given by equation (4) through a simulation study. We consider the proposed model with parameters $\theta_i = \exp(\beta_0 + \beta_1 x_i)$ $(i = 1, \ldots, n)$, where $\beta_0 = 0.976$, $\beta_1 = 0.459$ and $\gamma = 2.0$. The covariate x_i is generated from a Bernoulli distribution with success probability 0.5 (with these parameters the proportion of cured, for $x_i = 0$ and $x_i = 1$ results in 30% and 20%, respectively). For simplicity, we take an exponential distribution with parameter $\varphi = 1$ for the baseline distribution.

The censored times, C_i, are sampled from the exponential distribution with failure rate where $\tau_i = \varphi(p_c + p_{0i})/(1 - p_c - p_{0i})$, $p_{0i} = \exp\{\frac{1}{\gamma}\big(1 - \sqrt{1 + 2\gamma\theta_i}\big)\}$ (the mean cured proportion is 25% with these parameters) and p_c is the proportion of censored observations. We set $p_c = p_0 = 25$ % and $p_c = p_0 + 20 = 45$ %. The observed times are generated as follow:

(i) Generate $U_i \sim U(0, 1)$;
(ii) If $U_i < p_{0i}$, then $T_i = \infty$, else $T_i = F^{-1}(((1 - \gamma\log(u_i))^2 - 1)/2\gamma\theta_i; \lambda)$, where $F^{-1}(\cdot)$ is the qf of the exponential distribution with parameter $\varphi = 1$;
(iii) Generate a censoring time C_i from the exponential distribution with failure rate, $\tau_i = \varphi(pc + p_{0i})/(1 - p_c - p_{0i})$. 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, we take the sample sizes $n = 100$, 200, 400 and 800. For each configuration, we conduct 5000 simulations to calculate the averages of the MLEs (AE), standard errors (SE), biases, roots of the mean squared errors (RMSE) of the estimates and empirical coverage probabilities (CP) corresponding to the 95% confidence intervals for the parameters in model (4). The simulation results given in Table 1 reveal that the average of maximum likelihood estimates are close to the true values and the bias, SDs and RMSEs decrease as sample size increase 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 and asymptotically normal estimates.

Additionally, we perform a simulation study to investigate the null distribution of the LR statistic, Λ_n, to test the hypotheses $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$. Table 2 summarizes the results considering different sample sizes. The rejection rates are close to the nominal levels, 1%, 5% and 10% only for moderate sample sizes in agreement with first-order asymptotic theory.

6 Application

In this section, we provide an application of the model, described in Section 2, to a dataset on a colon cancer clinical trial reported in 1989 by the North Central Cancer Treatment Group (NCCTG).[3] The dataset comes from

Table 1. Average estimates (AE), SE, bias, RMSE and CP for the parameters in the promotion time cure model with dispersion.

n	Parameter	AE	SE	Bias	RMSE	CP
		25 %				
100	γ	2.774	3.159	0.774	3.251	0.965
	φ	1.060	0.368	0.060	0.372	0.962
	β_0	1.038	0.553	0.062	0.556	0.976
	β_1	0.458	0.369	-0.002	0.369	0.947
200	γ	2.340	1.778	0.340	1.809	0.937
	φ	1.029	0.264	0.029	0.266	0.962
	β_0	0.999	0.410	0.023	0.410	0.970
	β_1	0.468	0.262	0.009	0.262	0.945
400	γ	2.075	1.010	0.075	1.012	0.931
	φ	1.033	0.198	0.033	0.200	0.965
	β_0	0.959	0.302	-0.017	0.302	0.952
	β_1	0.461	0.180	0.002	0.180	0.950
800	γ	2.070	0.678	0.070	0.681	0.951
	φ	1.007	0.129	0.007	0.130	0.955
	β_0	0.978	0.209	0.002	0.209	0.958
	β_1	0.466	0.128	0.007	0.128	0.946
		45 %				
100	γ	3.044	4.087	1.044	4.217	0.987
	φ	1.020	0.478	0.020	0.478	0.968
	β_0	1.188	0.815	0.212	0.841	0.986
	β_1	0.464	0.391	0.005	0.391	0.949
200	γ	2.493	1.759	0.493	1.825	0.962
	φ	1.006	0.378	0.006	0.377	0.969
	β_0	1.079	0.548	0.103	0.558	0.979
	β_1	0.477	0.282	0.018	0.282	0.950
400	γ	2.192	1.152	0.192	1.168	0.956
	φ	1.017	0.279	0.017	0.280	0.953
	β_0	1.006	0.393	0.030	0.394	0.969
	β_1	0.462	0.185	0.003	0.185	0.961
800	γ	2.079	0.803	0.079	0.806	0.949
	φ	1.017	0.209	0.017	0.210	0.955
	β_0	0.984	0.281	0.009	0.281	0.957
	β_1	0.462	0.134	0.003	0.134	0.941

Table 2. Empirical rejection rates of the null hypothesis $H_0 : \gamma = 0$ at 1%, 5% and 10% nominal significance levels (α).

		N			
α	p_c(%)	100	200	400	800
1	25	0.82	1.18	1.00	0.99
	45	0.79	0.88	0.94	1.09
5	25	4.26	5.12	4.98	5.02
	45	4.05	4.78	4.96	5.02
10	25	9.38	10.02	10.02	9.98
	45	9.05	9.23	10.12	10.04

Table 3. MLEs of the parameters for the promotion cure rate model with dispersion.

Parameter	Estimate (est)	Standard error (se)	|est|/se	p-value
Γ	2.234	1.300	–	–
φ_1	1.480	0.098	–	–
φ_2	2.677	0.209	–	–
β_0	−0.856	0.636	1.346	0.178
β_{1_1}	−0.040	0.155	0.256	0.798
β_{1_2}	−0.692	0.177	3.915	0.000
β_{2_1}	0.249	0.645	0.387	0.699
β_{2_2}	1.053	0.614	1.713	0.087
β_{2_3}	1.822	0.684	2.663	0.008
β_3	0.322	0.147	2.186	0.029
β_4	1.244	0.177	7.010	0.000

an assay for the efficiency of treatment by 5-FU along Levamisole in randomly selected patients with Stage C colorectal cancer and the efficiency by Levamisole alone in order to prevent recurrence as already mentioned in Section 1. The dataset includes 929 patients observed after the full resection of the tumor surgically and subsequent follow-up of the patients for 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. For our purposes, we consider the time until recurrence of the disease as the response variable. The following variables are collected from each patient:

- t_i: observed time (in years);
- x_{i1}: treatment (Observation, Levamisole, Levamisole + 5-FU);
- x_{i2}: extent of local spread (1 = submucosa, 2 = muscle, 3 = serosa, 4 = contiguous structures);
- x_{i3}: time from surgery to registration (0 = short, 1 = long);
- x_{i4}: more than 4 positive lymph nodes (0 = no, 1 = yes), $i = 1, \ldots, 888$.

The Kaplan–Meier estimate of the survival function in Figure 1 (left panel) indicates disease-free patients. Further, the Kaplan–Meier estimate of the cumulative hazard function in Figure 1 (right panel) is limited and concave, which reveals that a distribution with monotone hazard function could be adequate for modeling these data. Thus, we consider the Weibull distribution with cdf $F(t; \boldsymbol{\varphi}) = 1 - \exp\{-(t/\varphi_2)^{\varphi_1}\}$, $\varphi_i > 0$ $(i = 1, 2)$, for modeling the nonobserved times or the baseline hazard function for the lifetime dataset.

We fit the model described in Section 2 with all covariates on the mean of clonogenic cells (θ), i.e.

$$\theta = \exp\{\beta_0 + \beta_{1_1} x_{1_1} + \beta_{2_1} x_{2_1} + \beta_{2_1} x_{2_2} + \beta_{2_3} x_{2_3} + \beta_3 x_3 + \beta_4 x_4\} \tag{18}$$

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

$$x_{1_1} = \begin{cases} 1, & \text{if Levamisole;} \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad x_{1_2} = \begin{cases} 1, & \text{if Levamisole} + 5 - \text{FU;} \\ 0, & \text{otherwise} \end{cases}$$

The MLEs of the model parameters are listed in Table 3. The estimate of the shape parameter (φ_1) provides evidence against the exponential distribution ($\varphi_1 = 1$) for the promotion times.

Further, we compute the LR statistic (Λ_n) to test the suitability of the promotion time cure model for this dataset, that is, $H_0 : \gamma = 0$ versus $H_1 : \gamma > 0$. The LR statistic Λ_n, under the null hypothesis H_0, is assumed to be asymptotically distributed as a symmetric mixture of a chi-squared distribution with one degree of freedom and a point-mass at zero. Thus, Λ_n is equal to 12.867, with a p-value = 0.0001, which provides strong evidence in favor of H_1, thus indicating that the promotion time cure rate model with dispersion is more adequate for the current data. Alternatively, we consider the Akaike Information Criteria (AIC) for comparing the models. The values of the AIC for the promotion time cure model (Rodrigues et al.[9]) and the proposed model are 2345.88 and 2335.013, respectively, thus indicating that the promotion time cure rate model with dispersion yields the best fit to these data. We can note that the covariate x_4 is statistically significant. So, there is a significant difference (5%) between

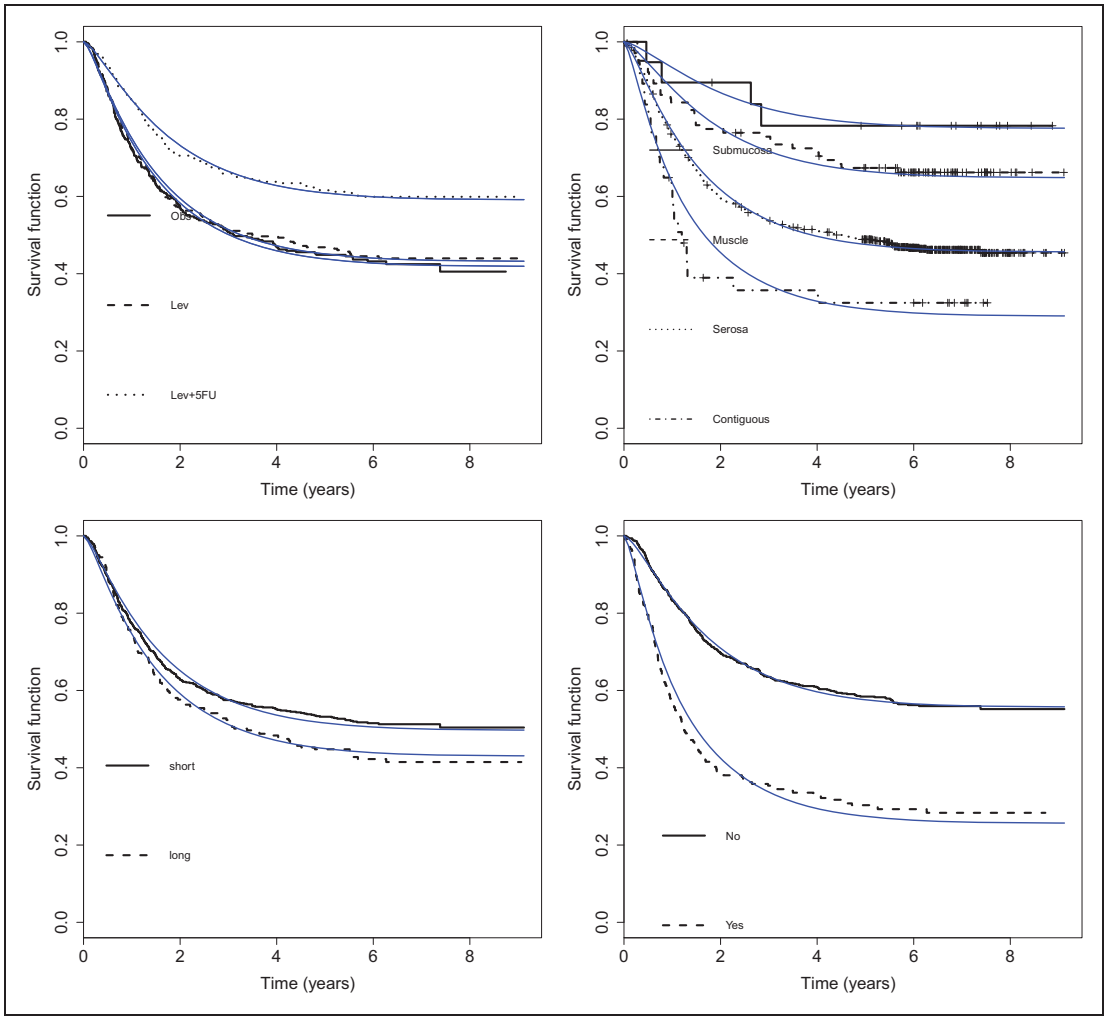

Figure 3. Kaplan–Meier curves stratified by explanatory variable and estimated survival functions to the colorectal carcinoma data, (left panel): extent of local spread and (right panel): more than four positive lymph nodes.

the levels of the covariate x_4 in relation to the cure proportion (p_0) of individuals. This can be seen in Figure 3. Also, since the parameter β_4 is related to this proportion, we can estimate this proportion. This fact can also be noted in Figure 3. We also fit the promotion time cure model with dispersion stratified by each covariate. In Figure 3, we plot the empirical survival function and the estimated survival function (2) for each covariate. We conclude that the promotion time cure model with dispersion provides a good fit to these data.

We estimate the probability of individuals disease-free (proportion being cured) after follow-up $t > 0$ from the MLEs of the model parameters by

$$\pi(t) = \exp\left\{\frac{1}{\gamma}\left(\sqrt{1 + 2\gamma\theta(1 - \exp[-(t/\varphi_2)^{\varphi_1}])} - \sqrt{1 + 2\gamma\theta}\right)\right\}$$

where θ is given by equation (18). Next, we estimate the proportion of patients disease-free for eight hypothetical patients A, B, C, D, E, F, G and H, who had resection of the tumor surgically and after treatment with values for the covariates given in Table 4. Figure 4 displays the plots with the probability of each patient being immune after treatment (left panel) for $t \in [0, 10]$ and the survival functions for recurrence times of each patient treated with colorectal cancer (right panel). Table 4 gives the MLEs and 95% asymptotic confidence intervals for the proportion of patients disease-free (cured rate) after treatment ($t = 0$). For example, for a 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 a patient E with extent of local spread in *submucosa*, time from surgery to

Table 4. MLEs, cured proportions and asymptotic 90% confidence intervals for eight hypothetical patients with colorectal cancer, for follow-up and after four years of the follow-up period.

					p_0		$\pi(4)$	
Patient	Treatment	Local Spread	time-surgery	Nodes > 4	MLE	IC (95%)	MLE	IC (95%)
A	Levamisole + 5-FU	Submucosa	Long	No	0.792	(0.636, 0.949)	0.968	(0.943, 0.993)
B	Levamisole + 5-FU	Muscle	Long	No	0.751	(0.674,0.829)	0.962	(0.946,0.978)
C	Levamisole + 5-FU	Serosa	Long	No	0.589	(0.524, 0.654)	0.937	(0.916,0.958)
D	Levamisole + 5-FU	Contiguous	Long	No	0.405	(0.268,0.542)	0.903	(0.863, 0.943)
E	Levamisole	Submucosa	Long	No	0.676	(0.467, 0.885)	0.951	(0.915 0.986)
F	Levamisole	Muscle	Long	No	0.623	(0.528, 0.718)	0.942	(0.920, 0.965)
G	Levamisole	Serosa	Long	No	0.434	(0.367, 0.500)	0.909	(0.880, 0.938)
H	Levamisole	Contiguous	Long	No	0.253	(0.132, 0.374)	0.864	(0.811 0.918)

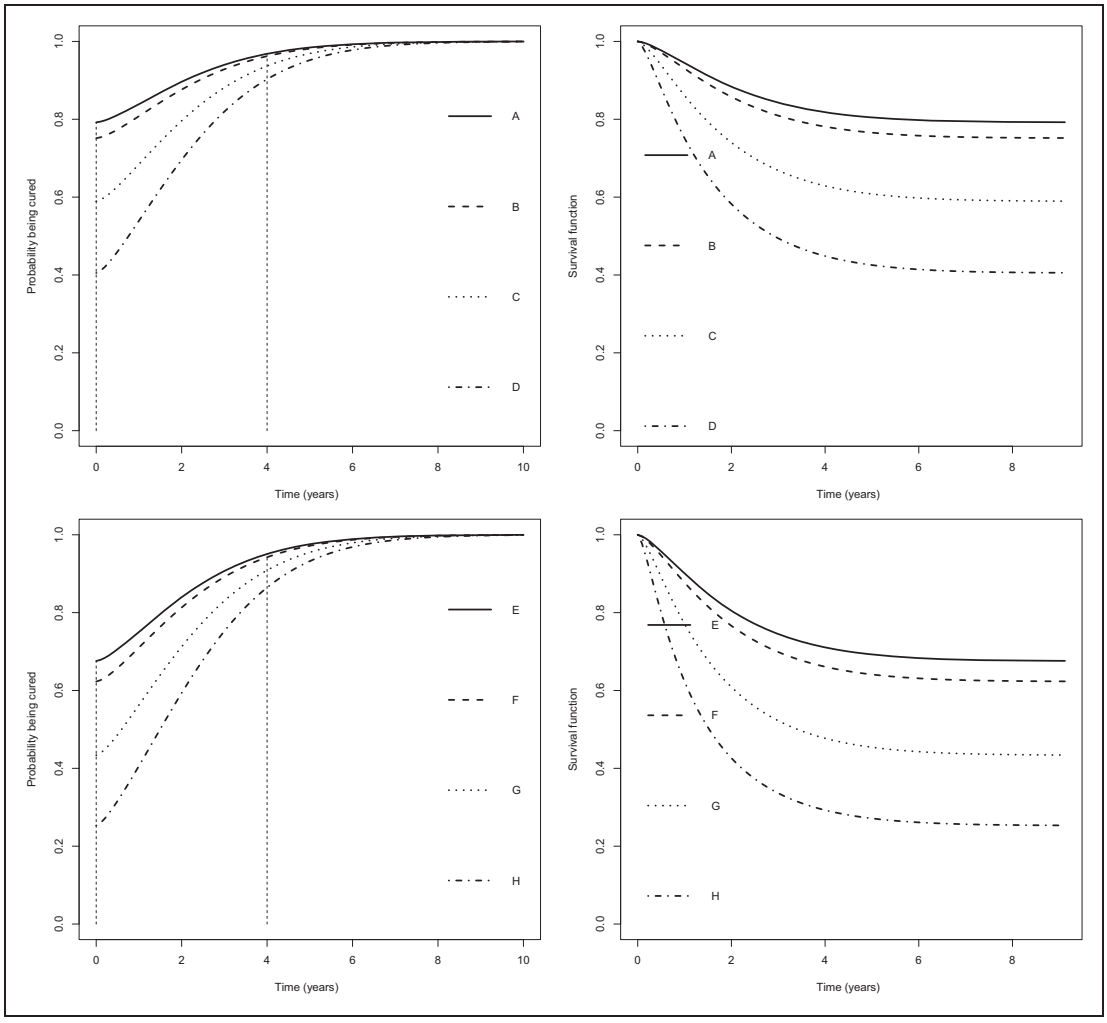

Figure 4. The probability being immune after treatment (left panel) and survival functions (right panel) for recurrence times of eight patients treated with colorectal cancer.

registration *long* and treatment *Levamisole*, we obtain different proportions, namely 0.792 for patient *A* and 0.676 for patient *E*. The probabilities of these patients being disease-free after four years are 0.968 and 0.942, respectively. The plots reveal that patients who are treated with Levamisole + 5-FU have mean time to recurrence disease larger than for those patients treated only with Levamisole, thus indicating the effectiveness of the Levamisole + 5-FU treatment.

7 Final comments

In this paper, we propose a new survival model with cure rate and examine some of its properties. The model is useful for modeling lifetime data with a cure rate fraction. Our model is an extension of the promotion time cure rate model introduced by Yakovlev and Tsodikov,[4] where we add a parameter to control the unobserved heterogeneity (or dependence) of the latent risks (or causes), which are responsible by activating the occurrence of the phenomenon of interest. We discuss its applicability in a colorectal cancer data, in which our model delivers the best fit. Moreover, from our modeling we observe that the therapy with Levamisole + 5-FU increases the lifetime of patients and cured fraction. The estimation of the model parameters has been carried out by using the maximum likelihood approach, which can be implemented straightforwardly by standard available software.[21] This fact makes the approach quite powerful and accessible to practitioners in the field.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Edwin Ortega http://orcid.org/0000-0003-3999-7402

References

1. Siegel R, DeSantis C and Jemal A. Colorectal cancer statistics. *CA: Cancer J Clinician* 2014; **64**: 104–117.
2. Kasper D, Fauci A, Hauser S, et al. *Harrison's principles of internal medicine*. USA: McGraw-Hill Education, 2015.
3. Moertel CG, Fleming TR, Macdonald JS, et al. Levamisole and fluorouracil for adjuvant therapy of resected coloncarcinoma. *New Engl J Med* 1990; **322**: 352–358.
4. Yakovlev AY and Tsodikov AD. *Stochastic models of tumor latency and their biostatistical applications*. Singapore: World Scientific, 1996.
5. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J Royal Stat Soc: Series B* 1949; **11**: 15–53.
6. Berkson J and Gage RP. Survival curve for cancer patients following treatment. *J Am Stat Assoc* 1952; **47**: 501–515.
7. Tsodikov AD, Ibrahim JG and Yakovlev AY. Estimating cure rates from survival data: An alternative to two-component mixture models. *J Am Stat Assoc* 2003; **98**: 1063–1078.
8. Ibrahim JG, Chen MH and Sinha D. *Bayesian survival analysis*. New York, NY: Springer, 2005.
9. Rodrigues J, Cancho VG, de Castro M, et al. On the unification of the long-term survival models. *Stat Probabil Lett* 2009; **79**: 753–759.
10. Ortega EMM, Cordeiro GM, Campelo AK, et al. A power series beta weibull regression model for predicting breastcarcinoma. *Stat Med* 2015; **34**: 1366–1388.
11. Cordeiro GM, Cancho VG, Ortega EMM, et al. A model with long-term survivors: negative binomial Birnbaum-Saunders. *Commun Stat-Theory Meth* 2016; **45**: 1370–1387.
12. Mudholkar GS, Srivastava DK and Friemer M. The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics* 1995; **37**: 436–445.
13. Gupta RD and Kundu D. Exponentiated exponential distribution: an alternative to gamma and Weibull distributions. *Biometric J* 2001; **43**: 117–130.
14. Nadarajah S and Kotz S. The beta exponential distribution. *Reliabil Eng Syst Safe* 2006; **91**: 689–597.
15. Press WH, Flannery BP, Teukolsky SA, et al. *Numerical recipes: the art of scientific computing*. New York, NY: Cambridge University Press, 2007.

16. Dempster AP, Laird NM and Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc Series B* 1977; **39**: 1–38.
17. McLachlan G and Krishnan T. *The EM algorithm and extensions*, Vol. 382, New Jersey: John Wiley & Sons, 2007.
18. Lange K. A gradient algorithm locally equivalent to the EM algorithm. *J Royal Stat Soc Ser B* 1995; **57**: 425–437.
19. Maller R and Zhou X. *Survival analysis with long-term survivors*. New York, NY: John Wiley & Sons, 1996.
20. Self S and Liang K. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 1987; **82**: 605–610.
21. R Core Team. *R: A Language and Environment for Statistical Computing*. Austria: R Foundation for Statistical Computing, 2016.

Appendix 1

The elements of the observed information matrix $J(\vartheta)$ for the parameters $\vartheta = (\gamma, \boldsymbol{\varphi}^\top, \beta^\top)^\top$ are given by

$$
\begin{aligned}
J_{\gamma\gamma}(\vartheta) &= \frac{2n}{\gamma^3} - 2\sum_{i=1}^{n}\frac{\delta_i\theta_i^2F^2(t_i;\boldsymbol{\varphi})}{u_i^2} + \frac{1}{\gamma^2}\sum_{i=1}^{n}\frac{\theta_iF(t_i;\boldsymbol{\varphi})[1+3\gamma\theta_iF(t_i;\boldsymbol{\varphi})]}{\sqrt{u_i^3}} \\
&\quad + \frac{1}{\gamma^2}\sum_{i=1}^{n}\frac{\theta_iF(t_i;\boldsymbol{\varphi})}{\sqrt{u_i}} - \frac{2}{\gamma^3}\sum_{i=1}^{n}\sqrt{u_i}, \\
J_{\gamma\varphi_q}(\vartheta) &= \sum_{i=1}^{n}\frac{\delta_i\theta_i[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q}}{u_i^2} + \sum_{i=1}^{n}\frac{\theta_i^2F(t_i;\boldsymbol{\varphi})[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q}}{\sqrt{u_i^3}}, \\
J_{\gamma\beta_j}(\vartheta) &= \sum_{i=1}^{n}\frac{\delta_iF(t_i;\boldsymbol{\varphi})[\dot{\theta}_i]_{\beta_j}}{u_i^2} + \left(1-\frac{1}{\gamma}\right)\sum_{i=1}^{n}\frac{F(t_i;\boldsymbol{\varphi})[\dot{\theta}_i]_{\beta_j}}{\sqrt{u_i}} \\
&\quad + \sum_{i=1}^{n}\frac{\theta_iF^2(t_i;\boldsymbol{\varphi})[\dot{\theta}_i]_{\beta_j}}{\sqrt{u_i^3}}, \\
J_{\varphi_q\varphi_{q'}}(\vartheta) &= \sum_{i=1}^{n}\frac{\delta_i[\ddot{f}(t_i;\boldsymbol{\varphi})]_{\varphi_q\varphi_{q'}}}{f(t_i;\boldsymbol{\varphi})} - \sum_{i=1}^{n}\frac{\delta_i[\dot{f}(t_i;\boldsymbol{\varphi})]_{\varphi_q}[\dot{f}(t_i;\boldsymbol{\varphi})]_{\varphi_{q'}}}{f^2(t_i;\boldsymbol{\varphi})} \\
&\quad - \gamma\sum_{i=1}^{n}\frac{\theta_i[\ddot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q\varphi_{q'}}}{u_i} + 2\gamma^2\sum_{i=1}^{n}\frac{\theta_i^2[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q}[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_{q'}}}{u_i^2} \\
&\quad - \sum_{i=1}^{n}\frac{\theta_i[\ddot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q\varphi_{q'}}}{\sqrt{u_i}} + \gamma\sum_{i=1}^{n}\frac{\theta_i^2[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q}[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_{q'}}}{\sqrt{u_i}}, \\
J_{\varphi_q\beta_j}(\vartheta) &= -\gamma\sum_{i=1}^{n}\frac{[\dot{\theta}_i]_{\beta_j}[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q}}{u_i} + 2\gamma^2\sum_{i=1}^{n}\frac{\theta_iF(t_i;\boldsymbol{\varphi})[\dot{\theta}_i]_{\beta_j}[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q}}{u_i^2} \\
&\quad - \sum_{i=1}^{n}\frac{[\dot{\theta}_i]_{\beta_j}[\ddot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q}}{\sqrt{u_i}} + \gamma\sum_{i=1}^{n}\frac{F(t_i;\boldsymbol{\varphi})[\dot{\theta}_i]_{\beta_j}[\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q}}{\sqrt{u_i^3}}
\end{aligned}
$$

and

$$
\begin{aligned}
J_{\beta_j\beta_{j'}}(\vartheta) &= \sum_{i=1}^{n}\frac{\delta_i[\ddot{\theta}_i]_{\beta_j\beta_{j'}}}{\theta_i} - \sum_{i=1}^{n}\frac{\delta_i[\dot{\theta}_i]_{\beta_j}[\dot{\theta}_i]_{\beta_{j'}}}{\theta_i^2} - \gamma\sum_{i=1}^{n}\frac{\delta_iF(t_i;\boldsymbol{\varphi})[\ddot{\theta}_i]_{\beta_j\beta_{j'}}}{u_i} \\
&\quad + 2\gamma^2\sum_{i=1}^{n}\frac{\delta_iF^2(t_i;\boldsymbol{\varphi})[\dot{\theta}_i]_{\beta_j}[\dot{\theta}_i]_{\beta_{j'}}}{u_i^2} - \sum\frac{F(t_i;\boldsymbol{\varphi})[\ddot{\theta}_i]_{\beta_j\beta_{j'}}}{\sqrt{u_i}} \\
&\quad + \gamma\sum_{i=1}^{n}\frac{F^2(t_i;\boldsymbol{\varphi})[\dot{\theta}_i]_{\beta_j}[\dot{\theta}_i]_{\beta_{j'}}}{\sqrt{u_i^3}}
\end{aligned}
$$

where

$$
\begin{aligned}
[\dot{f}(t_i;\boldsymbol{\varphi})]_{\varphi_q} &= \frac{\partial f(t_i;\boldsymbol{\varphi})}{\partial \varphi_q}, [\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q} = \frac{\partial F(t_i;\boldsymbol{\varphi})}{\partial \varphi_q} \\
[\dot{f}(t_i;\boldsymbol{\varphi})]_{\varphi_{q'}} &= \frac{\partial f(t_i;\boldsymbol{\varphi})}{\partial \varphi_{q'}}, [\dot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_{q'}} = \frac{\partial F(t_i;\boldsymbol{\varphi})}{\partial \varphi_{q'}}, \\
[\ddot{f}(t_i;\boldsymbol{\varphi})]_{\varphi_q\varphi_{q'}} &= \frac{\partial f(t_i;\boldsymbol{\varphi})}{\partial \varphi_q \partial \varphi_{q'}}, [\ddot{F}(t_i;\boldsymbol{\varphi})]_{\varphi_q\varphi_{q'}} = \frac{\partial F(t_i;\boldsymbol{\varphi})}{\partial \varphi_q \partial \varphi_{q'}}, \\
[\dot{\theta}_i]_{\beta_j} &= \frac{\partial g^{-1}(\mathbf{x}_i'\beta)}{\partial \beta_j}, [\dot{\theta}_i]_{\beta_{j'}} = \frac{\partial g^{-1}(\mathbf{x}_i'\beta)}{\partial \beta_{j'}}, \\
[\ddot{\theta}_i]_{\beta_j\beta_{j'}} &= \frac{\partial g^{-1}(\mathbf{x}_i'\beta)}{\partial \beta_j \partial \beta_{j'}}, \\
u_i &= 1 + 2\gamma\theta_i F(t_i;\boldsymbol{\varphi}), q, q' = 1, \ldots, k, j, j' = 0, 1, \ldots, p
\end{aligned}
$$