Published in final edited form as:

Stat Methodol. 2013 July ; 13: . doi:10.1016/j.stamet.2013.01.006.

The destructive negative binomial cure rate model with a latent activation scheme

Vicente G. Cancho^a, Dipankar Bandyopadhyay^b, Francisco Louzada^a, and Bao Yiqi^a

^aInstituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Brazil

^bDivision of Biostatistics, University of Minnesota, Minneapolis, MN, USA

Abstract

A new flexible cure rate survival model is developed where the initial number of competing causes of the event of interest (say lesions or altered cells) follow a compound negative binomial (NB) distribution. This model provides a realistic interpretation of the biological mechanism of the event of interest as it models a destructive process of the initial competing risk factors and records only the damaged portion of the original number of risk factors. Besides, it also accounts for the underlying mechanisms that leads to *cure* through various latent activation schemes. Our method of estimation exploits maximum likelihood (ML) tools. The methodology is illustrated on a real data set on malignant melanoma, and the finite sample behavior of parameter estimates are explored through simulation studies.

Keywords

competing risks; cure rate models; long-term survival models; negative binomial distribution

1. Introduction

The event of interest in many survival studies or cancer-relapse trials can be the death of a patient (due to various competing causes) or a tumor recurrence (attributed to metastasis-component tumor cells left active after an initial treatment). However with the recent advances in (cancer) treatment therapies, a high portion of the subjects are expected to be *cured*, i.e. remaining disease free after prolonged follow-ups. In this vein, there is now a vast literature on ‘cure rate models’ for survival data, also called ‘survival models with a surviving fraction’ or ‘long-term survival models’, though majority of these stems from either one of the *2-component Mixture Cure* model of Berkson & Gage (1952), or the *Bounded Cumulative Hazard* (or, BCH model) of Yakovlev & Tsodikov (1996). The books by Ibrahim *et al.* (2001) and Maller & Zhou (1996) are authoritative references of these two classes respectively, and covers a wide range of developments.

Recently, Rodrigues *et al.* (2010) have extended the BCH model of Chen *et al.* (1999) through a special case of the destructive (compound) weighted Poisson distribution. The

© 2013 Elsevier B.V. All rights reserved.

*Address for correspondence: Dipankar Bandyopadhyay, Division of Biostatistics, University of Minnesota School of Public Health, A452 Mayo MMC 303, 420 Delaware St. S.E., Minneapolis, MN 55455, USA. dbandyop@umn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

authors argue that this development is a much more realistic alternative to the unified cure rate model of Rodrigues *et al.* (2009) in explaining the biological mechanism undergoing the occurrence of the event of interest in presence of a cured fraction. This is because the proposed compound weighted cure rate survival model presumes that the original number of risk factors (i.e., lesions, or altered cells) are not repaired or eliminated after some intensive treatment, and this group of *damaged* or unrepaired cells (or latent factors) represented by a compound variable are potentially competing to give rise to tumor, or a *risk* of failure. Despite being a step in the right direction, the modeling scheme fails to unravel a (possibly) sequential progression scheme towards disease manifestation.

Motivated by the work of Cooner *et al.* (2007), here we introduce a destructive (compound) negative binomial (DNB) cure rate model which demonstrates greater flexibility than the proposition of Cancho *et al.* (2011) by considering a stochastic framework for the ordered sequence of latent factors leading to disease manifestation. Our parametrization allows us to model both under- and over-dispersion, commonly encountered in discrete data. In our proposition, the initial number of (unobservable) event-related competing causes are assumed to follow a negative binomial (NB) distribution, and the total number of competing causes (the sum of independently Bernoulli distributed random variables indicating presence/absence of a cause) conditional on the number of initial causes experience various (latent) activation mechanisms. The primary model parameters, i.e., the expected initial number of competing causes related to the event of interest (denoted as θ) and the probability of a competing cause not being eliminated by the initial treatment (denoted by p) are connected to covariates through appropriate link functions.

The rest of the paper proceeds as follows. In Section 2, we formulate the DNB cure rate model. Inference based on maximum-likelihood (ML) is developed in Section 3. Two simulation studies are presented in Section 4 to study some finite sample properties. In Section 5, our methodology is illustrated on a real dataset on malignant melanoma. Finally, Section 6 presents some concluding remarks.

2. Model

Let M be an unobservable NB random variable denoting the initial number of competing causes related to the occurrence of an event of interest, with probability mass function (p.m.f) given by

$$P[M=m; \theta, \eta] = \frac{\Gamma(\eta^{-1} + m)}{\Gamma(\eta^{-1}) m!} \left(\frac{\eta\theta}{1 + \eta\theta} \right)^m (1 + \eta\theta)^{-1/\eta}, m=0, 1, 2, \dots, \quad (1)$$

where $\theta > 0$, $\eta > -1$ and $\eta\theta + 1 > 0$. Negative values of $-1 < \eta < 0$ lead to a range for m from 0 to the largest integer less than $-1/\eta$ (Ross & Preece, 1985). Since $E(M) = \theta$ and $\text{var}(M) = \theta + \eta\theta^2$, the values of $\eta > 0$ (< 0) correspond to over (under)-dispersion relative to the Poisson distribution. As $\eta \rightarrow 0$ the NB approaches the Poisson distribution (Johnson *et al.*, 2005), and for $\eta = 1$, the NB reduces to the geometric distribution with parameter $1/(1 + \theta)$.

The immediate consequence of a prolonged treatment is the formation (or not) of precancerous lesions. Given $M = m$, let $W_j, j = 1, 2, \dots, m$ be independent random variables (independent of M) following a Bernoulli distribution with success probability p indicating presence of the j -th lesion (or competing cause), or the probability of an undestroyed clonogenic cell. The variable N , denoting the total number of altered cells among the M initial cells (competing causes) which are not destroyed or eliminated by the treatment is defined as:

$$N = \begin{cases} W_1 + W_2 + \dots + W_M, & \text{if } M > 0, \\ 0, & \text{if } M = 0. \end{cases} \quad (2)$$

By damaged or destruction, we mean $N = M$. Following Yang & Chen (1991), the initial M competing causes are primary initiated malignant cells, where W_j in (2) denotes the number of living malignant cells that are descendants of the initiated malignant cell j during some time interval. In this case, N denotes the total number of living malignant cells at some specific time. The conditional distribution of N (given $M = m$) is referred to as the damaged distribution, with probability generating function (p.g.f) of N is given by

$$\phi_N(s) = \{1 + \eta\theta p(1-s)\}^{-1/\eta}, \quad (3)$$

which is the p.g.f of a NB distribution with parameters η and θp .

The time to event for the j th competing cause is represented by Z_j , $j = 1, \dots, N$. We assume that conditional on N , the Z_j are i.i.d. with cumulative distribution function $F(t)$ and survival function $S(t) = 1 - F(t)$. Also, we assume that $Z_1, Z_2, \dots$ are independent of N and M . The total number of competing causes N and the time Z_j are not observable (latent). As in Cooner *et al.* (2006) and Cooner *et al.* (2007), cure rate models with latent activations assume that the observed failure time Y is generated by the *activation times* of M latent factors. $M = 0$ leads to $Y = \infty$, which means the individual is not exposed to any latent factors, and is considered *cured*. Hence, if we assume R out of M causes are required to produce the event of interest, the observable time to event is defined by $Y = Z_{(R)}$, where R depends on M , $Z_{(1)}, Z_{(2)}, \dots, Z_{(R)}, \dots, Z_{(M)}$ are the ordered Z_j . Hence R can be a fixed constant, a function of M or a random variable specified through a conditional distribution on M . Here, we consider three specifications for R . First, we assume that given $M \geq 1$, the conditional distribution of R is *Discrete Uniform* on $\{1, 2, \dots, M\}$ (random activation scheme). Under this setup, we obtain the DNB cure rate model with a random activation scheme (DNB-RA), with surviving function for the population given by

$$S_{\text{pop}}(y) = P(Y > y) = P(N = 0) + \{1 - P(M = 0)\} S(y) = \left(\frac{1}{1 + \eta\theta p}\right)^{1/\eta} + \left\{1 - \left(\frac{1}{1 + \eta\theta p}\right)^{1/\eta}\right\} S(y), \quad (4)$$

with the cured fraction $p_0 = P(M = 0) = \lim_{t \rightarrow \infty} S_{\text{pop}}(y) = \{1 + \eta\theta p\}^{-1/\eta}$. From (4) the

density function can be derived as $f_{\text{pop}}(y) = -S'_{\text{pop}}(y) = \left\{1 - \left(\frac{1}{1 + \eta\theta p}\right)^{1/\eta}\right\} f(y)$, where $f(y) = -S'(y)$ denotes the proper density function of the time to event Z . Furthermore, the corresponding hazard function is

$$h_{\text{pop}}(y) = \left[\left\{1 - \left(\frac{1}{1 + \eta\theta p}\right)^{1/\eta}\right\} f(y) \right] / \left[\left(\frac{1}{1 + \eta\theta p}\right)^{1/\eta} + \left\{1 - \left(\frac{1}{1 + \eta\theta p}\right)^{1/\eta}\right\} S(y) \right].$$

As our second setup, we consider the so-called *first activation* scheme (Cooner *et al.*, 2007) and obtain the DNB cure rate model with a first activation (DNB-FA). We assume the event of interest happens due to any one of the possible causes. Therefore, for $R = 1$, the time to event is $Y = Z_{(1)} = \min\{Z_1, \dots, Z_M\}$, implying that

$$S_{\text{pop}}(y) = \phi_M(S(y)) = \{1 + \eta\theta p F(y)\}^{-1/\eta}, \quad (5)$$

where $\phi_N(\cdot)$ is given in (3) with the cured fraction given by $p_0 = \{1 + \eta\theta p\}^{-1/\eta}$. The density function associated with (5) is given by $f_{\text{pop}}(y) = \theta p f(y) \{1 + \eta\theta p F(y)\}^{-1/\eta-1}$ with the hazard

function $h_{\text{pop}}(y) = \theta p f(y)(1 + \eta \theta p F(y))^{-1}$. The model in (5) with $p = 1$ implies the model proposed by Cancho *et al.* (2011).

Finally, as our third setup, we consider the *last activation* scheme where the event of interest only happens after all the M causes have occurred, such that $R = M$ and the observed failure time is $Y = Z_{(M)} = \max\{Z_1, \dots, Z_M\}$. According to Cooner *et al.* (2007), $S_{\text{pop}}(y) = 1 + \varphi_N(0) - \varphi_N(F(y))$. Hence from (3), we obtain the DNB cure rate model with a last activation (DNB-LA) with

$$S_{\text{pop}}(y) = 1 + \left(\frac{1}{1 + \eta \theta p} \right)^{1/\eta} - \left(\frac{1}{1 + \eta \theta p S(y)} \right)^{1/\eta}, \quad (6)$$

so that the cured fraction is $p_0 = \{1 + \eta \theta p\}^{-1/\eta}$. The surviving function in (6) leads to the density given by $f_{\text{pop}}(y) = \theta p f(y) \{1 + \eta \theta p S(y)\}^{-1/\eta - 1}$, and the hazard function

$$h_{\text{pop}}(y) = \left[\theta p f(y) \{1 + \eta \theta p S(y)\}^{-1/\eta - 1} \right] / \left[1 + \left(\frac{1}{1 + \eta \theta p} \right)^{1/\eta} - \left(\frac{1}{1 + \eta \theta p S(y)} \right)^{1/\eta} \right].$$

Observing p_0 for each one of the three activation schemes, we realize that the cured fraction remains the same irrespective of the activation schemes. The models however differ by its surviving, density and hazard functions. Of course, the particular cases of (4), (5) and (6) for $\eta \rightarrow 0$ and $\eta = 1$ leads to various activation schemes for the destructive Poisson cure rate model (DP-RA, DP-FA and DP-LA), and the destructive geometric cure rate model (DG-RA, DG-FA and DG-LA) respectively. The relationship between the population survival functions under different schemes is described in next proposition. Denoting the survival functions in (4), (5) and (6) by $S_{\text{pop}}^R(y)$, $S_{\text{pop}}^F(y)$ and $S_{\text{pop}}^L(y)$ respectively, we have the following proposition, whose proof is available in Appendix A.

Proposition 1. Under conditions for the survival functions as in (4), (5) and (6), we have $S_{\text{pop}}^F(y) \leq S_{\text{pop}}^R(y) \leq S_{\text{pop}}^L(y)$ for $y > 0$.

Hereafter, we assume a Weibull distribution for the unobserved time Z with $F(z; \gamma) = 1 - \exp(-z^{\gamma_1} e^{\gamma_2/2})$ and $f(z; \gamma) = \gamma_1 z^{\gamma_1 - 1} \exp(\gamma_2 - z^{\gamma_1} e^{\gamma_2/2})$, for $z > 0$, $\gamma_2 \in \mathbb{R}$ and $\gamma = (\gamma_1, \gamma_2)^T$. The Weibull distribution is considered one of the most versatile and widely used lifetime distributions in survival analysis. Depending on the value of its shape parameter γ_1 , the Weibull distribution can be used to model a variety of life behaviors. Its failure rate is monotone decreasing for $\gamma_1 < 1$, monotone increasing for $\gamma_1 > 1$, and for $\gamma_1 = 1$ it is constant, leading to the exponential distribution. Figure 1 (left panel) shows distinct behaviors of the surviving functions of models in (4), (5) and (6) with $\gamma_1 = 2$, $\gamma_2 = -1$, $\eta = 2$, $\theta = 3$ and $p = 0.5$ where we can observe that $S_{\text{pop}}^F(y) \leq S_{\text{pop}}^R(y) \leq S_{\text{pop}}^L(y)$. These plots illustrate the flexibility a orded by our proposal.

The (proper) surviving function for the non-cured population under the Negative Binomial-Weibull (NBW) framework, denoted by S_{NBW} , is computed as $S_{\text{NBW}}(y) = P(Y > y | M = 1)$. For the RA scheme, S_{NBW} has the same distribution as the latent random variable Z_j 's, i.e., $S_{\text{NBW}}(y) = S(y)$. For the FA scheme, the survival function for the non-cured population is given by

$$S_{\text{NBW}}(y) = \frac{[1 + p\eta\theta F(y)]^{-1/\eta} - (1 + p\eta\theta)^{-1/\eta}}{1 - (1 + p\eta\theta)^{-1/\eta}}, y > 0, \quad (7)$$

We note that $S_{NBW}(0) = 1$ and $S_{NBW}(\infty) = 0$, so that (7) is a proper survival function. The NBW pdf is given by

$$f_{NBW}(y) = \frac{p\theta f(y) [1+p\eta\theta F(y)]^{-(1/\eta+1)}}{1-(1+p\eta\theta)^{-1/\eta}}, y > 0 \quad (8)$$

In Figure 2, we plot the NBW FA density functions for some fixed values of $\gamma_1, \gamma_2, p, \eta$ and θ . These plots indicate that the NBW FA distribution is very flexible and that the values of γ_1, p, η and θ have a substantial effect on its skewness and kurtosis. From (7) and (8), it is easy to verify that the corresponding NBW hazard function for the population of non-cured can be derived as:

$$h_{NBW}(y) = \frac{p\theta h(t) S(t) [1+p\eta\theta F(y)]^{-(1/\eta+1)}}{[1+p\eta\theta F(y)]^{-1/\eta} - (1+p\eta\theta)^{-1/\eta}}, y > 0. \quad (9)$$

where $h(y)$ and $S(y)$ are hazard and survival functions of a Weibull distribution respectively. From (9) $h_{NBW}(y)/h(y)$ is increasing in $y > 0$ for all γ_1, p, η and θ . Beside, $h(y) \rightarrow h_{NBW}(y)$ and hence we have that the limit behavior of the hazard function of the NBW distribution is the same as the behavior of the hazard function of the Weibull distribution. Figure 3 display shapes of the NBW hazard function for some arbitrary parameter values.

The survival function for the non-cured population under the LA mechanism is given by

$$S_{NBW}(y) = \frac{1 - [1+p\eta\theta (1-F(y))]^{-1/\eta}}{1 - (1+p\eta\theta)^{-1/\eta}}, y > 0, \quad (10)$$

We note that $S_{NBW}(0) = 1$ and $S_{NBW}(\infty) = 0$, so that (10) is a proper survival function. The NBW pdf is then given by

$$f_{NBW}(y) = \frac{p\theta f(y) [1+p\eta\theta (1-F(y))]^{-(1/\eta+1)}}{1 - (1+p\eta\theta)^{-1/\eta}}, y > 0. \quad (11)$$

In Figure 4, we plot the NBW LA density functions for some arbitrary values of $\gamma_1, \gamma_2, p, \eta$ and θ . These plots indicate that the NBW LA distribution is very flexible and that the values of γ_1, p and θ have a substantial effect on its skewness and kurtosis. From (10) and (11), it is easy to verify that the corresponding NBW hazard function for the population of non-cured is given

$$h_{NBW}(y) = \frac{p\theta h(t) S(t) [1+p\eta\theta F(y)]^{-(1/\eta+1)}}{1 - [1+p\eta\theta (1-F(y))]^{-1/\eta}}, y > 0, \quad (12)$$

where $h(y)$ and $S(y)$ are hazard and survival functions of a Weibull distribution respectively. From (12), $h_{NBW}(y)/h(y)$ is decreasing in $y > 0$ for all a, p, η and θ . Beside, $h(y) \rightarrow h_{NBW}(y)$, and hence (as earlier), the limit behavior of the hazard function of the NBW distribution is the same as the behavior of the hazard function of the Weibull distribution. Figure 5 display shapes of the hazard function for the NBW distribution under LA for some arbitrary parameter values. The relationship between the non-cured population survival functions under different mechanisms is the same as in Proposition 1 (See Figure 1, right panel). There is a mathematical relationship between the models (5) and (6) and the mixture cure rate model (Boag, 1949; Berkson & Gage, 1952). We can write $S_{pop}(y) = (1 + p\eta\theta)^{-1/\eta} + (1 - (1 + p\eta\theta)^{-1/\eta}) S_{NBW}(y)$.

$+ p\eta\theta^{-1/\eta}S_{\text{NBW}}(y)$, where $S_{\text{NBW}}(y)$ is given by (7) or (10). Thus, $S_{\text{pop}}(y)$ is a mixture cure rate model with cure rate equal to $p_0 = (1 + p\eta\theta)^{-1/\eta}$ and survival function $S_{\text{NBW}}(y)$ for the non-cured population. This result implies that every mixture cure rate model corresponds to some model of the form (5) (FA mechanism) or (6) (LA mechanism) for any p , α and θ and $S(\cdot)$.

3. Inference

Let us now consider the situation where the lifetime in Section 2 is not completely observed and is subject to right-censoring, where C_i denotes the censoring time. For a sample of size n , we observe $T_i = \min\{Y_i, C_i\}$ and $\delta_i = I(Y_i \leq C_i)$, where $\delta_i = 1$ if it is an event-time and 0, if it is right-censored. Note that the destructive NB model in Section 2 is unidentifiable in the sense of Li *et al.* (2001). To circumvent this problem, we propose to relate the model parameters p and θ to covariates x_{1i} and x_{2i} , respectively, without common elements and without an intercept for x_{2i} . We adopt the common link functions given by

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_{1i}^T \beta_1 \text{ and } \log(\theta_i) = \mathbf{x}_{2i}^T \beta_2, \quad (13)$$

where β_1 and β_2 denote the respective parameter vectors. For our data consisting of n pairs of time to event and censoring indicators given by $\{(t_1, \delta_1), \dots, (t_n, \delta_n)\}$, the likelihood function under non-informative censoring is given by

$$L(\vartheta; \mathcal{D}) \propto \prod_{i=1}^n \{f_{\text{pop}}(t_i; \vartheta)\}^{\delta_i} \{S_{\text{pop}}(t_i; \vartheta)\}^{1-\delta_i}, \quad (14)$$

where $\vartheta = (\eta, \beta_1^T, \beta_2^T, \gamma^T)^T$, $\mathcal{D} = (\mathbf{t}, \delta, \mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{t} = (t_1, \dots, t_n)^T$, and $\delta = (\delta_1, \dots, \delta_n)^T$, $\mathbf{x}_1 = (x_{11}, \dots, x_{1n})^T$, $\mathbf{x}_2 = (x_{21}, \dots, x_{2n})^T$ whereas $f_{\text{pop}}(\cdot; \vartheta)$ and $S_{\text{pop}}(\cdot; \vartheta)$ are given in Section 2. The log-likelihood function $\ell = \ell_{\text{DNB-RA}}(\vartheta, \mathcal{D})$ for the DNB-RA case is given by

$$\ell = \sum_{i=1}^n \delta_i \log f_w(t_i) + \delta_i \log(1 - T_{(0,0,-1/\eta)}^{(i)}) + (1 - \delta_i) \log(T_{(0,0,-1/\eta)}^{(i)} + S_w(t_i)(1 - T_{(0,0,-1/\eta)}^{(i)})).$$

with the corresponding score function

$U_{\text{DNB-RA}}(\vartheta) = (\partial \ell / \partial \gamma_1, \partial \ell / \partial \gamma_2, \partial \ell / \partial \eta, \partial \ell / \partial \beta_1^T, \partial \ell / \partial \beta_2^T)^T$, whose components are given in the Appendix B1. Similarly, the log-likelihood function for the DNB-FA is $\ell = \ell_{\text{DNB-FA}}(\vartheta, \mathcal{D})$ given by

$$\ell = \sum_{i=1}^n \log \theta_i - \log p_i + \delta_i \log[f_w(t_i)] - (\delta_i + 1/\eta) \log T_{(1,0,1)}^{(i)},$$

with the corresponding score function

$U_{\text{DNB-FA}}(\vartheta) = (\partial \ell / \partial \gamma_1, \partial \ell / \partial \gamma_2, \partial \ell / \partial \eta, \partial \ell / \partial \beta_1^T, \partial \ell / \partial \beta_2^T)^T$, the components of the score appear in Appendix C1.

Once again, the log-likelihood function for the DNB-LA case $\ell = \ell_{\text{DNB-LA}}(\vartheta, \mathcal{D})$ is given by

$$\ell = \sum_{i=1}^n \log \theta_i + \log p_i + \sum_{i=1}^n \delta_i \log [f_w(z_i)] - (1 + \eta^{-1}) \sum_{i=1}^n \delta_i \log T_{0,1,1}^{(i)} + \sum_{i=1}^n (1 - \delta_i) \log \left(1 - T_{(0,0,-1/\eta)}^{(i)} - T_{(0,1,-1/\eta)}^{(i)} \right), \quad (15)$$

with the corresponding score function

$$U_{\text{DNB-LA}}(\vartheta) = \left(\partial \ell / \partial \gamma_1, \partial \ell / \partial \gamma_2, \partial \ell / \partial \eta, \partial \ell / \partial \beta_1^T, \partial \ell / \partial \beta_2^T \right)^T, \text{ whose components appear in Appendix D1.}$$

Maximum likelihood (ML) estimation for the parameter vector ϑ can be implemented via direct numerical maximization of the log-likelihood function $\ell(\vartheta) = \log L(\vartheta; \mathbf{D})$ for all the 3 schemes which can be accomplished using standard software R (R Development Core Team, 2010). Confidence intervals (CIs) and hypothesis testing can be conducted by using the large sample distribution of the MLE which is a normal distribution with the covariance matrix as the inverse of the Fisher information since regularity conditions are satisfied. More specifically, the asymptotic covariance matrix is given by $\mathbf{I}^{-1}(\vartheta)$ with $\mathbf{I}(\vartheta) = -E[\mathbf{J}(\vartheta)]$ where $\mathbf{J}(\vartheta) = \partial^2 \ell(\vartheta) / \partial \vartheta \partial \vartheta^T$. Since it is not possible to compute the Fisher information matrix $\mathbf{I}(\vartheta)$ due to the censored observations (censoring is random and noninformative), we resort to the matrix of second derivatives of the log likelihood, $-\mathbf{J}(\vartheta)$, evaluated at the MLE $\vartheta = \hat{\vartheta}$, which is a consistent estimator. The expressions of the matrix $\mathbf{J}(\vartheta)$ for the DNB-RA, DNB-FA and DNB-LA models are provided in Appendix B2, C2 and D2 respectively. The required second derivatives are computed numerically. Besides estimation, testing of hypothesis is another key issue. Let ϑ_1 and ϑ_2 be proper disjoint subsets of ϑ . We aim to test $H_0: \vartheta_1 = \vartheta_{01}$ against $H_1: \vartheta_1 \neq \vartheta_{01} / (\vartheta_2 \text{ unspecified})$. Let $\hat{\vartheta}_0$ maximize $L(\vartheta; \mathcal{D})$ constrained to H_0 and define the log-likelihood ratio (LR) statistic

$$w_n = 2 \left(\ell(\hat{\vartheta}) - \ell(\hat{\vartheta}_0) \right) \quad (16)$$

where $\ell(\cdot)$ is the log-likelihood. Under H_0 and some regularity conditions, w_n converges in distribution to the chi-square distribution with $\dim(\vartheta_1)$ degrees of freedom. Alternatively different models can be compared using the Akaike information criterion given by

$$AIC = -2\ell(\hat{\vartheta}) + 2\#(\vartheta) \text{ and the Schwartz-Bayesian criterion defined by}$$

$$SBC = -2\ell(\hat{\vartheta}) + \#(\vartheta) \log(n), \text{ where } \#(\vartheta) \text{ is the number of model parameters. The model with the smallest value of any of these criteria (among all models considered) is commonly taken as the preferred model for describing a given dataset.}$$

4. Simulation study

To evaluate the finite sample performance of parameter estimates in our proposition, we resort to two simulation studies. In the first one, we examine how our estimates compare with the true values using the mean squared error (MSE) and the coverage probability of the 95% CIs. In the second one, we compare how the DNB-FA and DNB-LA models identifies the true data generation scheme using the AIC criterion presented in Section 3. For the first simulation, we consider the DNB-FA model given in (5) with the Weibull distribution for the event times (Z) with parameter $\gamma_1 = 2$ and $\gamma_2 = -3.2$. For each individual i , $i = 1, \dots, n$, the number of causes (N_i) which are not destroyed or eliminated for this individual is generated from the NB distribution with parameter $\eta = 3$, and $\theta_i p_i$. In our simulations we have two covariates, x_{1i} and x_{2i} , such that, x_{1i} is generated from a Bernoulli(0.5) and x_{2i} generated from $N(0, 1)$. Thus, $p_i = \exp(\beta_{10} + \beta_{11}x_{1i}) / (1 + \exp(\beta_{10} + \beta_{11}x_{1i}))$ and $\theta_i = \exp(x_{2i}\beta_{21})$, where $\beta_{10} = -1$, $\beta_{11} = 0.5$ and $\beta_{21} = 0.7$. The censoring times were sampled from

Uniform(0, τ), where τ was set in order to control the proportion of censored observations on average to be approximately 78%.

We consider sample sizes of $n = 150, 300$ and 600 . For each of these schemes, we conducted 1000 iterations to calculate the average of MLEs, the mean squared error (MSE) of the MLEs and coverage probability of 95% CI for parameters in model (5). The simulation results are shown in Table 1. We observe that the averages MLEs are close to the true values, the MSEs decrease as sample size increases and the empirical coverage probabilities are closer to the nominal coverage level when sample size increases, which are all expected if the underlying estimation scheme is working correctly to produce consistent and asymptotically normal estimates. In our second simulation study of misspecification rates, we generated separately 1000 samples from each of the DNB-FA and DNB-LA schemes with $\eta = 0, 1, 3$ and the other parameters fixed at the same values as in the first simulation scheme and with sample sizes equals to $n = 150, 300$ and 600 . Both models (DNB-FA and DNB-LA) were then fitted to each samples. Table 4 presents the proportion of times the AIC criterion selected the true model from which the data was generated. The high proportions supports the fact that the AIC criterion can distinguish between the various activations. Similar results were obtained by considering the LRS (9), as well as the SBC criterion.

5. Application

In this section, we demonstrate an application of our models detailed in Section 2 to a well-known dataset on a Phase III cutaneous melanoma clinical trial (labeled as E1690 data), available at <http://merlot.stat.uconn.edu/~mhchen/survbook/>, and conducted by the Eastern Cooperative Oncology Group (Kirkwood *et al.*, 2000; Ibrahim *et al.*, 2001). The incidence of melanoma ranks among the highest among solid tumor growths, with high mortality rates (between 60%-75%) despite early detection and screening (Cooner *et al.*, 2007). The dataset here comes from an assay for the evaluation of postoperative treatment performance with a high dose of a certain drug (interferon alpha-2b) in order to prevent recurrence. Patients were included in the study from 1991 to 1995, and follow-up was conducted until 1998.

Here, we consider relapse-free survival (RFS) time (in years) as the response variable. After deleting subjects with incomplete data and missing observation times, we have a subset of $n = 408$ patients with approximately 43% of censoring. The following variables were collected from each patient: Observed time (in years, mean=2.31, standard deviation= 1.93); x_{1i} : treatment (0: observation, $n=198$; 1: interferon alfa-2b, $n=210$); x_{2i} : age (in years, mean=48.1, standard deviation= 13.1); x_{3i} : nodal category (1, $n = 110$; 2, $n = 131$; 3, $n = 86$; 4, $n = 81$) and x_{4i} : tumor thickness (in mm, mean=3.98 and standard deviation=3.22), $i = 1, \dots, 408$. As part of preliminary data analysis, Figure 6, which displays Kaplan–Meier survival curves stratified by treatment, confirms a well pronounced plateau with the survival curves leveling o between 0.25 and 0.42. This is indicative of the presence of a proportion of patients for whom the malignant melanoma will never recur (i.e., they are cured), which motivates fitting a cure-rate survival model to this dataset.

Due to issues of identifiability in the proposed model (as discussed in Section 3), the probability p of undestroyed causes (clonogenic cell) since initial stages of the treatment is related to x_{1i} , x_{3i} and x_{4i} through a logit link, while the parameter η (the expected number of initial competing causes, or clonogenic cells) is related to x_{2i} without intercept though a log link. The selection of covariates for p and η here remains an interesting issue. We considered all possible choices of covariates preserving identifiability, and this particular combination provided us the least AIC and SBC values for each of the categories of models (DP: Poisson, DG: Geometric, and DNB) and activation schemes (FA, LA and RA). Table 3

presents the *AIC* and *SBC* values for each candidate models. Overall, all the competing models under the first activation (FA) scheme outperforms their counterparts (LA and RA). Within the FA, the values of *AIC* and *SBC* are fairly close for the 3 models, though the DG model outperforms the other two slightly. The QQ plot of the normalized randomized quantile residuals (Dunn & Smyth, 1996; Rigby & Stasinopoulos, 2005) presented in the left panel of Figure 7 suggests that the DG-FA model produces an adequate fit. Considering the model selection criteria and the QQ plot, we will consider the DG-FA model as our working model, and present further analysis results based on this model.

Table 4 presents the MLEs and their standard errors for the parameters of the DG-FA model. The estimate of the shape parameter (γ_1) presents an evidence against the exponential distribution ($\gamma_1 = 1$) for the unobserved lifetimes. Considering a LR statistic (16), we test the effect of treatment in the probability of undestroyed causes, i.e., $H_0 : \beta_{1,\text{treatment}} = 0$ versus $H_A : \beta_{1,\text{treatment}} \neq 0$ resulting in $w_n = 0.700$ (p-value= 0.4027), indicating that the treatment effect is not significant. Hence, we also present the MLEs and their standard errors for the parameters in DG-FA model without the ‘treatment’ covariate in Table 4. The estimated probability (and standard error) of undestroyed causes of patients with tumor thickness of 3.175 mm (median thickness) and stratified by nodule category from 1 to 4 are given by 0.146538 (0.045623), 0.48480 (0.11697), 0.83759 (0.12614), and 0.965828 (0.047308) respectively, for this reduced DG-FA model without the covariate ‘treatment’. The right panel of Figure 7 display plots of these estimated probabilities stratified by nodule category and tumor thickness.

Figure 8 displays the estimated surviving function of patients with median tumor thickness (3.175 mm), stratified by nodal category (from 1 to 4) for patients with ages equal to 38, 47 and 58 years corresponding to the first, median and the third quartiles. These plots highlight the combined impact of the covariates on the cured fraction. Finally, Table 5 presents the MLE of the cured fraction (p_0) for cases displayed in Figure 8, with the 95% CIs obtained using the delta method. For the selected tumor thickness values, the 95% confidence intervals do not overlap. The plot and the table both reveals that the survival probability and cure rate diminishes more rapidly with increasing nodal category, and also with increasing age. Figure 9 displays the estimated non-cured surviving function under first activation.

6. Conclusion

In this paper, we proposed a destructive negative binomial (DNB) cure rate model which encompasses the cure rate model discussed in Cancho *et al.* (2011). The proposed model allows estimation of the proportion of causes not eliminated by initial treatment (undestroyed causes) and includes as special cases the destructive Poisson and geometric cure rate models. Our proposition exploits a (possible) latent activation scheme in developing a robust stochastic framework for disease (cancer) manifestation. We apply our methodology to a popular cutaneous melanoma study to illustrate how the procedure developed can be used to evaluate model assumptions and obtain parameter estimates.

The models considered in this article can be fitted using standard available software (viz. R), which makes the approach quite powerful and accessible to practitioners in the field. Connecting model parameters to important co-variates through appropriate link functions preserves model identifiability, and facilitates interpretation. Future developments might involve other parametric or semi-parametric extensions of the lifetime distribution under the proposed set up.

Acknowledgments

We thank the anonymous referee whose constructive comments led to an improved version of the manuscript. V. G. Cancho and F. Louzada acknowledges funding support from CNPq and FAPESP (Brazil). D. Bandyopadhyay is supported by grants P20RR017696-06, R03DE020114 and R03DE021762 from the US National Institutes of Health.

Appendix

Appendix A: Proof of Proposition 1

Proof: Let $S(y)$ be any arbitrary survival function. Then, the function $\frac{1-(1+p\eta\theta S(y))^{-1/\eta}}{S(y)}$ is increasing on y . Also,

$$\lim_{y \rightarrow 0} \frac{1-(1+p\eta\theta S(y))^{-1/\eta}}{S(y)} = 1-(1+p\theta\eta)^{-1/\eta} \Rightarrow \frac{1-(1+p\eta\theta S(y))^{-1/\eta}}{S(y)} \geq 1-(1+p\theta\eta)^{-1/\eta}, \forall y.$$

$$1-(1+p\eta\theta S(y))^{-1/\eta} \geq S(y) \left(1-(1+p\eta\theta)^{-1/\eta}\right) \Rightarrow 1$$

$$+(1+p\theta\eta)^{-1/\eta}$$

$$-(1+p\eta\theta S(y))^{-1/\eta} \geq (1+p\theta\eta)^{-1/\eta}$$

Hence, $+S(y) \left(1-(1+p\eta\theta)^{-1/\eta}\right)$, which implies that

$$S_{\text{pop}}^L(y) \geq S_{\text{pop}}^A(y). \text{ Similarly we can prove that } S_{\text{pop}}^F \leq S_{\text{pop}}^A.$$

Appendix B1: Components of the score function $U_{DNB-RA}(\vartheta)$ for the DNB-RA model

$$\begin{aligned} \frac{\partial \ell}{\partial \gamma_1} &= \sum_{i=1}^n \frac{\delta_i f_{\gamma_1}}{f_w(t_i)} + \sum_{i=1}^n \frac{(1-\delta_i) \left(1-T_{(0,0,-1/\eta)}^{(i)}\right) S_{\gamma_1}}{T_{(0,0,-1/\eta)}^{(i)} + S_w(t_i) \left(1-T_{(0,0,-1/\eta)}^{(i)}\right)}; \\ \frac{\partial \ell}{\partial \gamma_2} &= \sum_{i=1}^n \frac{\delta_i f_{\gamma_2}}{f_w(t_i)} + \sum_{i=1}^n \frac{(1-\delta_i) \left(1-T_{(0,0,-1/\eta)}^{(i)}\right) S_{\gamma_2}}{T_{(0,0,-1/\eta)}^{(i)} + S_w(z_i) \left(1-T_{(0,0,-1/\eta)}^{(i)}\right)}; \\ \frac{\partial \ell}{\partial \eta} &= - \sum_{i=1}^n \frac{\delta_i T_{\eta(0,0,-1/\eta)}^{(i)}}{1-T_{(0,0,-1/\eta)}^{(i)}} \\ &\quad - \frac{\left(1-\delta_i F_w(t_i) T_{\eta(0,0,-1/\eta)}^{(i)}\right)}{T_{(0,0,-1/\eta)}^{(i)} F_w(t_i) + S_w(t_i)}; \\ \frac{\partial \ell}{\partial \beta_{1j}} &= - \sum_{i=1}^n \frac{\delta_i p_{i\beta_{1j}} \theta_i T_{(0,0,-1/\eta-1)}^{(i)}}{1-T_{(0,0,-1/\eta)}^{(i)}} \\ &\quad - \frac{(1-\delta_i) p_{i\beta_{1j}} T_{(0,0,-1/\eta-1)}^{(i)} F_w(t_i)}{T_{(0,0,-1/\eta)}^{(i)} F_w(t_i) + S_w(t_i)}; \\ \frac{\partial \ell}{\partial \beta_{2l}} &= - \sum_{i=1}^n \frac{\delta_i \theta_i \beta_{2l} p_i T_{(0,0,-1/\eta-1)}^{(i)}}{1-T_{(0,0,-1/\eta)}^{(i)}} \\ &\quad - \frac{(1-\delta_i) p_{i\beta_{2l}} T_{(0,0,-1/\eta-1)}^{(i)} F_w(t_i)}{T_{(0,0,-1/\eta)}^{(i)} F_w(t_i) + S_w(t_i)}, \end{aligned}$$

where $f_W(t_i)$ and $S_W(t_i)$ are probability density and survival function of the Weibull distribution,

$$\begin{aligned} f_{\gamma_1} &= \frac{\partial f_W(t_i)}{\partial \gamma_1} = t_i^{\gamma_1-1} \exp(\gamma_2 - t_i^{\gamma_1} e^{\gamma_2}) \left(1 + \gamma_1(\gamma_1 - 1) t_i^{-1} - \gamma_1^2 t_i^{\gamma_1-1} e^{\gamma_2}\right); \\ f_{\gamma_2} &= \frac{\partial f_W(t_i)}{\partial \gamma_2} = \exp(\gamma_2 - t_i^{\gamma_1} e^{\gamma_2}) \left(\gamma_1 t_i^{\gamma_1-1} - t_i^{\gamma_1} e^{\gamma_2}\right); \\ F_{\gamma_1} &= \frac{\partial F_W(t_i)}{\partial \gamma_1} = t_i^{\gamma_1} \log t_i \exp(2\gamma_2 - t_i^{\gamma_1} e^{\gamma_2}); \\ F_{\gamma_2} &= \frac{\partial F_W(t_i)}{\partial \gamma_2} = \exp(\gamma_2 - t_i^{\gamma_1} e^{\gamma_2}) (t_i^{\gamma_1} e^{\gamma_2} - 1); \\ S_{\gamma_1} &= -F_{\gamma_1}; \text{ and } S_{\gamma_2} = -F_{\gamma_2}; \\ p_{i\beta_{1j}} &= \frac{\partial p_i}{\partial \beta_{1j}} = \frac{\exp(\mathbf{x}_{1i}^T \beta_1) x_{1ij}}{(1 + \exp(\mathbf{x}_{1i}^T \beta_1))^2}; j=0, \dots, n_1; \\ \theta_{i\beta_{2l}} &= \frac{\partial \theta_i}{\partial \beta_{2l}} = \exp(\mathbf{x}_{2i}^T \beta_2) x_{2il}; l=0, \dots, n_2; \end{aligned}$$

and we define

$$\begin{aligned} T_{(z,k,l)}^{(i)} &= \left[1 + p_i \eta \theta_i F_W(t_i)^z S_W(t_i)^k\right]^l, z, k, l \in \mathbb{R}, i=1 \dots n \\ T_{\beta_{1j}(z,k,-1/\eta)}^{(i)} &= \frac{\partial T_{(z,k,-1/\eta)}^{(i)}}{\partial \beta_{1j}} \\ &= -1/\eta T_{(z,k,-1/\eta-1)}^{(i)} p_{i\beta_{1j}} \eta \theta_i F_W(t_i)^z S_W(t_i)^k; \\ T_{\beta_{2l}(z,k,-1/\eta)}^{(i)} &= \frac{\partial T_{(z,k,-1/\eta)}^{(i)}}{\partial \beta_{2l}} = -1/\eta T_{(z,k,-1/\eta-1)}^{(i)} \theta_{i\beta_{2l}} \eta p_i F_W(t_i)^z S_W(t_i)^k; \\ T_{\eta(z,k,-1/\eta)}^{(i)} &= \frac{\partial T_{(z,k,-1/\eta)}^{(i)}}{\partial \eta} = \eta^{-2} T_{(z,k,-1/\eta)}^{(i)} \left[\log \left[T_{(z,k,1)}^{(i)} \right] - p_i \eta \theta_i F_W(t_i)^z S_W(t_i)^k T_{(z,k,-1)}^{(i)} \right]. \end{aligned} \quad (17)$$

Appendix B2: Matrix $J(\vartheta)$ of the DNB-RA model

We derive the necessary formulas to obtain the second-order partial derivatives of the log-likelihood function $\ell_{DNB-RA}(\vartheta, \mathcal{D})$. The components of the matrix are given by

$$\begin{aligned} \ell_{\gamma_1 \gamma_1} &= -\sum_{i=1}^n \delta_i f(t_i)^{-2} f_{\gamma_1} + \delta_i f(t_i)^{-1} f_{\gamma_1 \gamma_1} - \frac{(1-\delta_i) T_{(0,0,-1/\eta)}^{(i)} \left(1 - T_{(0,0,-1/\eta)}^{(i)}\right) S_{\gamma_1}^2}{\left(T_{(0,0,-1/\eta)}^{(i)} + S_W(t_i) \left(1 - T_{(0,0,-1/\eta)}^{(i)}\right)\right)^2}; \\ \ell_{\gamma_1 \gamma_2} &= -\sum_{i=1}^n \delta_i f(t_i)^{-2} f_{\gamma_2} + \delta_i f(t_i)^{-1} f_{\gamma_1 \gamma_2} - \frac{(1-\delta_i) T_{(0,0,-1/\eta)}^{(i)} \left(1 - T_{(0,0,-1/\eta)}^{(i)}\right) S_{\gamma_1} S_{\gamma_2}}{\left(T_{(0,0,-1/\eta)}^{(i)} + S_W(t_i) \left(1 - T_{(0,0,-1/\eta)}^{(i)}\right)\right)^2}; \\ \ell_{\gamma_1 \eta} &= -\sum_{i=1}^n (1-\delta_i) S_{\gamma_1} \left[\frac{p_i \theta_i \left(1 - T_{(0,0,-1/\eta)}^{(i)}\right) T_{(0,0,-1/\eta)}^{(i)} \log T_{(0,0,1)}^{(i)} F_W(t_i)}{\eta^2 \left(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)\right)^2} + \frac{p_i \eta^{-2} \theta_i T_{(0,0,-1/\eta)}^{(i)} \log T_{(0,0,1)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} \right]; \end{aligned}$$

$$\begin{aligned}
\ell_{\gamma_1\beta_{1j}} &= \sum_{i=1}^n S_{\gamma_1} \left[\frac{\theta_i p_i \beta_{1j} T_{(0,0,-1/\eta-1)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} + \frac{p_i \beta_{1j} \theta_i F_W(t_i) T_{(0,0,-1/\eta-1)}^{(i)} (1 - T_{(0,0,-1/\eta)}^{(i)})}{\left(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)\right)^2} \right]; \\
\ell_{\gamma_1\beta_{2l}} &= \sum_{i=1}^n S_{\gamma_1} \left[\frac{p_i \theta_i \beta_{2l} T_{(0,0,-1/\eta-1)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} + \frac{\theta_i \beta_{1j} p_i F_W(t_i) T_{(0,0,-1/\eta-1)}^{(i)} (1 - T_{(0,0,-1/\eta)}^{(i)})}{\left(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)\right)^2} \right]; \\
\ell_{\gamma_2\gamma_2} &= - \sum_{i=1}^n \delta_i f(t_i)^{-2} f_{\gamma_2} + \delta_i f(t_i)^{-1} f_{\gamma_2\gamma_2} - \frac{(1 - \delta_i) T_{(0,0,-1/\eta)}^{(i)} (1 - T_{(0,0,-1/\eta)}^{(i)}) S_{\gamma_2}^2}{\left(T_{(0,0,-1/\eta)}^{(i)} + S_W(t_i) (1 - T_{(0,0,-1/\eta)}^{(i)})\right)^2}; \\
\ell_{\gamma_2\eta} &= - \sum_{i=1}^n (1 - \delta_i) S_{\gamma_2} \left[\frac{p_i \theta_i (1 - T_{(0,0,-1/\eta)}^{(i)}) T_{(0,0,-1/\eta)}^{(i)} \log T_{(0,0,1)}^{(i)} F_W(t_i)}{\eta^2 \left(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)\right)^2} + \frac{p_i \eta^{-2} \theta_i T_{(0,0,-1/\eta)}^{(i)} \log T_{(0,0,1)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} \right]; \\
\ell_{\gamma_2\beta_{1j}} &= \sum_{i=1}^n S_{\gamma_2} \left[\frac{\theta_i p_i \beta_{1j} T_{(0,0,-1/\eta-1)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} + \frac{p_i \beta_{1j} \theta_i F_W(t_i) T_{(0,0,-1/\eta-1)}^{(i)} (1 - T_{(0,0,-1/\eta)}^{(i)})}{\left(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)\right)^2} \right];
\end{aligned}$$

$$\begin{aligned}
\ell_{\gamma_2\beta_{2l}} &= \sum_{i=1}^n S_{\gamma_2} \left[\frac{p_i \theta_{i\beta_{2l}} T_{(0,0,-1/\eta-1)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} + \frac{\theta_{i\beta_{1j}} p_i F_W(t_i) T_{(0,0,-1/\eta-1)}^{(i)} (1 - T_{(0,0,-1/\eta)}^{(i)})}{(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i))^2} \right]; \\
\ell_{\eta\eta} &= \sum_{i=1}^n \delta_i \left(\frac{T_{\eta,(0,0,-1/\eta)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)}} \right)^2 - \frac{\delta_i T_{\eta^2,(0,0,-1/\eta)}^{(i)}}{1 - T_{(0,0,-1/\eta)}^{(i)}} + \frac{(1 - \delta_i) F_W(t_i) T_{\eta^2,(0,0,-1/\eta)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} \\
&\quad - \frac{(1 - \delta^2) (F_W(t_i) T_{\eta,(0,0,-1/\eta)}^{(i)})^2}{(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i))^2}; \\
\ell_{\eta\beta_{1j}} &= \sum_{i=1}^n \frac{\delta_i \theta_{i\beta_{1j}} p_i T_{(0,0,-1/\eta-1)}^{(i)} T_{\eta,(0,0,-1/\eta)}^{(i)}}{(1 - T_{(0,0,-1/\eta)}^{(i)})^2} \\
&\quad - \frac{\delta_i T_{\eta\beta_{1j},(0,0,-1/\eta)}^{(i)}}{1 - T_{(0,0,-1/\eta)}^{(i)}} + \frac{(1 - \delta_i) F_W(t_i)}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} \left[\frac{T_{\eta\beta_{1j},(0,0,-1/\eta)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} + \frac{\theta_{i\beta_{1j}} F_W(t_i) T_{(0,0,-1/\eta-1)}^{(i)} T_{\eta,(0,0,-1/\eta)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} \right]; \\
\ell_{\eta\beta_{2l}} &= \sum_{i=1}^n \frac{\delta_i p_i \theta_{i\beta_{2l}} T_{(0,0,-1/\eta-1)}^{(i)} T_{\eta,(0,0,-1/\eta)}^{(i)}}{(1 - T_{(0,0,-1/\eta)}^{(i)})^2} \\
&\quad - \frac{\delta_i T_{\eta\beta_{2l},(0,0,-1/\eta)}^{(i)}}{1 - T_{(0,0,-1/\eta)}^{(i)}} + \frac{(1 - \delta_i) F_W(t_i)}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} \left[\frac{T_{\eta\beta_{2l},(0,0,-1/\eta)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} + \frac{p_i \theta_{i\beta_{2l}} F_W(t_i) T_{(0,0,-1/\eta-1)}^{(i)} T_{\eta,(0,0,-1/\eta)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} \right]; \\
\ell_{\beta_{1j}\beta_{1j}} &= \sum_{i=1}^n \delta_i \theta_i \left[\frac{(1 + \eta) T_{(0,0,-1/\eta-2)}^{(i)} (\theta_i (p_{i\beta_{1j}})^2 + T_{(0,0,-1)}^{(i)} p_{i\beta_{1j}^2})}{1 - T_{(0,0,-1/\eta)}^{(i)}} \right. \\
&\quad \left. - \frac{\theta_i (p_{i\beta_{1j}})^2 T_{(0,0,-2/\eta-2)}^{(i)}}{(1 - T_{(0,0,-1/\eta)}^{(i)})^2} \right] \\
&\quad + (1 - \delta_i) \theta_i F_W(t_i) \left[\frac{p_{i\beta_{1j}^2} T_{(0,0,-1/\eta-1)}^{(i)} + (1 + \eta) \theta_i p_{i\beta_{1j}} T_{(0,0,-1/\eta-2)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} + \frac{\theta_i (p_{i\beta_{1j}})^2 F_W(t_i)}{(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i))^2} \right]; \\
\ell_{\beta_{1j}\beta_{2l}} &= \sum_{i=1}^n \delta_i p_{i\beta_{1j}} \left[\frac{(1 + \eta) \theta_{i\beta_{2l}} (p_i \theta_i + T_{(0,0,1)}^{(i)}) T_{(0,0,-1/\eta-2)}^{(i)}}{1 - T_{(0,0,-1/\eta)}^{(i)}} + \frac{p_i \theta_i \theta_{i\beta_{2l}} T_{(0,0,-2/\eta-2)}^{(i)}}{(1 - T_{(0,0,-1/\eta)}^{(i)})^2} \right] \\
&\quad (1 - \delta_i) F_W(t_i) p_{i\beta_{1j}} \theta_{i\beta_{2l}} \left[\frac{T_{(0,0,-1/\eta-1)}^{(i)} + (\eta + 1) p_i T_{(0,0,-1/\eta-2)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} + \frac{\theta_i p_i T_{(0,0,-2/\eta-2)}^{(i)} F_W(t_i)}{(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i))^2} \right]; \\
\ell_{\beta_{2l}\beta_{2l}} &= \sum_{i=1}^n \delta_i p_i \left[\frac{(1 + \eta) T_{(0,0,-1/\eta-2)}^{(i)} (p_i (\theta_{i\beta_{2l}})^2 + T_{(0,0,1)}^{(i)} \theta_{i\beta_{2l}^2})}{1 - T_{(0,0,-1/\eta)}^{(i)}} - \frac{p_i (\theta_{i\beta_{2l}})^2 T_{(0,0,-2/\eta-2)}^{(i)}}{(1 - T_{(0,0,-1/\eta)}^{(i)})^2} \right] \\
&\quad + (1 - \delta_i) p_i F_W(t_i) \left[\frac{\theta_{i\beta_{2l}^2} T_{(0,0,-1/\eta-1)}^{(i)} + (1 + \eta) p_i \theta_{i\beta_{2l}} T_{(0,0,-1/\eta-2)}^{(i)}}{T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i)} \right. \\
&\quad \left. + \frac{p_i (\theta_{i\beta_{2l}})^2 F_W(t_i)}{(T_{(0,0,-1/\eta)}^{(i)} F_W(t_i) + S_W(t_i))^2} \right];
\end{aligned}$$

where

$$\begin{aligned}
 f_{\gamma_1 \gamma_1} &= \frac{\partial^2 \ell}{\partial \gamma_1^2} = t_i^{\gamma_1-1} \exp(\gamma_2 - t_i^{\gamma_1} e^{\gamma_2}) \left[\log t_i + \gamma_1 (\gamma_1 - 1) (2\gamma_1 - 1) t_i^{-1} \log t_i - 2\gamma_1 t_i^{\gamma_1-1} e^{\gamma_2} (\gamma_1 \log t_i + 1) \right]; \\
 f_{\gamma_1 \gamma_2} &= \frac{\partial^2 \ell}{\partial \gamma_1 \partial \gamma_2} = -t_i^{\gamma_1} \exp(\gamma_2 - t_i^{\gamma_1} e^{\gamma_2}) \left(\gamma_1^2 t_i^{\gamma_1-1} e^{\gamma_2} \right) (1 - t_i^{\gamma_1} e^{\gamma_2}) \left(1 + \gamma_1 (\gamma_1 - 1) t_i^{-1} - \gamma_1^2 t_i^{\gamma_1-1} e^{\gamma_2} \right); \\
 f_{\gamma_2 \gamma_2} &= \frac{\partial^2 \ell}{\partial \gamma_2^2} = -\exp(\gamma_2 - t_i^{\gamma_1} e^{\gamma_2}) (t_i^{\gamma_1} e^{\gamma_2}) (1 - t_i^{\gamma_1} e^{\gamma_2}) (\gamma_1 t_i^{\gamma_1} - t_i^{\gamma_1} e^{\gamma_2}); \\
 p_{i\beta_{1j}^2} &= \frac{\partial p_i}{\partial \beta_{1j}^2} = \frac{\exp(\mathbf{x}_{1i}^\top \beta_1) (x_{1ij})^2 (1 + 3\exp(\mathbf{x}_{1i}^\top \beta_1))}{(1 + \exp(\mathbf{x}_{1i}^\top \beta_1))^3}; \\
 \theta_{i\beta_{2l}^2} &= \frac{\partial \theta_i}{\partial \beta_{2l}^2} = \exp(\mathbf{x}_{2i}^\top \beta_2) (x_{2il})^2; \\
 T_{\eta^2, (0,0,-1/\eta)}^{(i)} &= \frac{\partial (T_{(0,0,-1/\eta)}^{(i)})^2}{\partial \eta^2} = \left(-2\eta^{-3} T_{(j,k,-1/\eta)}^{(i)} + \eta^{-2} T_{\eta, (j,k,-1/\eta)}^{(i)} \right) \left[\log [T_{(j,k,1)}^{(i)}] \right. \\
 &\quad \left. - p_i \eta \theta_i F_W(t_i) S_W(t_i) T_{(j,k,-1)}^{(i)} \right] \\
 &\quad + \left(\eta^{-2} T_{(j,k,-1/\eta)}^{(i)} \right) p_i \theta_i F_W(t_i) S_W(t_i) T_{(j,k,-1)}^{(i)} \left[1 - p_i \eta \theta_i F_W(t_i) S_W(t_i) T_{(j,k,-1)}^{(i)} \right]; \\
 T_{\eta \beta_{1j}, (0,0,-1/\eta)}^{(i)} &= \frac{\partial (T_{(0,0,-1/\eta)}^{(i)})^2}{\partial \eta \beta_{1j}} = \eta^{-2} T_{\beta_{1j}, (z,k,-1/\eta)}^{(i)} \left[\log [T_{(z,k,1)}^{(i)}] - p_i \eta \theta_i F_W(t_i)^z S_W(t_i)^k T_{(z,k,-1)}^{(i)} \right] \\
 &\quad + p_i (\theta_i F_W(t_i))^2 p_{i\beta_{1j}}; \\
 T_{\eta \beta_{2l}, (0,0,-1/\eta)}^{(i)} &= \frac{\partial (T_{(0,0,-1/\eta)}^{(i)})^2}{\partial \eta \beta_{2l}} = \eta^{-2} T_{\beta_{2l}, (z,k,-1/\eta)}^{(i)} \left[\log [T_{(z,k,1)}^{(i)}] - p_i \eta \theta_i F_W(t_i)^z S_W(t_i)^k T_{(z,k,-1)}^{(i)} \right] \\
 &\quad + \theta_i (p_i F_W(t_i) S_W(t_i))^2 \theta_{i\beta_{2l}};
 \end{aligned}$$

Appendix C1: Components of the score function $U_{DNB-FA}(\vartheta)$ of DNB-FA model

$$\begin{aligned}
 \frac{\partial \ell}{\partial \gamma_1} &= \sum_{i=1}^n \frac{\delta_i f_{\gamma_1}}{f_W(t_i)} - (\delta_i + 1/\eta) p_i \eta \theta_i T_{(1,0,-1)}^{(i)} F_{\gamma_1}; \\
 \frac{\partial \ell}{\partial \gamma_2} &= \sum_{i=1}^n \frac{\delta_i f_{\gamma_2}}{f_{textW}(t_i)} - (\delta_i + 1/\eta) p_i \eta \theta_i T_{(1,0,-1)}^{(i)} F_{\gamma_2}; \\
 \frac{\partial \ell}{\partial \eta} &= -\sum_{i=1}^n \eta^{-2} \log [T_{(1,0,1)}^{(i)}] + (\delta_i + 1/\eta) p_i \theta_i F_W(t_i) T_{(1,0,-1)}^{(i)}; \\
 \frac{\partial \ell}{\partial \beta_{1j}} &= \sum_{i=1}^n p_i^{-1} p_{i\beta_{1j}} - (\delta_i + 1/\eta) \eta \theta_i F_W(t_i) T_{(1,0,-1)}^{(i)} p_{i\beta_{1j}}; \\
 \frac{\partial \ell}{\partial \beta_{2l}} &= \sum_{i=1}^n \theta_i^{-1} \theta_{i\beta_{2l}} - (\delta_i + 1/\eta) \eta p_i F_W(t_i) T_{(1,0,-1)}^{(i)} \theta_{i\beta_{2l}}.
 \end{aligned}$$

Appendix C2: Matrix $J^{(\vartheta)}$ of the DNB-FA model

Now, we derive the necessary formulas to obtain the second-order partial derivatives of the log-likelihood function $\ell_{DNB-FA}(\vartheta, \mathcal{D})$. The components of the matrix are given by

$$\begin{aligned}
 \ell_{\gamma_1 \gamma_1} &= -\sum_{i=1}^n \delta_i \left(f_W(t_i)^{-2} f_{\gamma_1} - f_W(t_i)^{-1} f_{\gamma_1 \gamma_1} \right) + (\delta_i \eta + 1) p_i \theta_i \left[T_{(1,0,-1)}^{(i)} F_{\gamma_1 \gamma_1} - p_i \eta \theta_i F_{\gamma_1}^2 T_{(1,0,-2)}^{(i)} \right]; \\
 \ell_{\gamma_1 \gamma_2} &= -\sum_{i=1}^n \delta_i \left(f_W(t_i)^{-2} f_{\gamma_2} - f_W(t_i)^{-1} f_{\gamma_1 \gamma_2} \right) + (\delta_i \eta + 1) \eta (p_i \theta_i)^2 F_{\gamma_1} F_{\gamma_2} \left[T_{(1,0,-1)}^{(i)} - p_i \eta \theta_i T_{(1,0,-2)}^{(i)} \right]; \\
 \ell_{\gamma_1 \eta} &= -\sum_{i=1}^n p_i \theta_i F_{\gamma_1} \left(\delta_i T_{(1,0,-1)}^{(i)} - (\delta_i \eta + 1) p_i \theta_i F_{\gamma_1} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\gamma_1 \beta_{1j}} &= \sum_{i=1}^n (\delta_i + 1/\eta) p_i \beta_{1j} \eta \theta_i F_{\gamma_1} \left(T_{(1,0,-1)}^{(i)} - p_i \eta \theta_i F_{\gamma_1} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\gamma_1 \beta_{2l}} &= \sum_{i=1}^n (\delta_i + 1/\eta) \theta_i \beta_{2l} \eta p_i F_{\gamma_1} \left(T_{(1,0,-1)}^{(i)} - p_i \eta \theta_i F_{\gamma_1} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\gamma_2 \gamma_2} &= -\sum_{i=1}^n \delta_i \left(f_W(t_i)^{-2} f_{\gamma_2} - f_W(t_i)^{-1} f_{\gamma_2 \gamma_2} \right) + (\delta_i \eta + 1) p_i \theta_i \left[T_{(1,0,-1)}^{(i)} F_{\gamma_2 \gamma_2} - p_i \eta \theta_i F_{\gamma_2}^2 T_{(1,0,-2)}^{(i)} \right]; \\
 \ell_{\gamma_2 \eta} &= -\sum_{i=1}^n p_i \theta_i F_{\gamma_2} \left(\delta_i T_{(1,0,-1)}^{(i)} - (\delta_i \eta + 1) p_i \theta_i F_{\gamma_2} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\gamma_2 \beta_{1j}} &= \sum_{i=1}^n (\delta_i + 1/\eta) p_i \beta_{1j} \eta \theta_i F_{\gamma_2} \left(T_{(1,0,-1)}^{(i)} - p_i \eta \theta_i F_{\gamma_2} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\gamma_2 \beta_{2l}} &= \sum_{i=1}^n (\delta_i + 1/\eta) \theta_i \beta_{2l} \eta p_i F_{\gamma_2} \left(T_{(1,0,-1)}^{(i)} - p_i \eta \theta_i F_{\gamma_2} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\eta \eta} &= 2\eta^{-3} \log \left[T_{(1,0,1)}^{(i)} \right] - p_i \eta^{-2} \theta_i T_{(1,0,-1)}^{(i)} F_W(t_i) + \theta_i p_i F_W(t_i) \left[\eta^{-2} + (\delta_i + 1/\eta) p_i \theta_i F_{\gamma_1} T_{(1,0,-2)}^{(i)} \right]; \\
 \ell_{\eta \beta_{1j}} &= -\sum_{i=1}^n \eta^{-1} T_{(1,0,-1)}^{(i)} F_W(t_i) \theta_i p_i \beta_{1j} + (\delta_i + 1/\eta) F_W(t_i) \theta_i p_i \beta_{1j} \left(T_{(1,0,-1)}^{(i)} - p_i \eta \theta_i F_{\gamma_1} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\eta \beta_{2l}} &= -\sum_{i=1}^n \eta^{-1} T_{(1,0,-1)}^{(i)} F_W(t_i) p_i \theta_i \beta_{2l} + (\delta_i + 1/\eta) F_W(t_i) p_i \theta_i \beta_{2l} \left(T_{(1,0,-1)}^{(i)} - p_i \eta \theta_i F_{\gamma_1} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\beta_{1j} \beta_{1j}} &= \sum_{i=1}^n -\left(p_i^{-1} p_{i\beta_{1j}} \right)^2 + p_i^{-1} p_{i\beta_{1j}^2} - (\delta_i \eta + 1) \theta_i F_W(t_i) \left[p_{i\beta_{1j}^2} T_{(1,0,-1)}^{(i)} - \eta \theta_i (p_{i\beta_{1j}})^2 F_{\gamma_2} T_{(1,0,-2)}^{(i)} \right]; \\
 \ell_{\beta_{1j} \beta_{2l}} &= -\sum_{i=1}^n (\delta_i \eta + 1) F_W(t_i) p_i \beta_{1j} \theta_i \beta_{2l} \left(T_{(1,0,-1)}^{(i)} - p_i \eta \theta_i F_{\gamma_1} \theta_i \beta_{2l} T_{(1,0,-2)}^{(i)} \right); \\
 \ell_{\beta_{2l} \beta_{2l}} &= \sum_{i=1}^n -\left(\theta_i^{-1} \theta_{i\beta_{2l}} \right)^2 + \theta_i^{-1} \theta_{i\beta_{2l}^2} - (\delta_i \eta + 1) \theta_i F_W(t_i) \left[\theta_{i\beta_{2l}^2} T_{(1,0,-1)}^{(i)} - \eta p_i (\theta_{i\beta_{2l}})^2 F_{\gamma_2} T_{(1,0,-2)}^{(i)} \right].
 \end{aligned}$$

Appendix D1: Components of the score function $U_{DNB-LA}(\vartheta)$ of DNB-LA model

$$\begin{aligned}\frac{\partial \ell}{\partial \gamma_1} &= \sum_{i=1}^n \frac{\delta_i f_{\gamma_1}}{f_w(t_i)} - (\eta+1) \delta_i T_{0,1,-1}^{(i)} p_i \theta_i S_{\gamma_1} + \frac{(1-\delta_i) p_i \theta_i T_{(0,1,-1/\eta-1)}^{(i)} S_{\gamma_1}}{1 - T_{(0,0,-1/\eta)}^{(i)} - T_{0,1,-1/\eta}^{(i)}}; \\ \frac{\partial \ell}{\partial \gamma_2} &= \sum_{i=1}^n \frac{\delta_i f_{\gamma_2}}{f_w(t_i)} - (\eta+1) \delta_i T_{0,1,-1}^{(i)} p_i \theta_i S_{\gamma_2} + \frac{(1-\delta_i) p_i \theta_i T_{(0,1,-1/\eta-1)}^{(i)} S_{\gamma_2}}{1 - T_{(0,0,-1/\eta)}^{(i)} - T_{0,1,-1/\eta}^{(i)}}; \\ \frac{\partial \ell}{\partial \eta} &= \sum_{i=1}^n \eta^{-2} \delta_i \log T_{(0,1,1)}^{(i)} - \left(1 + \eta^{-1}\right) \delta_i p_i \theta_i T_{(0,1,-1)} S_w(t_i) \frac{(1-\delta_i) \left(T_{\eta(0,0,-1/\eta)}^{(i)} - T_{\eta(0,1,-1/\eta)}^{(i)}\right)}{1 + T_{(0,0,-1/\eta)}^{(i)} - T_{0,1,-1/\eta}^{(i)}}; \\ \frac{\partial \ell}{\partial \beta_{1j}} &= \sum_{i=1}^n p_{i\beta_{1j}} \left[p_i^{-1} - (\eta+1) \delta_i \theta_i S_w(t_i) T_{(0,1,-1)}^{(i)} \right] + \frac{(1-\delta_i) \theta_i p_{i\beta_{1j}}}{1 + T_{0,0,-1/\eta}^{(i)} - T_{-1/\eta}^{(i)}} \left(T_{(0,1,-1/\eta-1)}^{(i)} S_w(t_i) - T_{(0,0,-1/\eta-1)}^{(i)} \right); \\ \frac{\partial \ell}{\partial \beta_{2l}} &= \sum_{i=1}^n \theta_{i\beta_{2l}} \left[\theta_i^{-1} - (\eta+1) \delta_i p_i S_w(t_i) T_{(0,1,-1)}^{(i)} \right] + \frac{(1-\delta_i) p_i \theta_{i\beta_{2l}}}{1 + T_{0,0,-1/\eta}^{(i)} - T_{-1/\eta}^{(i)}} \left(T_{(0,1,-1/\eta-1)}^{(i)} S_w(t_i) - T_{(0,0,-1/\eta-1)}^{(i)} \right).\end{aligned}$$

Appendix D2: Matrix $J(\vartheta)$ of the DNB-LA model

Here, we derive the necessary formulas to obtain the second-order partial derivatives of the log-likelihood function $\ell_{DNB-LA}(\vartheta, \mathcal{D})$. The components of the matrix are given by

$$\begin{aligned}\ell_{\gamma_1 \gamma_1} &= \sum_{i=1}^n \delta_i \left(- \left(f_w(t_i)^{-1} f_{\gamma_1} \right)^2 + f_w(t_i)^{-1} f_{\gamma_1 \gamma_1} \right) - \delta_i (\eta+1) p_i \theta_i \left(T_{(0,1,-1)}^{(i)} S_{\gamma_1 \gamma_1} - p_i \eta \theta_i (S_{\gamma_1})^2 T_{(0,1,-2)}^{(i)} \right); \\ \ell_{\gamma_1 \gamma_2} &= \sum_{i=1}^n \delta_i f_{\gamma_1 \gamma_2} \left(- f_w(t_i)^{-2} + f_w(t_i)^{-1} \right) - \delta_i (\eta+1) p_i \theta_i \left(T_{(0,1,-1)}^{(i)} S_{\gamma_1 \gamma_2} - p_i \eta \theta_i S_{\gamma_1} S_{\gamma_2} T_{(0,1,-2)}^{(i)} \right); \\ \ell_{\gamma_1 \eta} &= \sum_{i=1}^n \delta_i p_i \theta_i S_{\gamma_1} T_{(0,1,-1)}^{(i)} \left[1 - (\eta+1) p_i \theta_i S_w(t_i) T_{(0,1,-1)}^{(i)} \right]; \\ \ell_{\gamma_1 \beta_{1j}} &= - \sum_{i=1}^n (\eta+1) \delta_i \theta_i p_{i\beta_{1j}} S_{\gamma_1} T_{(0,1,-1)}^{(i)} \left(1 - p_i \eta \theta_i S_w(t_i) T_{(0,1,-1)}^{(i)} \right) - (1-\delta_i) \theta_i p_{i\beta_{1j}}; \\ \ell_{\gamma_1 \beta_{2l}} &= - \sum_{i=1}^n (\eta+1) \delta_i p_i \theta_{i\beta_{2l}} S_{\gamma_1} T_{(0,1,-1)}^{(i)} \left(1 - p_i \eta \theta_i S_w(t_i) T_{(0,1,-1)}^{(i)} \right) - (1-\delta_i) p_i \theta_{i\beta_{2l}};\end{aligned}$$

where

$$L^{(i)} = T_{(0,0,-1/\eta)}^{(i)} \log T_{(0,0,1)}^{(i)} - T_{(0,1,-1/\eta)}^{(i)} \log T_{(0,1,1)}^{(i)} S_w(t_i).$$

References

- Berkson J, Gage RP. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*. 1952; 47(259):501–515.
- Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society B*. 1949; 11(1):15–53.
- Cancho V, Rodrigues J, de Castro M. A flexible model for survival data with a cure rate: a bayesian approach. *Journal of Applied Statistics*. 2011; 38(1):57–70.
- Chen M-H, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*. 1999; 94:909–919.
- Cooner F, Banerjee S, McBean AM. Modelling geographically referenced survival data with a cure fraction. *Statistical Methods in Medical Research*. 2006; 15:307–324. [PubMed: 16886733]
- Cooner F, Banerjee S, Carlin BP, Sinha D. Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*. 2007; 102:560–572. [PubMed: 21031152]
- Dunn PK, Smyth GK. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*. 1996; 5(3):236–244.
- Ibrahim, JG.; Chen, M-H.; Sinha, D. *Bayesian Survival Analysis*. Springer; New York: 2001.
- Johnson, LN.; Kemp, AW.; Kotz, S. *Univariate Discrete Distribution*. third edition. Wiley; New York, NY: 2005.
- Kirkwood JM, Ibrahim JG, Sondak VK, Richards J, Flaherty LE, Ernstoff MS, Smith TJ, Rao U, Steele M, Blum RH. High- and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial E1690/S9111/C9190. *Journal of Clinical Oncology*. 2000; 18:2444–2458. [PubMed: 10856105]
- Li CS, Taylor J, Sy J. Identifiability of cure models. *Statistics and Probability Letters*. 2001; 54:389–395.
- Maller, RA.; Zhou, X. *Survival Analysis with Long-Term Survivors*. Wiley; New York: 1996.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2010.
- Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*. 2005; 54(3):507–554.
- Rodrigues J, Cancho VG, de Castro M, Louzada-Neto F. On the unification of the long-term survival models. *Statistics & Probability Letters*. 2009; 79:753–759.
- Rodrigues J, de Castro M, Balakrishnan N, Cancho V. Destructive weighted Poisson cure rate models. *Lifetime Data Analysis*. 2010 DOI 10.1007/s10985-010-9189-2.
- Ross GJS, Preece DA. The negative binomial distribution. *The Statistician*. 1985; (3):323–336.
- Yakovlev, AY.; Tsodikov, AD. *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific; Singapore: 1996.
- Yang G, Chen C. A stochastic two-stage carcinogenesis model: a new approach to computing the probability of observing tumor in animal bioassays. *Mathematics Bioscience*. 1991; 104(1):247–258.

Figure 1. Surviving functions for the destructive negative binomial cure rate model (left panel) and non-cured surviving functions (right panel) for the Weibull distribution under different activations (First: dotted, Last: dashed and Random: solid).

Figure 2. Probability density (left panel) function and survival function (right panel) of the NBW distribution FA mechanism for some arbitrary parameter values.

Figure 3. Hazard function of the NBW distribution FA mechanism for some arbitrary parameter values.

Figure 4. Probability density (left panel) function and survival function (right panel) under LA mechanism for some arbitrary parameter values.

Figure 5. Hazard function of the NBW distribution under LA mechanism for some arbitrary parameter values.

Figure 6. Kaplan-Meier estimate of the surviving function of high-dose interferon and observation groups.

Figure 7.

Left Panel: QQ plot of the normalized randomized quantile residuals, with the identity line where each point corresponds to the median of 10 sets of ordered residuals ; Right panel: Probability of undestroyed causes stratified by nodule category and tumor thickness, for the DG-FA model.

Figure 8. Surviving function under the DG-FA model stratified by nodule category from 1 to 4 for median tumor thickness (3.175 mm) and age categories of (a) 38, (b) 47 and (c) 58 years, representing first, median and third quartiles.

Figure 9.

Non-cured surviving function under the DG-FA model stratified by nodule category from 1 to 4) for median tumor thickness (3.175 mm) and age categories of (a) 38, (b) 47 and (c) 58 years, representing first, median and third quartiles.

Table 1

Average of maximum likelihood estimates (AMLE), mean square error (MSE) and coverage probability (CP) of 95% CIs for the parameters in model (5).

<i>n</i>	Parameter	AMLE	MSE	CP
150	β_{10}	-0.330	2.648	0.928
	β_{11}	11.931	4.596	0.927
	β_{21}	0.787	0.313	0.937
	η	3.974	3.737	0.901
	γ_1	2.130	0.352	0.934
	γ_2	-3.482	0.777	0.930
300	β_{10}	-0.928	1.395	0.945
	β_{11}	0.722	1.314	0.931
	β_{21}	0.720	0.144	0.941
	η	3.529	2.934	0.956
	γ_1	2.074	0.268	0.933
	γ_2	-3.368	0.609	0.946
600	β_{10}	-1.036	0.565	0.959
	β_{11}	0.645	0.134	0.942
	β_{21}	0.730	0.051	0.954
	η	3.222	1.862	0.941
	γ_1	2.066	0.126	0.946
	γ_2	-3.045	0.346	0.951

Table 2

Proportion of times the AIC criterion selected the true underlying ‘data-generation’ model (DBN-FA/DBN-LA).

η	Sample Sizes		
	150	300	600
0 (Destructive Poisson)	0.957/0.963	0.978/0.975	0.986/0.984
1 (Destructive Geometric)	0.965/0.954	0.981/0.975	0.993/0.990
3 (Destructive BN)	0.964/0.967	0.983/0.979	0.995/0.996

Table 3

AIC and *SBC* values for the different candidate models: DNB-RA, DNB-FA and DNB-LA along with their special cases (Poisson and Geometric counterparts) for various activation schemes

Model	Activation scheme					
	FA			LA		
	AIC	SBC	AIC	SBC	AIC	SBC
DP	980.3	1008.3	1022.4	1050.5	1079.3	1107.4
DG	979.0	1007.0	1021.3	1049.3	1079.3	1107.4
DNB	980.5	1012.7	1024.2	1056.2	1081.3	1113.4

Table 4

Maximum likelihood (ML) estimates of the parameters for the DG-FA model with the covariate treatment.

DG-FA Model	Parameter	Estimate (est)	Standard error (se)	est / se
Complete Model	γ_1	1.267	0.072	-
	γ_2	-0.952	0.132	-
	$\beta_{1,\text{intercept}}$	-4.164	1.076	3.87
	$\beta_{1,\text{treatment}}$	-0.327	0.380	0.86
	$\beta_{1,\text{nodule}}$	1.583	0.486	3.26
	$\beta_{1,\text{thickness}}$	0.312	0.118	2.66
	$\beta_{2,\text{age}}$	0.024	0.004	6.26
Without 'Treatment' Covariate	γ_1	1.265	0.072	-
	γ_2	-0.956	0.133	-
	$\beta_{1,\text{intercept}}$	-4.542	1.105	4.11
	$\beta_{1,\text{nodule}}$	1.701	0.526	3.23
	$\beta_{1,\text{thickness}}$	0.340	0.128	2.66
	$\beta_{2,\text{age}}$	0.024	0.004	6.27

Table 5

Cure fraction (p_0) for various nodal categories stratified by age under the DG-FA model.

Age	Nodal Category	Mean	Standard Error	95% C.I.	
				LL	RL
38	1	0.735	0.057	0.622	0.848
	2	0.455	0.041	0.375	0.535
	3	0.326	0.028	0.271	0.245
	4	0.296	0.026	0.246	0.346
47	1	0.691	0.064	0.565	0.817
	2	0.403	0.039	0.327	0.478
	3	0.281	0.029	0.223	0.339
	4	0.253	0.030	0.195	0.311
58	1	0.637	0.072	0.494	0.778
	2	0.345	0.037	0.272	0.481
	3	0.233	0.031	0.173	0.294
	4	0.209	0.032	0.146	0.272