SORT 40 (1) January-June 2016, 31-54

Two alternative estimation procedures for the negative binomial cure rate model with a latent activation scheme

Diego I. Gallardo[1] and Heleno Bolfarine[2]

Abstract

In this paper two alternative estimation procedures based on the EM algorithm are proposed for the flexible negative binomial cure rate model with a latent activation scheme. The Weibull model as well as the log-normal and gamma distributions are also considered for the time-to-event data for the non-destroyed cells. Simulation studies show the satisfactory performance of the proposed methodology. The impact of misspecifying the survival function on both components of the model (cured and susceptible) is also evaluated. The use of the new methodology is illustrated with a real data set related to a clinical trial on Phase III cutaneous melanoma patients.

MSC: 62N01, 62N02, 62P10.

Keywords: Competing risks, EM algorithm, latent activation scheme.

1. Introduction

An implicit assumption with the ordinary survival model is that all individuals under study are susceptible to the event of interest, which is not always true given the improvements in disease treatments experienced in the last decades. For some types of cancer, for example, new treatments have significantly increased the probability that an individual is considered with the disease under control (typically called cured). The proportion of cured individuals after a treatment is usually known as the cure fraction.

[1] Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile. diego.gallardo@uantof.cl

[2] Departamento de Estatística, IME, Universidade de São Paulo, São Paulo, Brasil. hbolfar@ime.usp.br

Received: May 2015

Accepted: February 2016

Berkson and Gage (1952) developed a model that became known in the literature as the *mixture model*, which assumes that there is a proportion $1-q_0$ of susceptible individuals and, hence, a proportion q_0 of cured individuals. An alternative route was pursued by Yakovlev and Tsodikov (1996) and Chen et al. (1999). Their approach is based on the assumption that each individual has an unobserved (latent) number M of cells, each capable of triggering the event of interest. This model is known in the literature as the *promotion time cure rate model* and has been the subject of intense research activity. Rodrigues et al. (2009) unify the two approaches considering the negative binomial distribution for the variable M, known in the literature as the *negative binomial cure rate model*. Those models have a common element: both assume that the initial cells will produce the event of interest. In order to relax this assumption, Rodrigues et al. (2012) proposed the so-called destructive weighted Poisson cure rate model in which it is assumed that each one of the initial cells has a probability p of being able to produce the patient's death, so that only $D \leq M$ cells (usually called activated or non-destroyed cells) would remain in effect. Clearly, the case $p = 1$ (i.e., $M = D$) leads to the standard models above. Both destructive and non-destructive models mentioned above assume that one cell is sufficient to produce the event of interest, i.e., the time until the event occurs is considered as the minimum of the times related to each activated cell. This scheme is known as the first activation (FA) scheme.

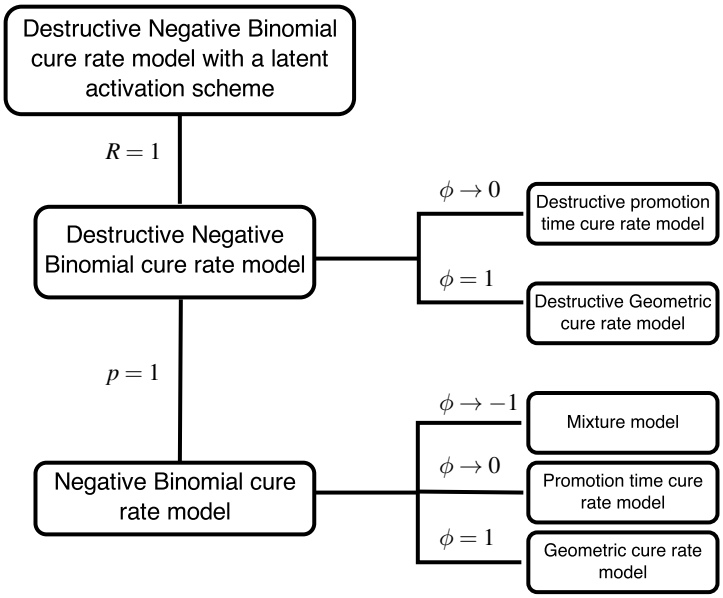

Figure 1: *Summary of some particular cases of the DNB model with a latent activation scheme.*

Cooner et al. (2007) proposed a more general activation scheme in a non-destructive context. This idea was used by Cancho et al. (2013) in the Destructive Negative Bino-

mial (DNB) cure rate model, where the negative binomial distribution with mean θ and variance $\theta(1+\phi\theta)$ is used for the initial number of cells. Accordingly, $\phi > 0 (\phi < 0)$ provides over-dispersion (sub-dispersion), including the Poisson model as particular case for $\phi = 0$. The idea is that the event of interest may be considered as the maximum of the times related to each one of the concurrent cells, i.e., all cells must be activated to produce the event of interest. This scheme is called last activation (LA) scheme. A third activation scheme is proposed assuming that a random number of factors (R) is needed to produce the event of interest, i.e., the time to the event of interest is defined as the R-th order statistics from the times related to the activated cells. A simple specification is to assume the discrete uniform distribution for R on the set $\{1,\dots,D\}$. This scheme is known as the random activation (RA) scheme. Figure 1 depicts a summary of the DNB in Cancho et al. (2013) and some particular cases of the model.

The main focus of this work is to develop two different ways of applying the EM algorithm for maximum likelihood estimation (MLE) for the DNB with different activation schemes. The first way is to compute directly the expected value of M and D, the number of initial and activated cells, respectively, and the second way is to write the model as a *mixture model* and to use the EM algorithm for this alternative version Lu (2010).

The paper is organized as follows. In Section 2 we describe the cutaneous melanoma data set. In Section 3, the DNB model with different activation schemes and some propositions about this model are stated. In Section 4, two estimation procedures based on the EM algorithm are proposed for the model in Section 3. Section 5 reveals results of two simulation studies aiming at investigating parameters recovery and assessing the time-to-event for the non-destroyed cells. Section 6 presents an application to a real data set referring to a clinical trial for patients with melanoma. Finally, in Section 7, the main conclusions and results obtained in this work are presented.

2. Cutaneous melanoma data set

The data set is related to a clinical trial on a Phase III cutaneous melanoma patients available at `http://merlot.stat.uconn.edu/~mhchen/survbook/`, labeled as E1690 data. The clinical trial was conducted by the Eastern Cooperative Oncology Group (see Ibrahim et al. (2001) for details). The incidence of melanoma is one of the highest among most types of cancer, with a high mortality rate even with early detection. The objective of this study was to evaluate a postoperative treatment performance with a high dose of the drug Interferon alpha-2b, in order to prevent recurrence. The study included patients between 1991 to 1995 and follow-up was conducted until 1998.

A characteristic of the disease (as in many other types of cancers) is the presence of a proportion of patients that can lead a normal life, comparable to patients without the disease. In other words, a proportion commonly known as "cured". After deleting patients with incomplete data and missing observation times, the data set is composed of

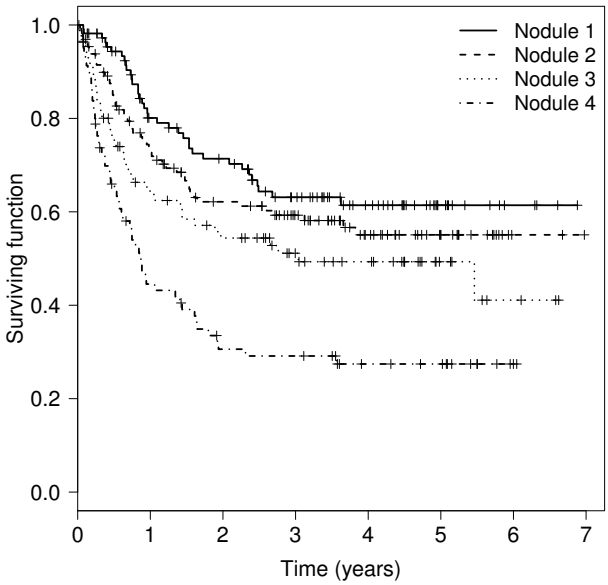

Figure 2: *Kaplan-Meier curves stratified by covariate Nodule.*

$n = 408$ individuals. The collected variables were: Observed time (in years, average = 2.31, median = 1.64, standard deviation = 1.93), treatment (0: control and 1: interferon alfa-2b with 198 and 210 patients respectively), age (in years, average = 48.1, median = 47.2 and standard deviation = 13.1), nodal category (categorical variable with levels 1-4 with 110, 131, 86 and 81 patients in each group, respectively, where 1 indicates the lower risk patients and 4 the higher risk patients) and tumour thickness (in mm, average = 3.98, median = 3.18 and standard deviation = 3.22).

Figure 2 depicts the Kaplan-Meier curves by nodule category, confirming a well pronounced plateau in all nodule categories. In the next Section, we present the model addressed for this particular problem.

3. Model specification

Following Cancho et al. (2013), let M be an unobservable random variable denoting the initial number of competing causes related to the occurrence of the event of interest. For the cutaneous melanoma data set, M represents the number of carcinogenic cells. Assume that M has negative binomial distribution with probability mass function (p.m.f.) given by

$$P(M = m;\theta,\phi) = \frac{\Gamma(\phi^{-1}+m)}{\Gamma(\phi^{-1})m!}\left(\frac{\phi\theta}{1+\phi\theta}\right)^{m}(1+\phi\theta)^{-1/\phi}, \quad m = 0,1,2,\ldots, \qquad (1)$$

where $\theta > 0$, $\phi \geq -1$ and $1 + \phi\theta > 0$. The distribution in (1) is denoted as $M \sim NB\left(\phi, \frac{\phi\theta}{1+\phi\theta}\right)$. Under this parametrization, $\mathbb{E}(M) = \theta$ and $\text{Var}(M) = \theta(1+\phi\theta)$. For this reason, $\phi > 0$ ($\phi < 0$) corresponds to over (under)-dispersion in relation to the Poisson distribution. For $\phi \to 0$, the p.m.f. in (1) is reduced to the p.m.f. of the Poisson distribution and $\phi = 1$ corresponds to the geometric distribution with parameter $1/(1+\theta)$.

Let ζ_j, $j = 1, \ldots, M$ be (conditionally) independent random variables given $M = m$, with Bernoulli distribution and success probability p indicating whether the j-th concurrent cause can produce or not the event. Contextualizing to the medical problem under study, $\zeta_j = 1$ ($\zeta_j = 0$) indicates that the j-th carcinogenic cell was (was not) activated or non-destroyed (destroyed), and each activated carcinogenic cell can produce the metastasis process. The (unobservable) total damaged D is defined as

$$D = \begin{cases} \zeta_1 + \ldots + \zeta_M & \text{, if } M > 0, \\ 0 & \text{, if } M = 0. \end{cases}$$

Note that D represents the total number of activated carcinogenic cells (among the M initials) which are activated. It is immediate that $D \mid M = m \sim Bin(m, p)$ for $m > 0$ and $P(D = 0 \mid M = 0) = 1$. Moreover, it is possible to show that $D \sim NB\left(\phi, \frac{\phi\theta p}{1+\phi\theta p}\right)$ Rodrigues et al. (2011). Define W_j, $j = 1, \ldots, D$ as the time to event for the j-th activated cell produces the metastasis process. Assume that W_j, $j = 1, \ldots, D$, are conditionally independent and identically distributed (*i.i.d.*) given D with common cumulative distribution function $F(\cdot;\lambda)$ and survival function $S(\cdot;\lambda) = 1 - F(\cdot;\lambda)$. Further, assume that $W_1, W_2, \ldots$, are independent of D and M. As discussed in Cooner et al. (2007), cure rate models with latent activation schemes assume that the failure time T^* is generated by the activation times of D latent factors. Thus, $D = 0$ implies $T^* = \infty$ and then the individual is considered cured. If $D > 0$ and it is assumed that R among the D cells are required to produce the event of interest, so the failure time to event is defined by $T^* = W_{(R)}$, where R depends (or not) on D and $W_{(R)}$ denotes the R-th order statistics corresponding to $W_1, \ldots, W_D$.

Assume that the data can be censored to the right. Thus, the observed data can be represented by $T = \min(T^*, C)$ and $\delta = I(T^* \leq C)$, with T^* and C denoting failure and censoring times, respectively, and $I(\cdot)$ the indicator function. Under this scheme and following similar arguments in Cooner et al. (2007), we can write the joint distribution of (T, δ, R, M, D) as

$$\begin{aligned} f(t,\delta,r,m,d;\theta,\phi,p,\lambda) =& f(t,\delta \mid D = d, R = r, \lambda) P(R = r \mid D = d) \times \\ & \times P(D = d \mid M = m; p) P(M = m, \phi, \theta), \end{aligned} \tag{2}$$

where $D \mid M = m; p \sim Bin(m, p)$, $P(M = m; \theta, \phi)$ is given in (1) and

$$f(t,\delta \mid D=d,R=r,\lambda) = \Big\{I(d=0)+I(m\geq d\geq r\geq 1)IB(S(t;\lambda),d-r+1,r)\Big\}^{1-\delta} \times \left\{d\binom{d-1}{r-1}f(t;\lambda)S(t;\lambda)^{d-r}F(t;\lambda)^{r-1}\right\}^{\delta} \quad (3)$$

with $IB(z,a,b)$ denoting the incomplete beta function defined as $IB(z,a,b)=\int_0^z u^{a-1}(1-u)^{b-1}du$. The population survival and density functions can be computed as

$$S_{pop}(t;\theta,\phi,p,\lambda) = P(D=0;\theta,\phi,p)+\sum_{m=1}^{\infty}\sum_{d=1}^{m}\sum_{r=1}^{d} f(t,\delta=0,r,m,d;\theta,\phi,p,\lambda)$$

$$f_{pop}(t;\theta,\phi,p,\lambda) = \sum_{m=1}^{\infty}\sum_{d=1}^{m}\sum_{r=1}^{d} f(t,\delta=1,r,m,d;\theta,\phi,p,\lambda)$$

It is immediate that $q_0 = S_{pop}(\infty;\theta,\phi,p,\lambda) = (1+\phi\theta p)^{-1/\phi}$, so that the cure rate does not depend on the choice of the (conditional) distribution of $R \mid D=d$.

Moreover, to contour the identifiability problems in the sense of Li et al. (2001) and Hanin and Huang (2014) and discussed in Rodrigues et al. (2011) in the context of the destructive weighted Poisson cure rate models, it is necessary to introduce a set of covariates z_{1i} (of dimension r_1) associated with the initial number of cells and z_{2i} (of dimension r_2) related to the activation probabilities for non-destroyed cells by

$$\log\theta_i = z_{1i}^{\top}\beta_1 \quad \text{and} \quad \log\left(\frac{p_i}{1-p_i}\right) = z_{2i}^{\top}\beta_2, \quad i=1,\ldots,n. \quad (4)$$

In addition, z_1 and z_2 shall not simultaneously include intercepts nor share common elements. Henceforth, in order to simplify the notation, define $\psi=(\beta_1,\beta_2,\phi,\lambda)$ as the vector of parameters to be estimated. Three typically used activation schemes are the random activation scheme (RA), first activation scheme (FA) and last activation scheme (LA), for which the p.m.f. for the conditional distribution $P(R=r \mid D=d)$ and the population survival function for DNB are given in Table 1. Those models are denoted by DNB-FA, DNB-LA and DNB-RA, respectively.

Table 1: *Conditional distribution of R given D = d for three activation schemes with DNB.*

Activation scheme	$P(R=r \mid D=d)$	$S_{pop}(t;\psi)$
RA	$\frac{1}{d}I(1\leq r\leq d)$	$q_0+\{1-q_0\}S(t;\lambda).$
FA	$I(r=1)$	$\{1+\phi\theta pF(t;\lambda)\}^{-1/\phi}$
LA	$I(r=d)$	$1+q_0-\{1+\phi\theta pS(t;\lambda)\}^{-1/\phi}$

Under the usual assumptions in survival analysis and right censoring (see Williams and Lagarkos, 1977), the contribution to the (observed) log-likelihood by the i-th individual is given by

$$f(t_i,\delta_i;\psi) = f_{pop}(t_i;\psi)^{\delta_i} S_{pop}(t_i;\psi)^{1-\delta_i}. \tag{5}$$

Based on (2) and (5), the following propositions are now stated.

Proposition 1 *For combinations DNB-FA and DNB-LA it follows that, given D_{obs}, the conditional distribution of R_i degenerates in the distribution of $R_i = 1$ and $R_i = D_i$ respectively. For the combination DNB-RA, that distribution is*

$$P(R_i = r_i \mid D_{obs};\psi) = \begin{cases} \dfrac{\sum_{k=0}^{r_i-1}\binom{r_i-1}{k}(-1)^k \mathbb{E}\left[\frac{S(t_i;\lambda)^{D_i-r_i+k+1}}{D_i(D_i-r_i+k+1)} I(D_i\geq r_i)\right]}{q_{0i}+(1-q_{0i})S(t_i;\lambda)} & , \text{if } \delta_i = 0 \\ \dfrac{F(t_i;\lambda)^{r_i-1}\mathbb{E}\left[\binom{D_i-1}{r_i-1} S(t_i;\lambda)^{D_i-r_i} I(D_i\geq r_i)\right]}{1-q_{0i}} & , \text{if } \delta_i = 1, \end{cases}$$

where $D_i \sim NB\left(\phi, \frac{\phi\theta_i p_i}{1+\phi\theta_i p_i}\right)$, and $r_i = 1,2,\ldots$.

Proof of proposition 1 is presented in the Appendix A.

Proposition 2 *For DNB in (2) and FA and LA schemes in Table 1, $P(D_i = d_i \mid D_{obs};\psi)$, $i = 1,\ldots,n$, have a closed form. Moreover, for the model DNB-FA,*

$$D_i - \delta_i \mid D_{obs};\psi \sim NB\left((\phi^{-1}+\delta_i)^{-1}, \frac{\phi\theta_i p_i S(t_i;\lambda)}{1+\phi\theta_i p_i}\right),$$

and for the DNB-LA

$$D_i - \delta_i \mid D_{obs};\psi \sim \begin{cases} NB\left((\phi^{-1}+1)^{-1}, \frac{\phi\theta_i p_i F(t_i;\lambda)}{1+\phi\theta_i p_i}\right) & , \text{if } \delta_i = 1, \\ a_i NB\left(\phi, \frac{\phi\theta_i p_i}{1+\phi\theta_i p_i}\right) + (1-a_i) NB\left(\phi, \frac{\phi\theta_i p_i F(t_i;\lambda)}{1+\phi\theta_i p_i}\right) & , \text{if } \delta_i = 0. \end{cases}$$

where $a_i = [1+q_{0i}-(1+\phi\theta_i p_i S(t_i;\lambda))^{-\phi^{-1}}]^{-1}$. For the DNB-RA combination, the conditional distribution is

$$P(D_i = d_i \mid D_{obs};\psi) = \begin{cases} \frac{\sum_{r_i=1}^{d_i}\sum_{k=0}^{r_i}(-1)^k\binom{r_i-1}{k}\frac{S(t_i;\lambda)^{d_i-r_i+k+1}}{d_i(d_i-r_i+k+1)}\frac{\Gamma(\phi^{-1}+d_i)}{\Gamma(\phi^{-1})d_i!}\left(\frac{\phi\theta_i p_i}{1+\phi\theta_i p_i}\right)^{d_i}}{1+[(1-q_{0i})/q_{0i}]S(t_i;\lambda)} & , \text{if } \delta_i = 0 \\ \frac{\Gamma(\phi^{-1}+d_i)}{\Gamma(\phi^{-1})d_i!}\left(\frac{\phi\theta_i p_i}{1+\phi\theta_i p_i}\right)^{d_i} I(d_i \geq 1) & , \text{if } \delta_i = 1, \end{cases}$$

Proposition 2 is proved in Appendix B.

Proposition 3 *For DNB in (2) and FA and LA schemes in Table 1, $P(M_i = m_i \mid D_{obs};\psi)$, $i = 1,\ldots,n$, have a closed form. Moreover, for the DNB-FA combination, we have*

$$M_i - \delta_i; D_{obs}, \psi \sim NB\left((\phi^{-1}+\delta_i)^{-1}, \frac{\phi\theta_i(1-p_iF(t_i;\lambda))}{1+\phi\theta_i}\right),$$

and for the DNB-LA,

$$M_i - \delta_i \mid D_{obs};\psi \sim \begin{cases} NB\left((\phi^{-1}+1)^{-1}, \frac{\phi\theta_i(1-p_iS(t_i;\lambda))}{1+\phi\theta_i}\right) & , \text{if } \delta_i = 1, \\ a_iNB\left(\phi, \frac{\phi\theta_i}{1+\phi\theta_i}\right) + (1-a_i)NB\left(\phi, \frac{\phi\theta_i(1-p_iS(t_i;\lambda))}{1+\phi\theta_i}\right) & , \text{if } \delta_i = 0. \end{cases}$$

where $a_i = [1+q_{0i}-(1+\phi\theta_i p_i S(t_i;\lambda))^{-\phi^{-1}}]^{-1}$. For the DNB-RA and $\delta_i = 0$ this conditional distribution is

$$P(M_i = m_i \mid D_{obs},\psi) = \frac{\sum_{d_i=0}^{m_i}\sum_{r_i=1}^{d_i}\sum_{k=0}^{r_i} v_i\left(\frac{p_i}{1-p_i}\right)^{d_i}\left(\frac{\phi\theta_i(1-p_i)}{1+\phi\theta_i}\right)^{m_i}}{1+[(1-q_{0i})/q_{0i}]S(t_i;\lambda)},$$

where $v_i = (-1)^k\binom{r_i-1}{k}\frac{S(t_i;\lambda)^{d_i-r_i+k+1}}{d_i(d_i-r_i+k+1)}\frac{\Gamma(\phi^{-1}+m_i)}{\Gamma(\phi^{-1})d_i!(m_i-d_i)!}$. On the other hand, for $\delta_i = 1$ we have that

$$P(M_i = m_i \mid D_{obs},\psi) = \frac{[1-(1-p_i)^{m_i}]\frac{\Gamma(\phi^{-1}+m_i)}{\Gamma(\phi^{-1})m_i!}\left(\frac{\phi\theta_i}{1+\phi\theta_i}\right)^{m_i}(1+\phi\theta_i)^{-1/\phi}I(m_i \geq 1)}{1-q_{0i}},$$

Proof of proposition 3 is presented in Appendix C.

Propositions 1-3 are very useful because they allow predicting the initial number of cells, the number of non-destroyed cells and the number of cells necessary to produce the event of interest in each individual. Moreover, they are useful in implementing the EM algorithm, to be discussed now.

Note that the complete log-likelihood function is given by

$$\ell(\psi \mid D_{comp}) = \sum_{i=1}^{n} f(t_i, \delta_i, R_i, M_i, D_i; \psi), \tag{6}$$

with $f(t_i, \delta_i, r_i, m_i, d_i; \psi)$ defined in (2). Specifically, for the DNB-FA the expression in (6), unless to a constant, assumes the form

$$\ell(\psi \mid D_{comp}) = \sum_{i=1}^{n} \Bigg[(D_i - \delta_i) \log S(t_i; \lambda) + \delta_i \log f(t_i; \lambda) + D_i \log(p_i) + M_i \log \theta_i + (M_i - D_i) \log(1 - p_i) + (M_i - \phi^{-1}) \log(1 + \phi\theta_i) \Bigg]. \tag{7}$$

From (7), it is simple to deduce that it is only necessary the expectations of M_i and D_i (given D_{obs}) to implement the E-step of the EM algorithm. Using Propositions 2 and 3, these expectations are

$$\mathbb{E}(M_i \mid D_{obs}; \beta_1, \beta_2, \phi, \lambda) = \delta_i + \frac{(1 + \phi\delta_i)\theta_i(1 - p_i F(t_i; \lambda))}{1 + \phi\theta_i p_i F(t_i; \lambda)} \quad \text{and} \tag{8}$$

$$\mathbb{E}(D_i \mid D_{obs}; \beta_1, \beta_2, \phi, \lambda) = \delta_i + \frac{(1 + \phi\delta_i)\theta_i p_i S(t_i; \lambda)}{1 + \phi\theta_i p_i F(t_i; \lambda)}. \tag{9}$$

On the other hand, the expression (6) for the DNB-LA assumes the form

$$\ell(\psi \mid D_{comp}) = \sum_{i=1}^{n} \Bigg[(1 - \delta_i) \log\left(1 - I(D_i \geq 1) F(t_i; \lambda)^{D_i}\right) + \delta_i \Big(\log D_i + \log f(t_i; \lambda) + (D_i - 1) \log F(t_i; \lambda) \Big) + D_i \log p_i + (M_i - D_i) \log(1 - p_i) + M_i \log \theta_i + (M_i - \phi^{-1}) \log(1 + \phi\theta_i) \Bigg]. \tag{10}$$

However, the expectation of $\log\left(1 - I(D_i \geq 1) F(t_i; \lambda)^{D_i}\right)$ does not have a closed form, hindering the application of the EM algorithm in this way. Finally, using a RA scheme the log-likelihood function of the model is even more complex, making it difficult the implementation of the EM algorithm in this form. For this reason, a second way is proposed to perform the estimation procedure in those models.

Following Tsodikov et al. (2003) and Rodrigues et al. (2009), all cure rate models can be expressed as a mixture model, i.e.,

$$S_{pop}(t;\psi) = q_0 + (1-q_0)S^*(t;\psi), \tag{11}$$

where $S^*(t;\psi)$ represents the survival function for susceptible individuals and q_0 is the cure rate. Table 2 presents this function for the three activation schemes considered in this work.

Table 2: *Survival and hazard functions for susceptible individuals for the DNB mixture model with three activation schemes.*

Act. Scheme	RA	FA	LA
$S^*(t;\psi)$	$S(t;\lambda)$	$\frac{(1+\phi\theta pF(t;\lambda))^{-1/\phi} - q_0}{1-q_0}$	$\frac{1-(1+\phi\theta pS(t;\lambda))^{-1/\phi}}{1-q_0}$
$h^*(t;\psi)$	$h(t;\lambda)$	$\frac{\theta pf(t;\lambda)(1+\phi\theta pF(t;\lambda))^{-1/\phi-1}}{(1+\phi\theta pF(t;\lambda))^{-1/\phi} - q_0}$	$\frac{\theta pf(t;\lambda)(1+\phi\theta pS(t;\lambda))^{-1/\phi-1}}{1-(1+\phi\theta pS(t;\lambda))^{-1/\phi}}$

Let Y_i the binary variable that indicates whether the individual is susceptible or cured ($Y_i = 1$ and $Y_i = 0$, respectively). Following Lu (2010), the complete log-likelihood function for this model is

$$\ell_c(\psi) = \sum_{i=1}^{n} \left[Y_i \log(1-q_{0i}) + (1-Y_i)\log q_{0i} + Y_i \log S^*(t_i;\psi) + \delta_i Y_i \log h^*(t_i;\psi) \right], \tag{12}$$

and the expected value for Y_i given D_{obs} is

$$\mathbb{E}(Y_i \mid D_{obs};\psi) = \delta_i + (1-\delta_i)\frac{(1-q_{0i})S^*(t_i;\psi)}{q_{0i} + (1-q_{0i})S^*(t_i;\psi)}. \tag{13}$$

Equations (12) and (13) provides a second way to implement the EM algorithm in any cure rate model, in particular, for the DNB with different activation schemes.

4. Estimation

In this Section it is discussed some inferential procedures for the parameters of the DNB with the three activation schemes discussed in Section 3. Parameter estimation is approached using the maximum likelihood method.

In Cancho et al. (2013), the estimation procedure was based on the direct maximization of the observed likelihood function given by

$$\ell(\psi \mid D_{obs}) = \sum_{i=1}^{n} \Big[\log S_{pop}(t_i;\psi) + \delta_i \log h_{pop}(t_i;\psi) \Big], \tag{14}$$

where $S_{pop}(\cdot)$ and $h_{pop}(\cdot)$ depend on the activation scheme used in Table 1. However, maximization of (14) is not simple because it is a function that involves all parameters.

The EM algorithm Dempster et al. (1977) is a very popular maximization alternative used to obtain the maximum likelihood estimators when the model has missing data. A further discussion about the EM algorithm in comparison with the direct maximization of the log-likelihood function is performed in MacDonald (2014). In the cure rate context, we found many recent works using this algorithm. For instance, Balakrishnan and Pal (2012, 2013, 2015) and Gallardo et al. (2016). Two different ways of applying this algorithm in the model considered will be presented in next subsection.

4.1. EM algorithm: implementation 1

Consider initially only the combination DNB-FA, i.e., $R = 1$. Moreover, it is assumed that ϕ is fixed. The first way to apply the EM algorithm in this model is to compute the expected values for M_i and D_i, $i = 1,\dots,n$ given D_{obs} and the parameters values in last iteration, namely $\psi^{(k-1)}$. Those values are denoted by $\widetilde{D}_i^{(k)}$ and $\widetilde{M}_i^{(k)}$, respectively, and they can be computed using equations (8) and (9). Then, it is necessary to replace those values in the complete log-likelihood function given in (7) and maximize it in relation to ψ. The algorithm is summarized as follows.

- **E-step**: For $i = 1,\dots,n$, compute

$$\widetilde{D}_i^{(k)} = \delta_i + \frac{(1+\phi\delta_i)\theta_i^{(k-1)} p_i^{(k-1)} S(t_i;\lambda^{(k-1)})}{1+\phi\theta_i^{(k-1)} p_i^{(k-1)} F(t_i;\lambda^{(k-1)})} \quad \text{and}$$

$$\widetilde{M}_i^{(k)} = \delta_i + \frac{(1+\phi\delta_i)\theta_i^{(k-1)} (1-p_i^{(k-1)} F(t_i;\lambda^{(k-1)}))}{1+\phi\theta_i^{(k-1)} p_i^{(k-1)} F(t_i;\lambda^{(k-1)})}.$$

- **M-step**: Given $\widetilde{D}^{(k)} = (\widetilde{D}_1^{(k)},\dots,\widetilde{D}_n^{(k)})$ and $\widetilde{M}^{(k)} = (\widetilde{M}_1^{(k)},\dots,\widetilde{M}_n^{(k)})$, find $\beta_1^{(k)}$, $\beta_2^{(k)}$ and $\lambda^{(k)}$ that maximize $Q_1(\beta_1 \mid \psi^k)$, $Q_2(\beta_2 \mid \psi^k)$ and $Q_3(\lambda \mid \psi^k)$ in relation to β_1, β_2 and λ, respectively, where

$$Q_1(\beta_1 \mid \psi^{(k)}) = \sum_{i=1}^{n} \left\{ \widetilde{M}_i^{(k)} \log\theta_i - \theta_i + \left(\widetilde{M}_i^{(k)} - \phi^{-1}\right) \log(1+\phi\theta_i) \right\}, \tag{15}$$

$$Q_2(\beta_2 \mid \psi^{(k)}) = \sum_{i=1}^{n} \left\{ \widetilde{D}_i^{(k)} \log(p_i) + \left(\widetilde{M}_i^{(k)} - \widetilde{D}_i^{(k)}\right) \log(1-p_i) \right\}, \tag{16}$$

$$Q_3(\lambda \mid \psi^{(k)}) = \sum_{i=1}^{n} \left\{ \left(\widetilde{D}_i^{(k)} - \delta_i\right) \log S(t_i;\lambda) + \delta_i \log f(t_i;\lambda) \right\}. \tag{17}$$

Then, define $\psi^{(k)} = \left(\beta_1^{(k)}, \beta_2^{(k)}, \lambda^{(k)}\right)$. The advantage of this approach is that functions in (15), (16) and (17) can be maximized separately with respect to β_1, β_2 and λ, respectively, instead of the joint maximization as occurs with the observed log-likelihood. Steps M and E are repeated until a suitable convergence rule is satisfied. For instance, $||\psi^{(k)} - \psi^{(k-1)}|| < \epsilon$, where $||\psi^{(k)} - \psi^{(k-1)}||$ represents the euclidian distance between $\psi^{(k)}$ and $\psi^{(k-1)}$ and ϵ is a prefixed value. For instance, we use $\epsilon = 0.0001$.

4.2. EM algorithm: implementation 2

For this approach, three activation schemes are considered in Table 1. As discussed in Section 3, models DBN-FA, DBN-LA and DBN-RA can be expressed as the mixture model with survival function for susceptible individuals given by $S^*(\cdot \mid \psi)$, according to Table 2, and cure rate given by $q_{0i} = (1 + \phi\theta_i p_i)^{-1/\phi}$ that is common for the three models.

Proceeding similarly as in the last procedure, the algorithm is summarized next.

- **E-step**: For $i = 1, \ldots, n$, compute

$$\widetilde{Y}_i^{(k)} = \delta_i + (1 - \delta_i)\frac{(1 - q_{0i}^{(k-1)})S^*\left(t_i; \psi^{(k-1)}\right)}{q_{0i}^{(k-1)} + (1 - q_{0i}^{(k-1)})S^*\left(t_i; \psi^{(k-1)}\right)}.$$

- **M-step**: Given $\widetilde{Y}^{(k)} = (\widetilde{Y}_1^{(k)}, \ldots, \widetilde{Y}_n^{(k)})$, find $\psi^{(k)}$ that maximizes

$$Q(\psi) = \sum_{i=1}^{n} \left[\widetilde{Y}_i^{(k)} \log(1 - q_{0i}^{(k)}) + (1 - \widetilde{Y}_i^{(k)}) \log q_{0i}^{(k)} + \widetilde{Y}_i^{(k)} \log S^*(t_i; \psi^{(k)}) + \delta_i \widetilde{Y}_i^{(k)} \log h^*(t_i; \psi^{(k)}) \right].$$

Then, steps M and E are repeated until a suitable convergence rule is satisfied. The advantage of this approach in relation to directly maximizing the observed log-likelihood in (14) is that the latent variables Y_i are completely observed for individuals with failure times because $\delta_i = 1$ implies $Y_i = 1$ (i.e., a failure time guarantees that the individual is susceptible). This information is lost when an approach based on the observed log-likelihood is used because the vector $Y = (Y_1, \ldots, Y_n)$ is removed when summing over $\{0,1\}^n$. Consequently, implementing the M-step, for fixed β_1 and β_2, which consists in maximizing the function $Q(\cdot)$ with respect to λ is easier than maximizing the observed log-likelihood function in (14). Thus, it seems more advantageous to use the EM algorithm over than a direct maximization of the observed log-likelihood function. Note

Table 3: *Distributions used for modelling the survival function of the non-destroyed cells.*

Distribution	$S(w;\lambda)$	$f(w;\lambda)$
Weibull	$\exp(-e^{\alpha}w^{\nu})$	$\nu w^{\nu-1}\exp(\alpha-e^{\alpha}w^{\nu})$
LN	$1-\Phi\left(\frac{\log(w)-\alpha}{\nu}\right)$	$\frac{1}{\nu w}\phi\left(\frac{\log(w)-\alpha}{\nu}\right)$
Gamma	$1-\frac{\gamma(\alpha,\nu w)}{\Gamma(\alpha)}$	$\frac{\nu^{\alpha}}{\Gamma(\alpha)}w^{\alpha-1}e^{-\nu w}$

NOTE: $\phi(\cdot)$ and $\Phi(\cdot)$ represent the density and the cumulative function of standard normal distribution. $\gamma(\cdot,\cdot)$ represents the lower incomplete gamma function.

that the EM procedures does not depend on a specific survival function considered for non-destroyed cells. In this work, it is used the Weibull, log-normal (LN) and gamma distributions with parametrizations in Table 3, where $\lambda=(\alpha,\nu)$.

Henceforth, the distribution of $S(\cdot \mid \lambda)$ will be specified jointly with the activation mechanism. For instance, DNB-FA/Weibull, DNB-LA/LN, DNB-RA/gamma, etc. Note that the asymptotic variances for the MLEs could be estimated using the inverse of the Hessian matrix (matrix of second derivatives of the log-likelihood function). The observed information matrix is then obtained from the Hessian matrix evaluated in the MLEs. The elements of the Hessian matrix are presented in the Appendix of Cancho et al. (2013) with the Weibull model considered for the times of the non-destroyed cells. Expressions relatives for the LN and gamma models will not be presented because they are slight modifications for the Weibull model.

Remark 1

1. *In the first version of the EM algorithm, it is assumed that ϕ is fixed. However, it is possible to relax this assumption by constructing a profile log-likelihood for ϕ and picking the value that maximize that function. On the other hand, the standard error for the estimator of ϕ can be estimated via Jackknife (Miller, 1974).*
2. *To avoid maximization problems with the constraint $1+\phi\theta>0$ (presented after eq. (1)), we use the same approach used by Cancho et al. (2013) considering $\phi\geq 0$, i.e., the over-dispersed case.*
3. *The maximization involved in the M-steps can be performed using software R (*R Development Core Team, 2015*), among others. The computational programs used in this work are available from the authors upon request.*
4. *Differently from the direct maximization of the log-likelihood function, the EM algorithm allows to obtain predictions for the number of initial cells and activated cells for each individual (M_i and D_i, $i=1,\ldots,n$, respectively) in the version 1 and to the chance for cure for each individual (Y_i, $i=1,\ldots,n$), in the version 2.*

5. Simulation studies

In this Section, two simulation studies are presented. The first study assess the performance of the two procedures through different elements as bias and coverage probabilities. The second study is designed to evaluate whether the AIC and BIC (Akaike's and Bayesian information) criteria are able to correctly pick the distribution for the non-destroyed cells, given the correct activation scheme.

5.1. Parameters recovery

For simulation purposes, the covariates z_1 and z_2 were drawn from the Bernoulli distribution with success probability 0.5. As discussed in Section 3, both vectors should not

Table 4: *Average of parameter estimates, standard errors (se), root of mean squared errors ($\sqrt{MSE}$) and coverage probability of 95% (CP) using the way 2 of defining the EM algorithm for DNB-FA, DNB-LA and DNB-RA models considering Weibull distribution for time-to-event of the non-destroyed cells. (CA denotes censoring average with their respective standard errors).*

DNB-FA									
		$n = 200$				$n = 400$			
Parameter	True	average	se	$\sqrt{MSE}$	CP	average	se	$\sqrt{MSE}$	CP
β_1	1.0	1.021	0.287	0.240	0.943	1.008	0.202	0.165	0.943
β_{20}	−0.5	−0.473	0.453	0.381	0.954	−0.484	0.304	0.259	0.947
β_{21}	0.5	0.577	0.519	0.489	0.970	0.524	0.386	0.309	0.961
ϕ	1.0	1.078	0.254	0.227	0.935	1.042	0.152	0.134	0.941
α	−1.3	−1.333	0.177	0.167	0.905	−1.317	0.124	0.114	0.914
ν	1.5	1.530	0.191	0.125	0.986	1.517	0.133	0.086	0.986
CA		0.636	0.039			0.612	0.024		
DNB-LA									
β_1	1.0	1.040	0.288	0.307	0.914	1.022	0.222	0.207	0.940
β_{20}	−0.5	−0.460	0.514	0.446	0.931	−0.496	0.307	0.295	0.947
β_{21}	0.5	0.646	0.801	0.694	0.923	0.542	0.420	0.402	0.950
ϕ	1.0	1.081	0.267	0.239	0.937	1.032	0.131	0.129	0.943
α	−1.3	−1.308	0.226	0.189	0.938	−1.305	0.158	0.132	0.942
ν	1.5	1.523	0.222	0.152	0.975	1.513	0.155	0.105	0.971
CA		0.660	0.034			0.659	0.024		
DNB-RA									
β_1	1.0	1.025	0.302	0.274	0.916	1.010	0.212	0.191	0.920
β_{20}	−0.5	−0.469	0.491	0.418	0.946	−0.485	0.315	0.276	0.937
β_{21}	0.5	0.615	0.654	0.590	0.960	0.530	0.429	0.360	0.950
ϕ	1.0	1.064	0.297	0.276	0.939	1.037	0.142	0.131	0.945
α	−1.3	−1.320	0.187	0.158	0.940	−1.313	0.132	0.111	0.939
ν	1.5	1.526	0.202	0.134	0.984	1.513	0.142	0.092	0.985
CA		0.632	0.034			0.632	0.024		

incorporate intercept at the same time. Thus, only z_2 has an intercept term. It is chosen $\beta_1 = 1, \beta_{20} = -0.5$ and $\beta_{21} = 0.5$, implying cure rates 0.73, 0.67, 0.49 and 0.42 for profiles $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ respectively. Parameters related to the time-to-event for non-destroyed cells where chosen as $\alpha = -1.3, \nu = 1.5$ for the Weibull model, $\alpha = 0.8, \nu = 0.4$ for the Log-Normal model and $\alpha = 3.5, \nu = 1.5$ for the gamma model. Those parameters were used with FA, LA and RA schemes. We assume $\phi = 1$ in all cases.

For scheme FA, the two methods exposed in Section 4 were used with sample sizes $n = 200$ and $n = 400$. For schemes LA and RA, the second method exposed in Section 4 was used with sample sizes $n = 200$ and $n = 400$. In each case, 10,000 replicates were considered. Tables 4 shows part of the results for the simulations. We report the average of the estimates obtained (average), the mean of the asymptotic standard errors (se), the root of the mean squared error ($\sqrt{MSE}$) and the asymptotic coverage probability with 95% (CP). Main conclusions are that the two ways of implementing the EM algorithm provide close results relation to average, se, $\sqrt{MSE}$ and CP for the three activation schemes. Results also reveals that the estimates are closer to the true values and $\sqrt{MSE}$ is decreased as n increases, suggesting that estimators are consistent. On the other hand, the se is greater than $\sqrt{MSE}$, suggesting that the standard errors are overestimated. Despite this, the CP are closer to the nominal value.

5.2. Misspecification of the distribution for the non-destroyed concurrent cells

In the survival analysis literature, it is common to consider the Weibull distribution as the survival model for the time-to-event for the non-destroyed cells because of its appropriateness in many medical and biological contexts. However, to the best of our knowledge, we were unable to trace studies on the effects on both susceptible and cured parts of the model, of an incorrect specification of the survival function for the time-to-event for the non-destroyed cells.

Bearing this in mind, a simulation study is conducted using the same specification for parameters used in the last subsection. The three activation schemes mentioned in Section 3 and the Weibull, LN and gamma distributions for the time-to-event for non-destroyed cells were used. For each activation scheme/distribution combination, 10,000 samples were simulated and, for each sample, parameter estimates were computed (including $S(\cdot \mid \lambda)$). Then, the mean and MSE of the estimates were computed for each parameter and for the cure rate. Additionally, the mean and MSE for the expected times for the non-destroyed cells were also computed. Furthermore, the AIC and BIC criteria were computed for the three distribution and which was the model choice based on those criteria. Since they provide similar results, data on AIC was presented. Results for FA scheme are shown in Table 5. It is expected that a wrong choice for $S(\cdot \mid \lambda)$ increases the bias and the MSE for the expected activation time for non-destroyed cells. However,

Table 5: *Estimated bias and MSE for cure rate and expected values for the non-destroyed cells in DNB-FA with different activation schemes.*

			Cure rate		$\mathbb{E}(W)$		
	True		First Activation Scheme				
n	Distribution	Distribution	bias	MSE	bias	MSE	% AIC
200	Weibull	Weibull	0.001	0.004	−0.049	0.129	0.912
		Log-Normal	0.087	0.014	−7.909	195.7	0.080
		Gamma	0.008	0.902	0.689	0.902	0.008
	Log-Normal	Weibull	0.005	0.004	0.337	0.138	0.025
		Log-Normal	0.000	0.004	0.294	0.118	0.932
		Gamma	0.294	0.004	−24.1	689.0	0.043
	Gamma	Weibull	0.000	0.004	2.974	8.886	0.083
		Log-Normal	0.019	0.005	2.321	5.857	0.086
		Gamma	0.001	0.004	−0.722	6.089	0.831
400	Weibull	Weibull	0.002	0.002	−0.030	0.084	0.920
		Log-Normal	0.094	0.012	−6.908	88.6	0.038
		Gamma	0.008	0.002	0.782	0.787	0.042
	Log-Normal	Weibull	0.005	0.002	0.336	0.125	0.001
		Log-Normal	0.000	0.002	0.301	0.106	0.923
		Gamma	0.000	0.002	−22.5	549.7	0.075
	Gamma	Weibull	0.000	0.002	2.974	8.865	0.074
		Log-Normal	0.019	0.002	2.363	5.754	0.074
		Gamma	0.000	0.002	−0.313	2.371	0.852

the wrong choice also impacts on the cure rate estimates. Except for the gamma model, the AIC and BIC criteria chose the correct model for more than 90% of generate samples, suggesting that those criteria are appropriate to this purpose. For other activation schemes, similar results are obtained.

6. Application

In this section we analyze the cutaneous melanoma data set described in Section 2. Models DNB-FA, DNB-LA and DNB-RA were fitted to the data, with the survival functions from the Weibull, LN and gamma distributions used as survival functions for the time-to-event for the non-destroyed cells. To avoid identifiability problems, the covariates treatment, age, nodule and thickness were incorporated into the model through the θ and p parameters. All possible combinations of covariates preserving identifiability were considered and the combination that provided the least AIC and BIC criteria was

Table 6: *AIC/BIC criteria for E1690 data set using the DNB with different activation schemes.*

	Activation Scheme		
$S(\cdot \mid \lambda)$	FA	LA	RA
Weibull	**827.6/863.7**	854.3/890.4	842.7/878.8
LN	834.6/870.7	851.8/888.0	842.0/878.1
gamma	828.0/864.1	854.5/890.6	841.8/877.9

selected, leading to the one assigning nodule and tumour thickness to θ and treatment to p (see equation (4)). Given that all considered patients have cutaneous melanoma, it is reasonable to assume that the nodule category is related to the number of initial cells (most advance stage, more initial cells) and the same with tumour thickness (greater tumour, more initial cells). On the other hand, treatment can be interpreted as an element that determines the chance of such cells be activated (patients receiving the treatment have reduced their probability of initial activation of the initial cells). Table 6 shows the AIC and BIC vales for those combinations of covariates. Based on those criteria, the DNB-FA/Weibull model was chosen as the one presenting the best fit. On the other hand, it makes sense to use this activation scheme in a biological context, because just one cell can trigger the metastasis process. The estimates for this model are presented in Table 7.

Table 7: *Parameter estimates for the DNB-FA/Weibull model.*

Parameter	est	se	|est|/se
$\beta_{1,nodule1}$	0.4690	0.4565	1.03
$\beta_{1,nodule2}$	1.5143	0.3661	4.14
$\beta_{1,nodule3}$	2.1539	0.4044	5.32
$\beta_{1,nodule4}$	3.0702	0.4210	7.29
$\beta_{1,thickness}$	0.0858	0.0473	1.81
$\beta_{2,treatment}$	-0.7965	0.4064	1.96
ϕ	3.1807	0.0785	
α	-1.3142	0.1977	
ν	1.5372	0.0273	

The estimated means of the initial number of cells are $1.60 \times 1.09^{thickness}$ (nodule 1), $4.55 \times 1.09^{thickness}$ (nodule 2), $8.62 \times 1.09^{thickness}$ (nodule 3) and $21.55 \times 1.09^{thickness}$ (nodule 4) and the probability of activation of those cells is 0.5 for patients in control group and 0.31 for patients in the treatment group.

Finally Figure 3 shows the estimated mean of non-destroyed cells (D) for each patients stratified by control and treatment group. Note that the estimated means of D vary on both group, agreeing with the fact that the treatment is effective. On the other hand, it is possible to conclude that patients with nodule 4 have more estimated non-destroyed

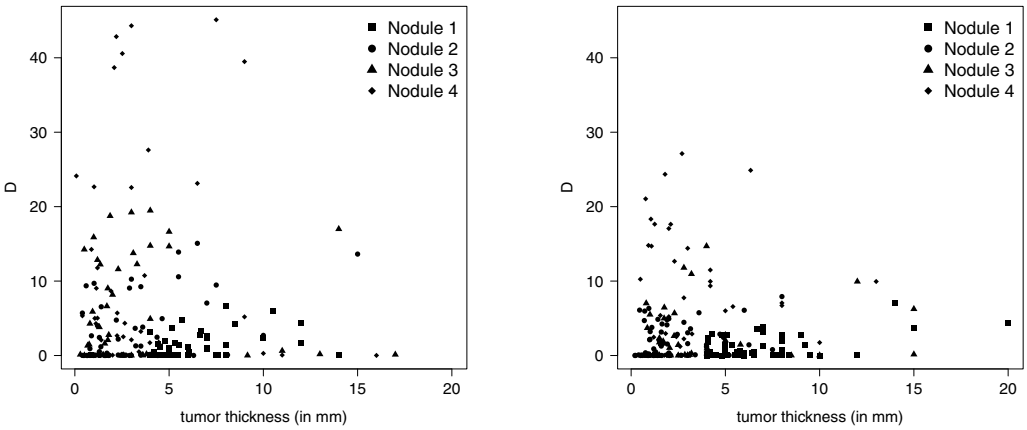

Figure 3: *Predicted means of the conditional distributions for all patients under the DNB-FA/Weibull model for the number of activated cells (D), stratified by nodule categories and belonging to the control group (left panel) and to the treatment group (right panel), respectively.*

cells. This is expected because patients in this stage of disease are more susceptible to die faster than patients in others stages of disease.

7. Final discussion

In this paper, an alternative estimation procedure based on the EM algorithm is proposed for the destructive Negative Binomial cure rate model introduced in Cancho et al. (2013). Two different ways of implementing the algorithm are investigated. Simulation studies indicate that those procedures work satisfactorily. It also investigated other alternatives (besides the Weibull distribution) for the survival function for the time for non-destroyed cells $S(\cdot \mid \lambda)$, and through the use of simulation studies evaluating the performances of the AIC/BIC criteria to correctly choose the model that provides the best fit to the data. Using simulation studies we assess the performances of the AIC/BIC criteria to correctly choose the model that provides the best fit to the data. However, a wrong choice for $S(\cdot \mid \lambda)$ can lead to incorrect estimates in both, the parameters related to the cure rate and the ones related to the survival function of the time-to-event for non-destroyed cells. Thus, precision loss is incurred if the wrong model is selected, that is, one has to be careful when selecting the working model. For this reason, it will be proposed non-parametric frameworks to estimate $S(\cdot \mid \lambda)$. Finally, the proposed approach was illustrated using real data related to a clinical trial on Phase III cutaneous melanoma patients.

Acknowledgments

We thank two referees and Editor for comments and suggestions that substantially improved the presentation of the work. The research of Diego I. Gallardo was supported by FAPESP grant number 2013/23684-2.

8. Appendix: Proofs of propositions

8.1. Appendix A: Proposition 1

For DNB-FA and DNB-LA the result is trivial. On the other hand, note it is possible to show that the marginal distribution of $D_i \mid \theta_i, p_i, \phi$ is $NB(\theta_i p_i, \phi)$. Thus, for the DNB-RA we have that for $r_1 \in \{1,2,\dots,\}$

$$P(R_i = r_i \mid D_{obs}, \psi) = \frac{\sum_{d_i=r_i}^{\infty} f(t_i, \delta_i \mid D_i = d_i, R_i = r_i)}{[(1-q_{0i})f(t_i;\lambda)]^{\delta_i}[q_{0i} + (1-q_{0i})S(t_i;\lambda)]^{1-\delta_i}},$$

where $f(t_i, \delta_i \mid D_i = d_i, R_i = r_i)$ is defined in (3). For $\delta_i = 1$, the expression takes the form

$$P(R_i = r_i \mid D_{obs}, \psi) = \frac{1}{(1-q_{0i})} \sum_{d_i=r_i}^{\infty} d_i \binom{d_i - 1}{r_i - 1} S(t_i;\lambda)^{d_i - r_i} F(t_i;\lambda)^{r_i - 1} P(D_i = d_i; \theta_i, p_i, \phi)$$

$$= \frac{F(t_i;\lambda)^{r_i-1}}{(1-q_{0i})} \mathbb{E}\left[D_i \binom{D_i - 1}{r_i - 1} S(t_i;\lambda)^{D_i - r_i} I(D_i \geq r_i) \right].$$

For $\delta_i = 0$,

$$P(R_i = r_i \mid D_{obs}, \psi) = \frac{\sum_{d_i=r_i}^{\infty} IB(S(t_i;\lambda), d_i - r_i + 1, r_i) P(D_i = d_i; \theta_i, p_i, \phi)}{q_{0i} + (1-q_{0i})S(t_i;\lambda)}.$$

On the other hand, by using the binomial theorem, it can be shown that $IB(S(t_i;\lambda), d_i - r_i + 1, r_i) = \sum_{k=0}^{r_i-1} \binom{r_i-1}{k} (-1)^k \frac{S(t_i;\lambda)^{d_i - r_i + k + 1}}{d_i - r_i + k + 1}$. In other words,

$$P(R_i = r_i \mid D_{obs}, \psi) = \frac{\sum_{k=0}^{r_i-1} \binom{r_i-1}{k} (-1)^k \mathbb{E}\left[\frac{S(t_i;\lambda)^{D_i - r_i + k + 1}}{D_i - r_i + k + 1} I(D_i \geq r_i) \right]}{q_{0i} + (1-q_{0i})S(t_i;\lambda)}.$$

8.2. Appendix B: proposition 2

Consider now the DNB-FA model ($R_i = 1$, $i = 1,\ldots,n$). Thus, by (2) and (5) the expression $P(D_i = d_i \mid D_{obs}, \psi)$ assumes the following form

$$P(D_i = d_i \mid D_{obs}, \psi) = \frac{S(t_i;\lambda)^{d_i-\delta_i}[d_i f(t_i;\lambda)]^{\delta_i}\frac{\Gamma(\phi^{-1}+d_i)}{\Gamma(\phi^{-1})d_i!}\left(\frac{\phi\theta_i p_i}{1+\phi\theta_i p_i}\right)^{d_i}(1+\phi\theta_i p_i)^{-\phi^{-1}}}{(\theta_i p_i f(t_i;\lambda))^{\delta_i}\left(1+\phi\theta_i p_i F(t_i;\lambda)\right)^{-(\phi^{-1}+\delta_i)}}$$

$$= \frac{\Gamma((\phi^{-1}+\delta_i)+d_i-\delta_i)}{\Gamma(\phi^{-1}+\delta_i)(d_i-\delta_i)!}\theta_{1i}^{d_i-\delta_i}\left(1-\theta_{1i}\right)^{(\phi^{-1}+\delta_i)},$$

i.e., $D_i - \delta_i \mid D_{obs} \mid \psi \sim NB\left((\phi^{-1}+\delta_i)^{-1}, \theta_{1i}\right)$, where $\theta_{1i} = \frac{\phi\theta_i p_i S(t_i;\lambda)}{1+\phi\theta_i p_i}$. For the DNB-LA, $R_i = D_i$, $i = 1,\ldots,n$ and then

$$P(D_i = d_i \mid D_{obs}, \psi) = \frac{\left\{d_i F(t_i;\lambda)^{d_i-1} f(t_i;\lambda)\right\}^{\delta_i}\left\{1-F(t_i;\lambda)^{d_i}\right\}^{1-\delta_i}}{\left\{\theta_i p_i f(t_i;\lambda)\left(1+\phi\theta_i p_i S(t_i;\lambda)\right)^{-(\phi^{-1}+1)}\right\}^{\delta_i}}\times$$

$$\times\frac{\frac{\Gamma(\phi^{-1}+d_i)}{\Gamma(\phi^{-1})d_i!}\left(\frac{\phi\theta_i p_i}{1+\phi\theta_i p_i}\right)^{d_i}(1+\phi\theta_i p_i)^{-\phi^{-1}}}{\left\{1+q_{0i}-\left(1+\phi\theta_i p_i S(t_i;\lambda)\right)^{-\phi^{-1}}\right\}^{1-\delta_i}}.$$

For $\delta_i = 1$, this expression takes the form

$$P(D_i = d_i \mid D_{obs}, \psi) = \frac{\Gamma((\phi^{-1}+1)+(d_i-1))}{\Gamma(\phi^{-1}+1)(d_i-1)!}\theta_{2i}^{d_i-1}\left(1-\theta_{2i}\right)^{-(\phi^{-1}+1)},$$

i.e., $(D_i - 1) \mid D_{obs}, \psi \sim NB\left((\phi^{-1}+1)^{-1}, \theta_{2i}\right)$, where $\theta_{2i} = \frac{\phi\theta_i p_i F(t_i;\lambda)}{1+\phi\theta_i p_i}$. For $\delta_i = 0$, this expression is reduced to

$$P(D_i = d_i \mid D_{obs}, \psi) = a_i\frac{\Gamma(\phi^{-1}+d_i)}{\Gamma(\phi^{-1})d_i!}\left(\frac{\phi\theta_i p_i}{1+\phi\theta_i p_i}\right)^{d_i}(1+\phi\theta_i p_i)^{-\phi^{-1}}$$

$$+(1-a_i)\frac{\Gamma(\phi^{-1}+d_i)}{\Gamma(\phi^{-1})d_i!}\theta_{2i}^{d_i}\left(1-\theta_{2i}\right)^{\phi^{-1}},$$

where $a_i = (1+q_{0i}-(1+\phi\theta_i p_i S(t_i;\lambda))^{-\phi^{-1}})^{-1}$, i.e., $D_i \mid D_{obs}, \psi \sim a_i NB\left(\phi, \frac{\phi\theta_i p_i}{1+\phi\theta_i p_i}\right) + (1-a_i)NB(\phi, \theta_{2i})$. Finally, for DNB-RA we have that

$$P(D_i = d_i \mid D_{obs}, \psi) = \frac{\sum_{r_i=1}^{d_i} \left\{ d_i \binom{d_i-1}{r_i-1} f(t_i;\lambda) S(t_i;\lambda)^{d_i-r_i} F(t_i;\lambda)^{r_i-1} \right\}^{\delta_i} \frac{1}{d_i}}{[q_{0i} f(t_i;\lambda)]^{\delta_i} [q_{0i} + (1 - q_{0i} S(t_i;\lambda))]^{1-\delta_i}}$$
$$\times \{IB(S(t_i;\lambda), d_i - r_i + 1, r_i)\}^{1-\delta_i}$$
$$\frac{\Gamma(\phi^{-1} + d_i)}{\Gamma(\phi^{-1}) d_i!} \left(\frac{\phi\theta_i p_i}{1 + \phi\theta_i p_i} \right)^{d_i} (1 + \phi\theta_i p_i)^{-\phi^{-1}}.$$

For $\delta_i = 1$, it is immediate that

$$P(D_i = d_i \mid D_{obs}, \psi) = \frac{\Gamma(\phi^{-1} + d_i)}{\Gamma(\phi^{-1}) d_i!} \left(\frac{\phi\theta_i p_i}{1 + \phi\theta_i p_i} \right)^{d_i} I(d_i \geq 1),$$

i.e., $(D_i - 1) \mid D_{obs}, \psi \sim NB(\phi, \frac{\phi\theta_i p_i}{1+\phi\theta_i p_i})$. Finally, for $\delta_i = 0$, using the binomial theorem, the expression is reduced to

$$P(D_i = d_i \mid D_{obs}, \psi) = \frac{\sum_{r_i=1}^{d_i} \sum_{k=0}^{r_i} (-1)^k \binom{r_i-1}{k} \frac{S(t_i;\lambda)^{d_i-r_i+k+1}}{d_i(d_i-r_i+k+1)} \frac{\Gamma(\phi^{-1}+d_i)}{\Gamma(\phi^{-1}) d_i!} \left(\frac{\phi\theta_i p_i}{1+\phi\theta_i p_i} \right)^{d_i}}{1 + [(1 - q_{0i})/q_{0i}] S(t_i;\lambda)}.$$

8.3. Appendix C: proposition 3

Considering the DNB-FA model ($R_i = 1,\ i = 1, \ldots, n$), and by (2) and (5) the expression $P(M_i = m_i \mid D_{obs}, \psi)$ assume the following form

$$P(M_i = m_i \mid D_{obs}, \psi) = \frac{\sum_{d_i=\delta_i}^{m_i} S(t_i;\lambda)^{d_i-\delta_i} [d_i f(t_i;\lambda)]^{\delta_i} \binom{m_i}{d_i} p_i^{d_i} (1 - p_i)^{m_i - d_i}}{(\theta_i p_i f(t_i;\lambda))^{\delta_i} (1 + \phi\theta_i p_i F(t_i;\lambda))^{-(\phi^{-1}+\delta_i)}} \times$$
$$\times \frac{\Gamma(\phi^{-1} + m_i)}{\Gamma(\phi^{-1}) m_i!} \left(\frac{\phi\theta_i}{1 + \phi\theta_i} \right)^{m_i} (1 + \phi\theta_i)^{-\phi^{-1}}$$
$$= \frac{\Gamma((\phi^{-1} + \delta_i) + m_i - \delta_i)}{\Gamma(\phi^{-1} + \delta_i)(m_i - \delta_i)!} \theta_{3i}^{m_i - \delta_i} (1 - \theta_{3i})^{(\phi^{-1}+\delta_i)},$$

i.e., $M_i - \delta_i \mid D_{obs} \mid \psi \sim NB\left((\phi^{-1} + \delta_i)^{-1}, \theta_{3i}\right)$, where $\theta_{3i} = \frac{\phi\theta_i(1 - p_i F(t_i;\lambda))}{1+\phi\theta_i}$. For the DNB-LA, $R_i = D_i$, $i = 1, \ldots, n$ and then

$$P(M_i = m_i \mid D_{obs}, \psi) = \sum_{d_i=\delta_i}^{m_i} \left[\frac{\left\{ d_i F(t_i;\lambda)^{d_i-1} f(t_i;\lambda) \right\}^{\delta_i} \left\{ 1 - F(t_i;\lambda)^{d_i} \right\}^{1-\delta_i}}{\left\{ \theta_i p_i f(t_i;\lambda) \left(1 + \phi\theta_i p_i S(t_i;\lambda)\right)^{-(\phi^{-1}+1)} \right\}^{\delta_i}} \times \right.$$

$$\left. \times \frac{\binom{m_i}{d_i} p_i^{d_i} (1-p_i)^{m_i-d_i} \frac{\Gamma(\phi^{-1}+m_i)}{\Gamma(\phi^{-1})m_i!} \left(\frac{\phi\theta_i}{1+\phi\theta_i}\right)^{m_i} (1+\phi\theta_i)^{-\phi^{-1}}}{\left\{ 1 + q_{0i} - \left(1 + \phi\theta_i p_i S(t_i;\lambda)\right)^{-\phi^{-1}} \right\}^{1-\delta_i}} \right]$$

For $\delta_i = 1$, this expression is reduced to

$$P(M_i = m_i \mid D_{obs}, \psi) = \frac{\Gamma((\phi^{-1}+1)+m_i-1)}{\Gamma(\phi^{-1}+1)(m_i-1)!} \theta_{4i}^{m_i-1} \left(1-\theta_{4i}\right)^{(\phi^{-1}+1)},$$

i.e., $(M_i - 1) \mid D_{obs}, \psi \sim NB\left((\phi^{-1}+1)^{-1}, \theta_{4i}\right)$, where $\theta_{4i} = \frac{\phi\theta_i(1-p_i S(t_i;\lambda))}{1+\phi\theta_i}$. For $\delta_i = 0$, this expression takes the form

$$P(M_i = m_i \mid D_{obs}, \psi) = a_i \frac{\Gamma(\phi^{-1}+m_i)}{\Gamma(\phi^{-1})m_i!} \left(\frac{\phi\theta_i}{1+\phi\theta_i}\right)^{m_i} (1+\phi\theta_i)^{-\phi^{-1}}$$

$$+ (1-a_i) \frac{\Gamma(\phi^{-1}+m_i)}{\Gamma(\phi^{-1})m_i!} \theta_{4i}^{m_i} \left(1-\theta_{4i}\right)^{\phi^{-1}},$$

where $a_i = (1 + q_{0i} - (1+\phi\theta_i p_i S(t_i;\lambda))^{-\phi^{-1}})^{-1}$, i.e., $M_i \mid D_{obs}, \psi \sim a_i NB\left(\phi, \frac{\phi\theta_i}{1+\phi\theta_i}\right) + (1-a_i) NB(\phi, \theta_{4i})$. Finally, for DNB-RA we have that

$$P(M_i = m_i \mid D_{obs}, \psi) = \frac{\sum_{d_i=\delta_i}^{m_i} \sum_{r_i=1}^{d_i} \left\{ d_i \binom{d_i-1}{r_i-1} f(t_i;\lambda) S(t_i;\lambda)^{d_i-r_i} F(t_i;\lambda)^{r_i-1} \right\}^{\delta_i} \times \frac{1}{d_i}}{[q_{0i} f(t_i;\lambda)]^{\delta_i} [q_{0i} + (1 - q_{0i} S(t_i;\lambda))]^{1-\delta_i}}$$

$$\times \left\{ IB(S(t_i;\lambda), d_i - r_i + 1, r_i) \right\}^{1-\delta_i} \times$$

$$\times \binom{m_i}{d_i} p_i^{d_i} (1-p_i)^{m_i-d_i} \frac{\Gamma(\phi^{-1}+m_i)}{\Gamma(\phi^{-1})m_i!} \left(\frac{\phi\theta_i}{1+\phi\theta_i}\right)^{m_i} (1+\phi\theta_i)^{-\phi^{-1}}.$$

For $\delta_i = 1$, the expression is reduced to

$$P(M_i = m_i \mid D_{obs}, \psi) = \left[1 - (1-p_i)^{m_i}\right] \frac{\Gamma(\phi^{-1}+m_i)}{\Gamma(\phi^{-1})m_i!} \left(\frac{\phi\theta_i}{1+\phi\theta_i}\right)^{m_i} I(m_i \geq 1).$$

Finally, for $\delta_i = 0$,

$$P(M_i = m_i \mid D_{obs}, \psi) = \frac{\sum_{d_i=0}^{m_i} \sum_{r_i=1}^{d_i} \sum_{k=0}^{r_i} v_i \left(\frac{p_i}{1-p_i}\right)^{d_i} \left(\frac{\phi\theta_i(1-p_i)}{1+\phi\theta_i}\right)^{m_i}}{1 + [(1-q_{0i})/q_{0i}]S(t_i;\lambda)}.$$

References

Balakrishnan, N. and Pal, S. (2009). EM algorithm-based likelihood estimation for some cure rate models. *Journal of Statistical Theory and Practice*, 6, 698–724.

— (2013). Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes. *Statistical Methods in Medical Research.* DOI:10.1177/0962280213491641.

— (2015). An EM algorithm for the estimation of parameters of a flexible cure rate model with generalized gamma lifetime and model discrimination using likelihood-and information-based methods. *Computational Statistics*, 30, 151–189.

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

Cancho, V., Bandyopadhyay, D., Louzada, F. and Yiqi, B. (2013). The destructive negative binomial cure rate model with a latent activation scheme. *Statistical Methodology*, 13, 48–68.

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

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

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B*, 39, 1–38.

Gallardo, D.I., Bolfarine, H. and Pedroso-de-Lima, A.C. (2016). An EM algorithm for estimating the destructive weighted Poisson cure rate model. *Journal of Statistical Computation and Simulation*, 86, 1497–1515.

Hanin, L. and Huang, L.S. (2014). Identifiability of cure models revisited. *Multivariate Data Analysis*, 130, 261–274.

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

Li, C.S., Taylor, J. and Sy, J. (2001). Identifiability of cure models. *Statistics and Probability Letters*, 54, 389–395.

Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure estimation. *Statistica Sinica*, 20, 661–674.

MacDonald, I.L. (2014). Numerical Maximisation of Likelihood: A Neglected Alternative to EM? *International Statistical Review*, 82, 296–308.

Miller, R.G. (1974). The jackknife: a review. *Biometrika*, 61, 1–15.

R Development Core Team. (2015). *R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing*. ISBN 3-900051-07-0, version 3.0.3.

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

Rodrigues, J., Castro, M., Balakrishnan, N. and Cancho, V.G. (2011). Destructive weighted Poisson cure rate models. *Lifetime Data Analysis*, 17, 333–346.

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

Williams, J.S. and Lagakos, S.W. (1977). Models for censored survival analysis: constant-sum and variable-sum models. *Biometrika*, 64, 215–224.

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