Lifetime Data Anal (2019) 25:1–25
https://doi.org/10.1007/s10985-018-9417-8

Vertical modeling: analysis of competing risks data with a cure fraction

Mioara Alina Nicolaie[1] · **Jeremy M. G. Taylor**[2] · **Catherine Legrand**[1]

Received: 12 August 2015 / Accepted: 6 January 2018 / Published online: 31 January 2018

Abstract In this paper, we extend the vertical modeling approach for the analysis of survival data with competing risks to incorporate a cure fraction in the population, that is, a proportion of the population for which none of the competing events can occur. The proposed method has three components: the proportion of cure, the risk of failure, irrespective of the cause, and the relative risk of a certain cause of failure, given a failure occurred. Covariates may affect each of these components. An appealing aspect of the method is that it is a natural extension to competing risks of the semi-parametric mixture cure model in ordinary survival analysis; thus, causes of failure are assigned only if a failure occurs. This contrasts with the existing mixture cure model for competing risks of Larson and Dinse, which conditions at the onset on the future status presumably attained. Regression parameter estimates are obtained using an EM-algorithm. The performance of the estimators is evaluated in a simulation study. The method is illustrated using a melanoma cancer data set.

Keywords Mixture cure model · Competing risks · Cumulative incidences

1 Introduction

In medicine, the risk of failure from a given disease or the chance of getting cured of it are of interest to diseased patients as well as to the treating physicians; this helps

✉ Mioara Alina Nicolaie
mioara.nicolaie@uclouvain.be; alina.nicolaie@rivm.nl

[1] Institute of Statistics, Biostatistics and Actuarial Sciences, Catholic University of Louvain, Voie du Roman Pays 20, bte L1.04.01, 1348 Louvain-la-Neuve, Belgium

[2] School of Public Health, University of Michigan, M4509 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109-2029, USA

patients at risk to make important life decisions and it helps physicians in treatment selection. However, investigating this type of information in a clinical study involving time-to-event data requires accounting for multiple possible outcomes: either failure, due to the disease or due to some other cause, or cure. From the statistical point of view, the analysis of several types of failures is incorporated in the competing risks framework. These methods usually assume that all patients will eventually experience one of the possible types of failure if there is sufficient follow-up and, therefore, do not accomodate cure. The presence of cure is suggested when the data includes a considerable number of long-term event-free survivors (censored patients with long follow-up times). Therefore, extending modeling of competing risks to accommodate a cure proportion is an important issue in understanding this type of data.

In the analysis of single outcome survival data with a cured fraction, cure models address the problem of cure rate estimation, as well as the estimation of the probability of failure due to the disease of interest. They have received a lot of attention both in terms of methodological developments (Farewell 1986; Sy and Taylor 2000; Li and Taylor 2002; Yu et al. 2004; Kim et al. 2013) and applications (Andersson et al. 2011; Andrae et al. 2012). These cure models are mixture models which specify a conditional model for the survival component, given that failure may occur, and a marginal distribution of the binary indicator for whether or not cure can occur. They are formulated either in a parametric or in a semi-parametric way (Kuk and Chen 1992; Taylor 1995; Sy and Taylor 2000; Peng and Dear 2000; Peng 2003; Corbiere et al. 2009). They share the common feature of allowing the cure rate to be determined from the onset, but to be observed only later in the course of the follow-up, leading to the presence of a sub-population of event-free survivors beyond sufficiently long follow-up.

Analysis of competing risks data with a cure fraction is not well developed. Several authors caste the competing risks model of Larson and Dinse (1985) into the cure framework. These models express the mixing joint distribution of the time-to-event and type of event variables as the sum of the marginal cause-specific distributions multiplied by the associated mixing proportions. This approach is formulated under the strong, unverifiable assumption that the mixing proportions for failure types and cure indicator is determined but unobserved at the onset. In terms of formulas, this amounts to the following decomposition of the joint distribution of time of failure T and failure type D:

$$P(T, D) = P(T|D) \cdot P(D). \tag{1}$$

Examples include the approach of Chao (1998), which imputes the cure indicator for censored patients and uses a Gibbs sampling algorithm for estimation; Ng and McLachlan (1998) propose a parametric version, to be used when there are only few failures from the competing causes and Choi and Zhou (2002) discuss extensively a class of multivariate parametric models. This approach could be adopted in contexts where the interest is in assessing the parameters of the conditional failure time distribution given failure type or in the mixing distribution of the different competing failure types.

However, from the inference and interpretation points of view, the case where the model parameters translate directly into natural observable quantities in competing risks is appealing. In this paper, we adopt this perspective and introduce a semi-parametric approach for the analysis of competing risks data with a cure fraction. This approach extends the idea of vertical modeling formulated earlier by Nicolaie et al. (2010) in the competing risks framework, and it is based on the following decomposition of the joint distribution of time of failure T and failure type D:

$$P(T, D) = P(D|T) \cdot P(T). \tag{2}$$

In the remainder of the article, we demonstrate how this can be applied to the analysis of mixture cure data with several competing causes of failure in the presence of right-censored data. In Sect. 2 we introduce in detail our approach. Simulation studies are presented in Sect. 3. In Sect. 4 we illustrate our methods of analysis in a clinical study of melanoma cancer. Section 5 concludes with some points for discussion.

2 Vertical modeling with a cure fraction

2.1 Notation

Suppose that data are available from n individuals each of whom can experience one of J terminal, competing events during the period under study or can be subjected to noninformative right censoring. Assume that a non-negligible proportion of individuals do not experience any of the terminal events by the end of the follow-up and assume further that follow-up is long enough to be able to consider individuals with full follow-up as cured from the disease of interest on the basis of clinical knowledge. Such a population can naturally be regarded as a mixture population in which two categories of individuals are combined: susceptible (the individual experiences failure, irrespective of the cause) and non-susceptible or cured (the individual is immune to all causes of failure). Let T denote the time-to-event variable, C the right-censoring time variable, and D the terminal event type, where $D \in \{1, \ldots, J\}$. Assign $D = 1$ for the cause of interest and $D \geq 2$ for the competing causes. Let $\mathbf{Z}$ denote an l-vector of covariates measured at baseline.

Consider a binary random variable Y such that $Y = 1$ corresponds to a susceptible individual and $Y = 0$ corresponds to a nonsusceptible/cured individual. Note that in this type of study Y is partially observed; it equals 1 in the case of an event, and it is unobserved in the case of right-censoring. If the latter occurs, the individual has no event observed during the study period, but either the event will eventually take place (the individual is censored and susceptible) or the event will never take place in the future (the individual is censored and non-susceptible).

The observed data for individual i is $\mathcal{O}_i = (\tilde{T}_i, \Delta_i, Z_i)$, where $\tilde{T}_i = \min(T_i, C_i)$ is the earliest of time-to-event T_i and censoring time C_i, and $\Delta_i = \mathbf{1}\{T_i < C_i\}D_i$ is the type of terminal event in the case a terminal event occurs and 0 in the case of censoring, for $i = 1, \ldots, n$. Data from different individuals are assumed to be independent. Assume that (T, D) and C are independent given $\mathbf{Z}$.

2.2 Model formulation

Our goal is to develop and implement an approach to determine whether a terminal event would occur (whether an individual is susceptible, which hereafter is referred to as *incidence*) and, conditional on being susceptible, when the event might occur and which type of event it might be, given a failure occurred (which hereafter are jointly referred to as *latency*). In addition, we are interested in assessing covariate effects on incidence and latency.

For ease of notation in formulating the model and describing the method we will omit covariates, and note that they can be included in an obvious way, as will be done for the melanoma dataset in Sect. 4. For the incidence part let $P(Y = 1) = p$; thus, $1 - p$ represents the proportion of individuals who get cured. For the latency part, we aim to extend the vertical modeling approach, earlier proposed by Nicolaie et al. (2010), to the mixture cure model framework. We shall refer to this competing risks mixture cure model as vertical modeling with a cure fraction VMCF [in contrast to vertical modeling (VM) which does not take a cure fraction into account].

The main idea behind the modeling of the latency part of VMCF is to specify the conditional (on $Y = 1$) joint distribution $P(T, D|Y = 1)$ as

$$P(T, D|Y = 1) = P(T|Y = 1) \cdot P(D|T, Y = 1), \tag{3}$$

that is, the product of the conditional (on $Y = 1$) failure rate $P(T|Y = 1)$ and of the conditional distribution of the causes of failure, given a failure occurred $P(D|T, Y = 1)$. If we assume that the survival time T is continuous, we define the conditional (on $Y = 1$) total hazard by

$$\lambda_\bullet(t|Y = 1) = \lim_{\Delta t \to 0} \frac{P(t \le T \le t + \Delta t | T \ge t, Y = 1)}{\Delta t} \tag{4}$$

and its cumulative counterpart by $\Lambda_\bullet(t|Y = 1) = \int_0^t \lambda_\bullet(u|Y = 1)du$, which then gives $P(T|Y = 1)$. We will also define the conditional (on $Y = 1$) cause-specific hazard of cause j, representing the instantaneous risk of experiencing cause j in the susceptible population:

$$\lambda_j(t|Y = 1) = \lim_{\Delta t \to 0} \frac{P(t \le T \le t + \Delta t, D = j | T \ge t, Y = 1)}{\Delta t}, \qquad j = 1, \ldots, J.$$

In turn, the conditional (on $Y = 1$) survival function of susceptible individuals, defined as $S(t|Y = 1) = P(T > t|Y = 1)$ is given by

$$S(t|Y = 1) = \exp(-\Lambda_\bullet(t|Y = 1)). \tag{5}$$

We emphasize that $S(t|Y = 1)$ is a proper survival function in the sense that $\lim_{t\to\infty} S(t|Y = 1) = 0$. Note that the conditional (on $Y = 0$) survival function of non-susceptibles is degenerate, that is, $P(T > t|Y = 0) = 1$. We define the conditional (on $Y = 1$) relative cause-specific hazard of cause j at time t by

$$\pi_j(t|Y=1) = P(D=j|T=t, Y=1), \tag{6}$$

which can be shown to equal $\lambda_j(t|Y=1)/\lambda_\bullet(t|Y=1)$, $j=1,\ldots,J$. Note that the probability $\pi_j(t|Y=1)$ involves both the failure time and the cause, therefore its estimation involves only the (observed) susceptible individuals. This implies that (6) can be expressed as:

$$P(D=j|T=t, Y=1) = P(D=j|T=t), \qquad j=1,\ldots,J. \tag{7}$$

As a consequence, we suppress the dependence on $Y=1$ of $\pi_j(t|Y=1)$ and, in the following, we will denote it by $\pi_j(t)$. Thus the vector $(\pi_j(t))_{j=1,\ldots,J}$ denotes the conditional (on $T=t$) distribution $P(D|T=t)$, with $\Sigma_j \pi_j(t) = 1$ for all t. Thus the vector $(\lambda_\bullet(t|Y=1), (\pi_j(t))_{j=1,\ldots,J})$ completely specifies the latency part of VMCF.

The conditional (on $Y=1$) cumulative incidence function of cause j, defined as $F_j(t|Y=1) = P(T \le t, D=j|Y=1)$, can be obtained as

$$F_j(t|Y=1) = \int_0^t \lambda_\bullet(u|Y=1)\pi_j(u)S(u-|Y=1)du, \qquad j=1,\ldots,J. \tag{8}$$

The marginal survival function, which is defined as $S_{pop}(t) = P(T \ge t)$ can be expressed as

$$\begin{aligned} S_{pop}(t) &= P(Y=1)P(T \ge t|Y=1) + P(Y=0)P(T \ge t|Y=0) \\ &= p \cdot S(t|Y=1) + 1 - p. \end{aligned} \tag{9}$$

Note that $S_{pop}(t)$ is an improper survival function in the sense that $\lim_{t\to\infty} S_{pop}(t) = 1-p$. For this marginal survival $S_{pop}(t)$ we can define a marginal (population) total hazard as

$$\lambda_{\bullet,\mathrm{pop}}(t) = \lim_{\Delta t \to 0} \frac{P(t \le T \le t + \Delta t | T \ge t)}{\Delta t}.$$

The relationship between the marginal total hazard $\lambda_{\bullet,\mathrm{pop}}(t)$ and the conditional (on $Y=1$) total hazard $\lambda_\bullet(t|Y=1)$ is given by

$$\lambda_{\bullet,\mathrm{pop}}(t) = \frac{-S'_{pop}(t)}{S_{pop}(t)} = \frac{pS(t|Y=1)\lambda_\bullet(t|Y=1)}{pS(t|Y=1)+1-p}, \tag{10}$$

where S' stands for the derivative of S.

The marginal cumulative incidence function of cause j, defined as $F_j(t) = P(T \le t, D=j)$, is given by

$$F_j(t) = \int_0^t \lambda_{\bullet,\mathrm{pop}}(u)\pi_j(u)S_{pop}(u-)du, \qquad j=1,\ldots,J. \tag{11}$$

2.3 Specific models

For the incidence $p(\mathbf{X}) = P(Y = 1|\mathbf{X})$, where $\mathbf{X} \subseteq \mathbf{Z}$, we postulate a binary regression model:

$$g(p(\mathbf{X})) = \beta^\top \mathbf{X}^*, \tag{12}$$

where $\mathbf{X}^* = (1, \mathbf{X})$, β stands for a vector of unknown regression parameters, including an intercept, and g is a known differentiable link function, such as the logistic link $g(p) = \log \frac{p}{1-p}$.

For the conditional (on $Y = 1$) total hazard we postulate a Cox proportional hazards model:

$$\lambda_\bullet(t|Y = 1, \mathbf{Z}) = \lambda_0(t|Y = 1) \exp(\boldsymbol{\gamma}^\top \mathbf{Z}), \tag{13}$$

where $\lambda_0(\cdot|Y = 1)$ is an unspecified conditional (on $Y = 1$) baseline hazard and γ stands for a vector of unknown regression parameters.

For the relative hazards we specify

$$\pi_j(t|\mathbf{U}) = \frac{\exp\left(\kappa_j^\top \mathbf{B}(t) + \upsilon_j^\top \mathbf{U}\right)}{\sum_{l=1}^J \exp\left(\kappa_l^\top \mathbf{B}(t) + \upsilon_l^\top \mathbf{U}\right)}, \qquad j = 1, \ldots, J, \tag{14}$$

where $\mathbf{U} \subseteq \mathbf{Z}$, $\mathbf{B}(t)$ is an r-vector of pre-specified time functions and $\eta_j = (\kappa_j, \upsilon_j)$ stands for an m-vector of unknown regression parameters, $j = 1, \ldots, J$. For identifiability, we set $\eta_J \equiv 0$. Examples of $\mathbf{B}(t)$ are polynomial or spline functions. This yields a flexible parametric model which can provide smooth estimates of $\pi_j(t|\mathbf{U})$; model choice of $\mathbf{B}(t)$ can be guided by the Akaike information criterion. Denote by $\boldsymbol{\theta}$ the vector $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \eta_1, \ldots, \eta_J)$ of all regression parameters characterizing the components of VMCF.

2.4 Marginal model

Omitting from the notation the dependence on covariates, it can be shown that the conditional expectation of Y given $T \geq t$ is determined by

$$\begin{aligned} E[Y|T \geq t] = P(Y = 1|T \geq t) &= \frac{P(Y = 1)P(T \geq t|Y = 1)}{P(T \geq t)} \\ &= \frac{p \cdot S(t|Y = 1)}{pS(t|Y = 1) + 1 - p}. \end{aligned} \tag{15}$$

Thus the relationship in (10) between the marginal total hazard $\lambda_{\bullet,\text{pop}}(t)$ and the conditional (on $Y = 1$) total hazard $\lambda_\bullet(t|Y = 1)$ can be written as

$$\lambda_{\bullet,\text{pop}}(t) = E[Y|T \geq t]\lambda_\bullet(t|Y = 1), \tag{16}$$

which, intuitively, expresses the fact that $\lambda_{\bullet,\text{pop}}(t)$ is the average of the $Y\lambda_\bullet(t|Y = 1)$ taken over all the individuals at risk just before $T = t$.

Note that model (13) implies that in the presence of covariates the marginal total hazard is given by

$$\lambda_{\bullet,\text{pop}}(t|\mathbf{Z},\mathbf{X}) = \lambda_0(t|Y=1)\exp\left(\boldsymbol{\gamma}^\top\mathbf{Z}\right) \times \frac{g^{-1}\left(\boldsymbol{\beta}^\top\mathbf{X}^*\right)S_0(t|Y=1)^{\exp\left(\boldsymbol{\gamma}^\top\mathbf{Z}\right)}}{g^{-1}\left(\boldsymbol{\beta}^\top\mathbf{X}^*\right)S_0(t|Y=1)^{\exp\left(\boldsymbol{\gamma}^\top\mathbf{Z}\right)} + 1 - g^{-1}\left(\boldsymbol{\beta}^\top\mathbf{X}^*\right)}, \quad (17)$$

where $S_0(t|Y=1) = \exp(-\Lambda_0(t|Y=1))$. This clearly shows that at the population level, the proportional hazards assumption postulated in the strata $\{Y=1\}$ no longer holds. For two individuals with covariate vectors $\mathbf{Z}$ and $\widetilde{\mathbf{Z}}$, the following relationship holds:

$$\log\frac{\lambda_{\bullet,\text{pop}}(t|\mathbf{Z})}{\lambda_{\bullet,\text{pop}}(t|\widetilde{\mathbf{Z}})} = \boldsymbol{\gamma}(\mathbf{Z}-\widetilde{\mathbf{Z}}) + \log\frac{E[Y|T\geq t,\mathbf{Z},\mathbf{X}]}{E[Y|T\geq t,\widetilde{\mathbf{Z}},\widetilde{\mathbf{X}}]}, \quad (18)$$

which expresses that the log hazard ratio at the population level equals the sum of the log hazard ratio in the strata $\{Y=1\}$ and a time-varying term.

2.5 Likelihood

The observed data for individual i is $(\tilde{T}_i, \Delta_i, Z_i)$, and let $t_0 = 0$ and denote by $t_1 \leq t_2 \leq \ldots \leq t_K$ the K ordered event times. Omitting the dependence on Z_i from the notation, the likelihood is the product of contributions of individuals from two categories: an individual i who fails at time t_i due to cause j contributes

$$P(T_i = t_i, D_i = j, Y_i = 1) = P(Y_i = 1)\cdot P(T_i = t_i|Y_i = 1)\cdot P(D_i = j|T_i = t_i),$$

and an individual i who is censored at time t contributes

$$P(T_i > t) = \left[P(Y_i = 1)\cdot P(T_i > t|Y_i = 1) + P(Y_i = 0)\right].$$

Assuming that the distributions of T and C have no common parameters, then the observed data likelihood is proportional to

$$L = \prod_{i=1}^{n}\left[p_i P(T_i = \tilde{T}_i|Y_i = 1)\prod_{j=1}^{J}P(D_i = j|T_i = \tilde{T}_i)^{\mathbf{1}\{\Delta_i = j\}}\right]^{\mathbf{1}\{\Delta_i>0\}} \cdot\prod_{i=1}^{n}\left[p_i P(T_i > \tilde{T}_i|Y_i = 1) + (1-p_i)\right]^{\mathbf{1}\{\Delta_i=0\}}.$$

This can be rewritten as

$$L = \prod_{i=1}^{n} \left\{ p_i \lambda_\bullet(\widetilde{T}_i|Y_i = 1) \exp\left[-\Lambda_\bullet(\widetilde{T}_i|Y_i = 1)\right] \prod_{j=1}^{J} \pi_j(\widetilde{T}_i)^{\mathbf{1}\{\Delta_i = j\}} \right\}^{\mathbf{1}\{\Delta_i > 0\}} \cdot \prod_{i=1}^{n} \left\{ p_i \exp\left[-\Lambda_\bullet(\widetilde{T}_i|Y_i = 1)\right] + (1 - p_i) \right\}^{\mathbf{1}\{\Delta_i = 0\}}.$$

In the special case of absence of covariates $\mathbf{Z}$ in the models specified in Sect. 2.3, L is a function of the conditional baseline hazard $\lambda_0(t|Y = 1)$ and of the regression parameters $(\kappa_1, \ldots, \kappa_J)$, that is, $L = L(\kappa_1, \ldots, \kappa_J, \lambda_0(t|Y = 1))$, which can be separated in the following way:

$$L(\kappa_1, \ldots, \kappa_J, \lambda_0(t|Y = 1)) = L_1(\lambda_0(t|Y = 1)) \cdot L_2(\kappa_1, \ldots, \kappa_J), \tag{19}$$

where

$$L_1(\lambda_0(t|Y = 1)) = \prod_{i=1}^{n} \left\{ p_i \lambda_0(\widetilde{T}_i|Y_i = 1) \exp\left[-\Lambda_0(\widetilde{T}_i|Y_i = 1)\right] \right\}^{\mathbf{1}\{\Delta_i > 0\}} \cdot \prod_{i=1}^{n} \left\{ p_i \exp\left[-\Lambda_0(\widetilde{T}_i|Y_i = 1)\right] + (1 - p_i) \right\}^{\mathbf{1}\{\Delta_i = 0\}}$$

and

$$L_2(\kappa_1, \ldots, \kappa_J) = \prod_{i=1}^{n} \prod_{j=1}^{J} \pi_j(\widetilde{T}_i)^{\mathbf{1}\{\Delta_i = j\}} = \prod_{i=1}^{n} \prod_{j=1}^{J} \left\{ \frac{\exp\left[\kappa_j^\top \mathbf{B}(t_i)\right]}{\sum_{l=1}^{J} \exp\left[\kappa_l^\top \mathbf{B}(t_i)\right]} \right\}^{\mathbf{1}\{\Delta_i = j\}}.$$

In the presence of covariates $\mathbf{Z}$, the dependence of the likelihood function on covariates will explicitly take the form of a joint conditional distribution. By employing the models specified in Sect. 2.3, L is a function of the conditional baseline hazard $\lambda_0(t|Y = 1)$ and of the regression parameters $\boldsymbol{\theta}$, that is, $L = L(\boldsymbol{\theta}, \lambda_0(t|Y = 1))$, which can be separated in the following way:

$$L(\boldsymbol{\theta}, \lambda_0(t|Y = 1)) = L_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y = 1)) \cdot L_2(\eta_1, \ldots, \eta_J), \tag{20}$$

where

$$L_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y = 1)) = \prod_{i=1}^{n} \left\{ p_i(\mathbf{X}_i) \lambda_\bullet(\widetilde{T}_i|Y_i = 1, \mathbf{Z}_i) \times \exp\left[-\Lambda_\bullet(\widetilde{T}_i|Y_i = 1, \mathbf{Z}_i)\right] \right\}^{\mathbf{1}\{\Delta_i > 0\}}$$

$$\cdot \prod_{i=1}^{n} \left\{ p_i(\mathbf{X}_i) \exp\left[-\Lambda_\bullet(\widetilde{T}_i|Y_i=1,\mathbf{Z}_i)\right] + (1-p_i(\mathbf{X}_i)) \right\}^{\mathbf{1}\{\Delta_i=0\}}$$

and

$$L_2(\eta_1,\ldots,\eta_J) = \prod_{i=1}^{n}\prod_{j=1}^{J} \pi_j(\widetilde{T}_i|\mathbf{U}_i)^{\mathbf{1}\{\Delta_i=j\}}.$$

2.6 Estimation

We will use a variant of the EM algorithm given in Sy and Taylor (2000) to estimate the parameters $(\boldsymbol{\theta}, \lambda_0(t_1|Y=1),\ldots,\lambda_0(t_K|Y=1))$ of the VMCF. The technique consists of an adaptation of the EM algorithm to deal with the latent Y and to accommodate our semi-parametric approach. Typically, it is assumed that all event-free survivors after t_K get cured. This amounts to imposing the zero-tail constraint, that is, $S(t|Y=1)=0$ for $t \geq t_K$, a condition which assures that the observed likelihood is well-behaved (see also Sy and Taylor 2000).

Denote by $\mathcal{C}$ the complete data, that is, the sample data when the latent Y would be observed for all cases. The complete data likelihood function is given by:

$$L_{\mathcal{C}}(\boldsymbol{\theta}, \lambda_0(t|Y=1)) = L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1); Y) \cdot L_2(\eta_1,\ldots,\eta_J),$$

where

$$\begin{aligned} L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1); Y) = & \prod_{i=1}^{n} \left\{ p_i \lambda_\bullet(\widetilde{T}_i|Y_i=1) \exp\left[-\Lambda_\bullet(\widetilde{T}_i|Y_i=1)\right] \right\}^{\mathbf{1}\{\Delta_i>0\}} \\ & \cdot \prod_{i=1}^{n} \left\{ \left\{ p_i \exp\left[-\Lambda_\bullet(\widetilde{T}_i|Y_i=1)\right] \right\}^{\mathbf{1}\{Y_i=1\}} \right. \\ & \left. \cdot (1-p_i)^{\mathbf{1}\{Y_i=0\}} \right\}^{\mathbf{1}\{\Delta_i=0\}}. \end{aligned}$$

Note that $\Delta_i > 0$ implies $Y_i = 1$, and $Y_i = 0$ implies $\Delta_i = 0$. After rearranging the factors, we get:

$$\begin{aligned} L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1); Y) = & \prod_{i=1}^{n} p_i^{\mathbf{1}\{Y_i=1\}} (1-p_i)^{\mathbf{1}\{Y_i=0\}} \\ & \cdot \prod_{i=1}^{n} \lambda_\bullet(\widetilde{T}_i|Y_i=1)^{\mathbf{1}\{\Delta_i>0\}} \exp\left[-Y_i \Lambda_\bullet(\widetilde{T}_i|Y_i=1)\right] \\ = & L_{31}(\boldsymbol{\beta}; Y) \cdot L_{32}(\boldsymbol{\gamma}, \lambda_0(t|Y=1); Y). \end{aligned} \tag{21}$$

The maximization of $\log L_{\mathcal{C}}(\boldsymbol{\theta}, \lambda_0(t|Y=1))$ is enhanced by the observation that the multinomial logistic structure embedded in $\log L_2(\eta_1, \ldots, \eta_J)$ can be maximized independently of $\log L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1); Y)$. An appealing fact is that this can be achieved by means of standard software like the PROC GLM in SAS or the **glm** function in R (R Development Core Team 2010). The use of these functions counts on specifying the event type as dependent variable and the time-dependent terms and baseline covariates as independent variables and also the family of probability distributions to use, by the family option family=binomial defaults to logistic regression. Denote by $(\widehat{\eta}_1, \ldots, \widehat{\eta}_J)$ the estimator of $(\eta_1, \ldots, \eta_J)$. Maximization of $\log L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1); Y)$ is more involved and uses the EM-algorithm following the approach in Sy and Taylor (2000). Technical details are given in Appendix A.

2.7 Standard errors

An approximation of the asymptotic variance of $(\boldsymbol{\theta}, \lambda_0(t_1|Y=1), \ldots, \lambda_0(t_K|Y=1))$ can be obtained as the inverse of the observed full information matrix $\Im_{\boldsymbol{\theta}, \lambda_0(t|Y=1)}$ of $L(\theta, \lambda_0(t|Y=1))$. Note that due to the factorisation (20), we can write:

$$\Im_{\boldsymbol{\theta}, \lambda_0(t|Y=1)} = \begin{pmatrix} \Im_{\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1)} & \vert & 0 \\ - & \vert & - \\ 0 & \vert & \Im_{\boldsymbol{\eta}} \end{pmatrix}$$

where $\Im_{\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_\bullet(t|Y=1)}$ is the observed full information matrix of $L_1(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1); Y)$ and $\Im_{\boldsymbol{\eta}}$ is the observed full information matrix of $L_2(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_J)$.

Sy and Taylor (2000) derived a formula for $\Im_{\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_\bullet(\cdot|Y=1)}$. In Appendix B, we use their results to obtain the standard errors of the conditional (on $Y=1$) cause-specific cumulative hazards from vertical modeling with a cure fraction.

3 Simulations

In this simulation study, we assess the performance of our estimators in the semi-parametric setting when the assumed model is correct. We simulate data for $n = 500$ individuals, each of whom can fail due to the disease (cause 1) or due to a competing cause (cause 2), or can be right-censored. Assume a non-negligible proportion of individuals get cured. Individuals are followed over a period of at most 15 years; random right-censoring, which is independent of survival time, occurred uniformly in the interval [7, 15] years. Assume $Z \sim N(0, 1)$ is a continuous, normally distributed baseline covariate. We generate data as random samples drawn from a VMCF model (the "true" model), where we assume that the cure indicator Y follows a logistic regression model:

$$\text{logit}(P(Y=1|Z)) = \beta_0 + \beta_1 \cdot Z,$$

for $(\beta_0, \beta_1) \in \{(-0.62, 1.24), (-1.38, 0), (1.38, 0)\}$ leading to different amounts of cure proportions, that is, 65% for $Z = 1$, 80 and 20% respectively. The latency part is specified by

$$\lambda_\bullet(t|Y=1,Z)=\lambda_0(t|Y=1)\exp(\gamma Z),$$

where $\gamma = 0.3$ and the baseline hazard $\lambda_0(t|Y=1)$ is either constant, equal to 0.4 (scenario 1) or piecewise constant, equal to 0.3 for $t \in [0,3]$ and equal to 0.4 for $t \in (3,15]$ (scenario 2). In scenario 1, the relative hazards are constant such that $\pi_1(t) = 0.25$ and $\pi_2(t) = 0.75$, favoring cause 2 over cause 1. This corresponds to model (14), where $B(t) = 1$ for all t, $\kappa_1 = -[\kappa_2 = -1.0986$ and no covariates are included in the model. In scenario 2, the relative hazards follow model (14), where we include covariate Z and set $\mathbf{B}(t) = (\mathbf{1}_{[0,3]}(t), \mathbf{1}_{(3,15]}(t)]$ such that:

$$\begin{aligned}\pi_1(t|Z) &= \frac{\exp(-1.4\times \mathbf{1}_{[0,3]}(t) - 2.2\times \mathbf{1}_{(3,15]}(t) + 2.5\times Z)}{1+\exp(-1.4\times \mathbf{1}_{[0,3]}(t) - 2.2\times \mathbf{1}_{(3,15]}(t) + 2.5\times Z)},\\ \pi_2(t|Z) &= 1-\pi_1(t|Z).\end{aligned}$$

VMCF and VM are fitted to the simulated data generated in the two scenarios. The incidence component of VMCF is assumed to have the same form as in the true model and the conditional (on $Y = 1$) total hazard follows a Cox proportional hazards model including Z as covariate and with unspecified baseline hazard. When fitting models to the data generated in scenario 1, the relative hazards are assumed constant over time, with no covariate effect; the observed proportions of cause specific events relative to the total number of events were employed as their estimates. When fitting models to the data generated in scenario 2, the relative hazards are assumed to follow model (14) with the same set of covariates as in the true model. The covariate Z was therefore included in the two mixture components (incidence and latency). The probabilities $\widehat{F}_j(t|Y_i = 1, Z_i)$, $j = 1, 2$, were computed by plugging in the estimates $\widehat{\pi}_j(t|Z)$, $j = 1, 2$, and $\widehat{\lambda}_\bullet(t|Y = 1, Z)$ in Eq. (8). The probabilities $\widehat{F}_j(t|Z_i)$, $j = 1, 2$, were computed by plugging in the estimates $\widehat{\pi}_j(t|Z)$, $j = 1, 2$, $\widehat{\lambda}_\bullet(t|Y = 1, Z)$ and $\widehat{p}(Z)$ in Eqs. (10) and (11). They are reported either at each of the prediction time points $\{1, 2, 5, 7, 11\}$ (scenario 1) or at each of the prediction time points $\{1, 3, 10\}$ (scenario 2) for an individual i with $Z_i = 1$. In VM, a Cox proportional hazards model with Z as predictor and an unspecified baseline hazard is postulated for the total hazard. The relative hazards are identical with those in VMCF for each scenario in turn. The estimates $\widehat{F}_j(t|Z_i)$, $j = 1, 2$, are reported either at each of the prediction time points $\{1, 2, 5, 7, 11\}$ (scenario 1) or at each of the prediction time points $\{1, 3, 10\}$ (scenario 2) for an individual i with $Z_i = 1$.

We report in Tables 1 and 2 the estimated bias and root mean squared error (RMSE) obtained for our estimators under various proportions of susceptibles in the two scenarios. Each instance was run 10,000 times.

The proposed method of estimation VMCF shows little bias in estimating both $\widehat{F}_j(t|Y = 1, Z = 1)$ and $\widehat{F}_j(t|Z = 1)$, with bias (RMSE) increasing with the increase in the cure proportion. The bias contributes little to the RMSE. The behaviour at later time points is more uncertain due to the smaller number of events by the end of the follow-up. Comparisons of $\widehat{F}_j(t|Z = 1)$ derived from VMCF and VM with the true cause-specific cumulative incidences consistently show smaller bias (RMSE) for VMCF, thus supporting the idea that fitting VMCF to data from a population with a

Table 1 Estimated bias (root mean squared error) of $\widehat{F}_j(t|Y=1,Z=1)$ with respect to the true cumulative incidence function $F_j(t|Y=1,Z=1)$, $j=1,2$, for the three instances of cure proportions

t	65% Cure in $\{Z=1\}$		80% Cure		20% Cure	
	$\widehat{F}_1(t\|Y=1,Z=1)$	$\widehat{F}_2(t\|Y=1,Z=1)$	$\widehat{F}_1(t\|Y=1,Z=1)$	$\widehat{F}_2(t\|Y=1,Z=1)$	$\widehat{F}_1(t\|Y=1,Z=1)$	$\widehat{F}_2(t\|Y=1,Z=1)$
Scenario 1						
1	−0.0001 (0.0162)	−0.0005 (0.0314)	−0.0003 (0.0239)	−0.0008 (0.0500)	−0.0004 (0.0119)	−0.0012 (0.0247)
2	0.0001 (0.0229)	0.0003 (0.0350)	0.0000 (0.0328)	0.0002 (0.0559)	−0.0002 (0.0164)	−0.0006 (0.0276)
5	0.0012 (0.0302)	0.0036 (0.0337)	0.0018 (0.0419)	0.0057 (0.0495)	0.0004 (0.0208)	0.0015 (0.0240)
7	0.0017 (0.0313)	0.0052 (0.0335)	0.0030 (0.0435)	0.0092 (0.0472)	0.0008 (0.0215)	0.0027 (0.0228)
10	0.0029 (0.0321)	0.0088 (0.0339)	0.0049 (0.0447)	0.0151 (0.0476)	0.0013 (0.0219)	0.0041 (0.0225)
Scenario 2						
1	0.0004 (0.0322)	−0.0011 (0.0199)	0.0002 (0.0374)	−0.0019 (0.0234)	0.0000 (0.0264)	−0.0010 (0.0166)
3	0.0025 (0.0462)	−0.0019 (0.0384)	0.0032 (0.0564)	−0.0033 (0.0459)	0.0009 (0.0395)	−0.0018 (0.0324)
10	0.0082 (0.0514)	0.0017 (0.0502)	0.0084 (0.0634)	0.0006 (0.0623)	0.0047 (0.0442)	0.0004 (0.0437)

The data were generated from VMCF according to the two scenarios 1 or 2

Table 2 Estimated bias (root mean squared error) of $\widehat{F}_j(t|Z=1)$ with respect to the true cumulative incidence function $F_j(t|Z=1)$, $j=1,2$, when the VM and VMCF models are used

t	VM		VMCF	
	$\widehat{F}_1(t\|Z=1)$	$\widehat{F}_2(t\|Z=1)$	$\widehat{F}_1(t\|Z=1)$	$\widehat{F}_2(t\|Z=1)$
Scenario 1				
1	−0.0064 (0.0119)	−0.0193 (0.0284)	0.0000 (0.0112)	0.0002 (0.0229)
2	−0.0086 (0.0170)	−0.0260 (0.0366)	0.0004 (0.0160)	0.0012 (0.0281)
5	−0.0084 (0.0217)	−0.0254 (0.0393)	0.0012 (0.0214)	0.0037 (0.0322)
7	−0.0071 (0.0221)	−0.0214 (0.0376)	0.0016 (0.0222)	0.0049 (0.0330)
10	−0.0053 (0.0222)	−0.0159 (0.0353)	0.0024 (0.0228)	0.0073 (0.0338)
Scenario 2				
1	−0.0155 (0.0255)	−0.0059 (0.0133)	0.0008 (0.0226)	−0.0005 (0.0132)
3	−0.0256 (0.0410)	−0.0101 (0.0258)	0.0027 (0.0350)	−0.0008 (0.0258)
10	−0.0184 (0.0427)	−0.0052 (0.0334)	0.0069 (0.0412)	0.0018 (0.0344)

The data were generated from VMCF according to the two scenarios 1 or 2, with true parameter values $\beta_0=-0.62$, $\beta_1=1.24$ and $\gamma=0.3$, leading to a proportion of cure of 0.65

Table 3 Estimated coverage probabilities of $\widehat{F}_j(t|Y=1,Z=1)$ and $\widehat{F}_j(t|Z=1)$, $j=1,2$ when the VM and VMCF models are used

t	VMCF		VM		VMCF	
	$\widehat{F}_1(t\|Y=1,Z=1)$	$\widehat{F}_2(t\|Y=1,Z=1)$	$\widehat{F}_1(t\|Z=1)$	$\widehat{F}_2(t\|Z=1)$	$\widehat{F}_1(t\|Z=1)$	$\widehat{F}_2(t\|Z=1)$
Scenario 1						
1	95.2	95.4	90.8	85.1	95.2	95.5
2	95.1	95.1	91.6	82.9	95.0	95.0
5	95.1	95.1	93.3	86.8	95.1	95.2
7	95.2	95.0	94.2	89.9	95.1	94.8
10	95.1	94.4	94.8	92.5	95.1	94.6
Scenario 2						
1	94.87	95.12	88.83	92.47	95.06	95.12
3	95.02	94.72	87.93	92.92	94.89	94.88
10	94.70	94.67	92.61	94.59	94.87	94.90

The data were generated from VMCF according to the two scenarios 1 or 2, with true parameter values $\beta_0=-0.62$, $\beta_1=1.24$ and $\gamma=0.3$, leading to a proportion of cure of 0.65

non-negligable proportion of cure can replace a conventional survival model when the interest is to estimate the disease frequency.

The coverage rates for the normal approximations 95% confidence intervals of $\widehat{F}_j(t|Y=1,Z=1)$ and $\widehat{F}_j(t|Z=1)$ are higher for VMCF compared to VM. They are reported in Table 3 for each of the two scenarios when the proportion of cure amounts to 65% for $Z=1$. The other two instances, with heavy or small amount of cure, show the same pattern (not reported). The poor coverage rates when fitting VM, given that VMCF is the true model, might be due to the loss of accuracy in estimating

covariate effects in the total hazard model. This is readily clarified if we look at Eq. (18); what VM estimates is the quantity γ instead of the full right hand term. The lack of fit reduces by the end of the follow-up interval, leading to an increase in accuracy of the effect estimate in the model of total hazard from VM and, therefore, increased coverage rates at later time points for VM.

The performance of estimators $\widehat{\beta_0}$, $\widehat{\beta_1}$ and $\widehat{\gamma}$ was studied in Sy and Taylor (2000). Since these parameters are separated from the parameters of the relative hazards model, the results obtained in Sy and Taylor (2000) are applicable to and were confirmed in our setting and consequently have not been explored further.

4 Data analysis

The data comprises 205 patients with malignant melanoma who had their tumor removed by surgery between 1962 and 1977. These data from Andersen et al. (1993) are part of the **boot** package (Canty and Ripley 2008) for the R software. This was a prospective clinical study to assess the effect of risk factors on survival. Baseline covariates are gender, age (in years) at operation, year of operation, tumor thickness and ulceration of the tumor tissue. Age (scaled by 10), year of operation (centered at 1970 and scaled by 10) and thickness were considered continuous covariates; age ranged from 4 to 95 years and thickness ranged from 0.10 to 17.42 mm. 61% of patients were men and 44% of patients presented with ulceration of the tissue at the time of surgery. The follow-up time ranged from 10 days to 15.23 years. The survival time is known only for those patients who have had their event before the end of 1977. The rest of the patients are censored at the end of 1977. Endpoint of interest is the time from operation to death due to melanoma (cause 1), in the presence of a competing cause, that is, death due to other causes, unrelated to melanoma (cause 2). The total number of deaths was 71:57 (80%) deaths due to cause 1 and 14 (20%) deaths due to cause 2. Of the 134 censored patients, 27 were censored with longer follow-up time than the largest death time. Figure 1 shows the Kaplan–Meier estimate of the survival function of time to any terminal event. This survival curve appears to reach a plateau 10 years post-surgery, indicating the presence of a sub-population which survives event-free by the end of the follow-up and clearly suggesting the appropriateness of a competing risks mixture cure model. For these data clinical interest is to understand how covariates affect the survival distribution, after surgery for melanoma, in the presence of competing risks.

First a vertical model ignoring cure (VM) was fit to the data. This model serves two purposes: (1) to estimate the effect of covariates on time to death due to each of melanoma cancer and other causes and (2) to estimate the probabilities of each type of events since surgery. The model aggregates the covariate effects on latency and incidence in an intricate way [see, for instance, Eq. (17)]. A Cox proportional hazards model was used for the total hazard including all covariates:

$$\lambda_\bullet(t|\mathbf{Z}) = \lambda_0(t) \exp\left(\boldsymbol{\gamma}^\top \mathbf{Z}\right),$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard and γ stands for a vector of unknown regression parameters, and a logistic regression model was used for the relative hazards including all covariates and piecewise constant time functions, with cut-off points at

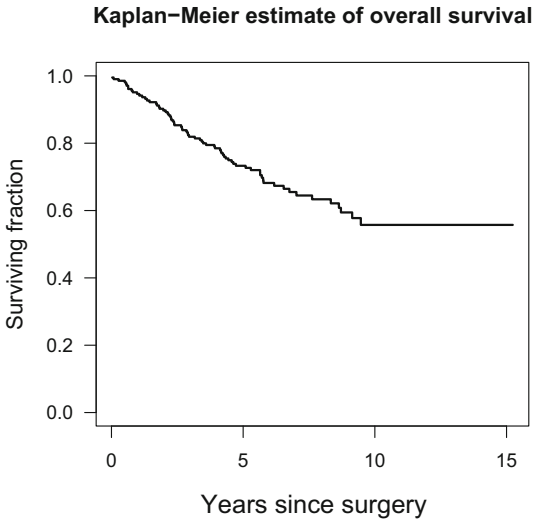

Fig. 1 The Kaplan–Meier estimate of time to all-causes failure

Table 4 Estimated regression parameters in VM and VMCF

Covariate	Regression parameters (SE)			
	VM	VMCF		Both VM and VMCF
	γ in Cox model	β in incidence model	γ in conditional Cox model	(υ_1, κ_1) in logistic model
Intercept		− 2.62 (0.88)		
Thickness	0.09 (0.04)	0.07 (0.11)	0.10 (0.06)	0.01 (0.12)
Ulcer	0.98 (0.26)	0.95 (0.64)	0.87 (0.50)	1.46 (0.86)
Age	0.02 (0.08)	0.03 (0.15)	− 0.0006 (0.11)	− 0.04 (0.24)
Year (stand)	− 0.88 (0.55)	0.51 (0.89)	− 1.49 (1.03)	− 0.95 (1.70)
Sex	0.42 (0.24)	0.57 (0.47)	0.42 (0.50)	0.30 (0.72)
$\mathbf{1}\{t \in (0, 1.76]\}$				1.93 (1.87)
$\mathbf{1}\{t \in (1.76, 2.90]\}$				4.24 (2.08)
$\mathbf{1}\{t \in (2.90, 4.67]\}$				3.84 (1.80)
$\mathbf{1}\{t \in (4.67, 15.23]\}$				2.61 (1.67)

the 0.25, 0.5, 0.75 quartiles of the non-parametric failure time distribution estimator leading to:

$$\text{logit}(\pi_1(t|\mathbf{Z})) = \kappa_1^\top \mathbf{B}(t) + \upsilon_1^\top \mathbf{Z}, \tag{22}$$

where $\mathbf{B}(t) = (\mathbf{1}\{t \in (0, 1.76]\}, \mathbf{1}\{t \in (1.76, 2.90]\}, \mathbf{1}\{t \in (2.90, 4.67]\}, \mathbf{1}\{t \in (4.67, 15.23]\})$ and $\eta_1 = (\kappa_1, \upsilon_1)$ stands for a vector of unknown regression parameters. The estimated regression parameters are reported in Table 4.

For illustration, we present the results of the model fit for a male patient with ulceration and for average values of the continuous covariates. Table 5 gives the estimated

Table 5 Estimated piecewise constant relative hazard of cause 1 and standard errors for a male patient with ulceration and for average values of the continuous covariates derived by plugging-in the estimates of $(\boldsymbol{\gamma}, \upsilon_1, \kappa_1)$ from Table 4 in model formula 22

	$\pi_1(t \in (0, 1.76]\|\mathbf{Z})$	$\pi_1(t \in (1.76, 2.90]\|\mathbf{Z})$	$\pi_1(t \in (2.90, 4.67]\|\mathbf{Z})$	$\pi_1(t \in (4.67, 15.23]\|\mathbf{Z})$
Estimate (SE)	0.754 (0.141)	0.968 (0.038)	0.954 (0.044)	0.858 (0.106)

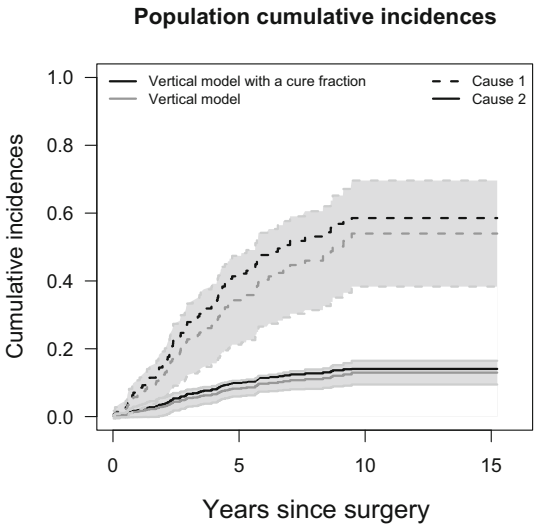

Fig. 2 The estimated cumulative incidences of time to melanoma and time to other causes based on VM (in gray) and on VMCF (in black) for a male individual with ulceration and for the mean values of the continuous covariates, together with the 95% confidence intervals derived by fitting the former model

relative hazards implied by fitting model (22) with associated standard errors. It can be seen that the dominating cause of death post surgery is melanoma cancer.

Figure 2 shows (in gray lines) the estimated cumulative incidence functions of time to type 1-event and time to type 2-event.

We then applied the vertical modeling with a cure fraction (VMCF) to these data. This model serves three purposes: (1) to estimate the effect of covariates on time to death due to each of melanoma cancer and other causes in the population of uncured patients, (2) to estimate the cause-specific cumulative incidences in the population of uncured patients and (3) to estimate the effect of covariates on the probability of being cured. A logistic regression model was postulated for the incidence part and for the latency part, a Cox proportional hazards model for the conditional (on $Y = 1$) total hazard and a logistic model for the relative hazards where piece-wise constant time functions were used with cut-off points at the quartiles of the failure time distribution function. Note that this model for the relative hazards coincides with the one used in VM where the evidence of cure was ignored, because the conditional (on $T = t$) distribution of causes of failure depends only on the actual causes of failure observed. All baseline covariates were included in both parts (incidence and latency) of the model. No selection of covariates was performed to test whether some covariates could be removed. The estimated regression parameters are reported in Table 4. The

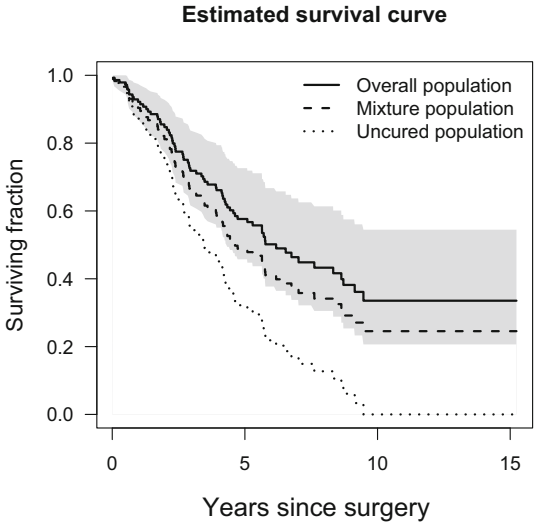

Fig. 3 The estimated survival curve of time to any event at the mean values of the continuous covariates for a male individual with ulceration, based on VM (solid line) and on VMCF for the combined population (dashed line) and for the susceptible sub-population (dotted line), together with the 95% confidence intervals derived by fitting the former model

results suggest that association of thickness with death due to melanoma found by the VM approach is best explained by the latency term in the VMCF model, whereas the associations of ulcer and sex in the VM model is explained by a combination of both incidence and latency in the VMCF model.

The estimators of the survival curve $S_{\text{pop}}(t|\mathbf{Z})$ and of the conditional (on $Y = 1$) survival curve $S(t|Y = 1, \mathbf{Z})$, for a male individual with ulceration and for average values of the continuous covariates, are plotted in Figure 3 in dashed and dotted lines, respectively, and compared with the estimator of the survival curve $S(t|\mathbf{Z})$ derived from VM (solid line).

The estimated conditional (on $Y = 1$) cumulative incidence function for a susceptible male patient with ulceration and for the mean values of the continuous covariates are plotted in Figure 4. The corresponding predicted probability of cure is $1-\widehat{p} = 0.25$. We also plotted in Figure 2 (in black) the estimated (unconditional) cumulative incidence functions of time to melanoma and time to other causes derived from VMCF, for a male individual with ulceration and average values of his other covariates. As expected in Figure 3, the estimates for the population $S_{\text{pop}}(t|\mathbf{Z})$ based on our VMCF model is similar to that based on VM.

Finally, it is worth noting the difference in estimates in Table 4. The coefficients in the Cox models differ because they are for different models, one is for $P(t|Y = 1, \mathbf{Z})$ and the other is for $P(t|\mathbf{Z})$. Also as expected the SE's are larger for $P(t|Y = 1, \mathbf{Z})$ than for $P(t|\mathbf{Z})$ because implicitly they are based on a smaller population of those with $Y = 1$.

5 Discussion

Little has been published about cure models in combination with competing risks. Several reviews have been recently published (Jia et al. 2013; Peng and Taylor 2013;

Fig. 4 Model-based estimates of the conditional (on $Y = 1$) cumulative incidences of time to melanoma and time to other causes for susceptible male patient with ulceration and for the mean values of the continuous covariates

Yu et al. 2013) to highlight the advantages of cure models over the limitations of standard methods like Kaplan–Meier, Cox models or parametric models for survival data when statistical cure is a reasonable assumption. In this paper, we have illustrated a strategy for the statistical analysis of competing risks data with a cure fraction. In this way, we have argued that, when there is clinical evidence of a cure proportion in a cohort, special attention should be given to the heterogeneity present among individuals.

Identifying covariates which contribute to failure-specific risk and time-to-event are sensible research goals for which traditional methods exist. The risk of failure is commonly analyzed using logistic or multinomial regression which allows one to estimate the effects of covariates as odds ratios, but ignores the shape of the survival function. Time-to-event data is typically analyzed using ordinary survival analysis or its extension to competing risks. In traditional competing risks studies, the focus is on the analysis of type-specific times to the event, while assuming all individuals will experience a type-specific failure, if followed long enough, and therefore, leaving out the scenario in which individuals remain event-free long enough or even during their entire lifetime. Therefore, this analysis treats all the individuals similarly, while there might be differences between subjects with a certain risk of failure and subjects that will never develop such a risk.

A general limitation of the approach we describe is that it relies on three models, a logistic model for cure, a Cox PH model for the time-to-event and a multinomial logistic for the type of event. The soundness of any conclusions will be dependent on the validity of the assumed structure of these three models and how well they fit the data. The multinomial logistic regression model used for the type of event is quite flexible, allowing various specifications of covariate effects (main effects or interactions) and diverse time functions (polynomials, splines). Since this part of the overall model is a separate component of the likelihood and is a model for observable quantities (the type

of failure amongst those that did fail) standard methods for assessing its fit can be used. However, the situation is harder for the model for cure and for the time-to-event model for the susceptibles. The complication arises because cure is not explicitly observed for everyone, so standard methods of assessing goodness-of-fit don't apply. Recent papers (Peng and Taylor 2017; Scolas et al. 2016a) suggest ways to examine residuals and do diagnostic checks for a mixture cure model, but further work is needed in this area. Logistic models and Cox PH models have been the most common choices in mixture cure models. But other choices could be made, such as complementary log-log instead of logistic and an accelerated failure time (AFT) or parametric survival model, instead of a Cox PH model. Metrics, such as Akaike Information Criteria, could be used to assess whether alternative choices give a better overall fit. The logistic and complementary log-log could be embedded in a larger class of models (Taylor and Liu 2007) to aid in determining which is preferred. The Cox PH model could be replaced by a parametric PH model, such as the Weibull model. A comparison of the general performance of these in a mixture cure model has been performed (Sy and Taylor 2000), and this comparison is relevant to the VMCF. The merits of using an AFT model instead of a Cox PH model has been discussed and there is literature on using them in a mixture cure model (Chen et al. 2013; Scolas et al. 2016b; Wei 1992; Yamaguchi 1992).

An important advantage of the competing risks mixture cure model (VMCF) over the standard competing risks model is that it accomodates this feature and simultaneously (1) it allows inference of the susceptible sub-population, and therefore a better understanding and interpretation of the variability of the data and (2) it allows estimation and direct modeling of the risk of failure or of the cure indicator. Summary measures like these can be a useful tool to complement the existing statistical measures encountered in traditional competing risks setting. Each covariate in VMCF can contribute up to three regression parameters, one parameter reflecting how the covariate affects the risk of failure or the chance of cure, one parameter for the time to event, irrespective of the cause of failure, and one parameter for the relative magnitude of each failure type among all failure types.

An appealing technical feature of VMCF resides in its parametrisation, which is illustrated in Eq. (20). VMCF naturally separates the observed likelihood into two factors: one where the cure indicator distribution is irrelevant and one where the cure indicator distribution is relevant. The former factor is pertinent to causes of failure when a failure occurs; in the absence of competing events, the observed likelihood (20) reduces to the observed likelihood of the Cox proportional hazards mixture cure model of Sy and Taylor (2000). The latter factor uses information on the failure and censoring times and, therefore, it is sensitive to the joint distribution of (T, Y). This feature makes our method straightforward to implement by means of smcure package available in R Development Core Team (2010).

We believe that our approach can play a useful role, as it can accommodate complex competing risks data. It is worth mentioning that the approach can be naturally extended to deal with missing causes of failure in competing risks. We refer to the work of Nicolaie et al. (2015) for a description of how this can be achieved.

An important issue is that semi-parametric mixture cure models are by construction non-identifiable. We cannot precisely tell apart individuals who are cured or not among

those who are censored. However, the presence of individuals with long follow-up and event-free works as empirical evidence of the existence of a cured subgroup. We adopt the strategy of Sy and Taylor (2000), who approach the non-identifiability problem through the use of the zero-tail constraint on the baseline failure time distribution. Another strategy has been adopted by Peng (2003), who impose a parametric shape on the tail of the failure time distribution.

The performance to reliably estimate a cure fraction is dependent on whether there is a sufficient follow-up for the subjects. A heuristic used in the literature is whether there are "lots of" censored observations after the latest event time, and whether the Kaplan–Meier plot appears to "level off", despite the fact that neither of them is very precisely defined. A more reasonable way to assess sufficiency of the follow-up is through hypothesis testing (Laska and Meisner 1992; Maller and Zhou 1994; Klebanov and Yakolev 2007). These tests can be used for testing sufficient follow-up in VMCF as well. The reason is that, in the modeling of the cause-specific survival in VMCF, event time distribution separates its parameters from those of the type of failure distribution, so the competing risks aspect can be neglected in the inference of the cure rate. Indeed, the critical connection for assessing sufficient follow-up is between the probability of cure and the total hazard function.

Another aspect goes to the core of how cure is perceived in the presence of competing risks, other than the risk that is related to the disease. It might be the case that a diseased patient is at risk of several mutually exclusive causes of failure and their potential cure from the disease cannot be observed if they die from accidental causes. For instance, a different formulation is proposed by Basu and Tiwari (2010), where separate competing risks structures are considered for the cure and the susceptible latent groups and a joint prior distribution is assumed on the collection of parameters. Cure is defined as not having experienced death due to the disease-related cause; therefore, the status of cure is observed only for individuals subjected to death not due to the cause of interest. We are currently adapting the vertical modeling approach to this situation.

Acknowledgements Funding was provided by IAP grant (Grant No. P7/06).

Appendix A: The EM-algorithm

In the E-step of the algorithm the conditional expectation of $\log L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y = 1); Y)$ is computed with respect to the distribution of the unobserved Y_i's, given the current parameters values and the observed data $\mathcal{O} = (\mathcal{O}_i)_{i=1,\ldots,n}$. As Y_i's contribute as linear terms in $\log L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y = 1); Y)$, it is enough to compute, at a given iteration m, the weight $w_i^{(m)} = E(Y_i|\boldsymbol{\theta}^{(m)}, \lambda_0^{(m)}(t|Y = 1), \mathcal{O})$. For an individual i who experiences an event at time t_i (irrespective of its cause) the corresponding weight is

$$\begin{aligned} w_i^{(m)} &= E(Y|\boldsymbol{\theta}^{(m)}, \lambda_0^{(m)}(t_i|Y = 1), T = t_i, D_i \in \{1, \ldots, J\}) \\ &= P(Y_i = 1|\boldsymbol{\theta}^{(m)}, \lambda_0^{(m)}(t_i|Y = 1), T = t_i, D_i \in \{1, \ldots, J\}) \\ &= 1, \end{aligned}$$

while for an individual i who is censored at time t the corresponding weight is

$$\begin{aligned} w_i^{(m)} &= E\left(Y|\boldsymbol{\theta}^{(m)}, \lambda_0^{(m)}(t_i|Y=1), T>t, D_i=0\right) \\ &= P\left(Y_i=1|\boldsymbol{\theta}^{(m)}, \lambda_0^{(m)}(t_i|Y=1), T>t, D_i=0\right) \\ &= \frac{g^{-1}\left(\boldsymbol{\beta}^\top \mathbf{X}_i^*\right)\cdot S_0(t|Y=1)^{\exp(\boldsymbol{\gamma}^\top \mathbf{Z}_i)}}{g^{-1}\left(\boldsymbol{\beta}^\top \mathbf{X}_i^*\right)\cdot S_0(t|Y=1)^{\exp(\boldsymbol{\gamma}^\top \mathbf{Z}_i)}+1-g^{-1}\left(\boldsymbol{\beta}^\top \mathbf{X}_i^*\right)}\Bigg|_{(\boldsymbol{\theta},\lambda_0(t|Y=1))=\left(\boldsymbol{\theta}^{(m)},\lambda_0^{(m)}(t|Y=1)\right)}. \end{aligned}$$

Denote the expected complete log-likelihood by $\mathbf{E}_{\mathrm{p}}[\log L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t_i|Y=1); w^{(m)})|\mathcal{O}]$, where $w^{(m)} = (w_i^{(m)})_{i=1,\ldots,n}$.

In the M-step of the algorithm, $\mathbf{E}_{\mathrm{p}}[\log L_3(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1); w^{(m)})|\mathcal{O}]$ is maximized with respect to $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(t|Y=1)$, given $w^{(m)}$. Unlike in the standard Cox proportional hazards model, where the baseline hazard is seen as a nuisance parameter and eliminated in the procedure of estimating $\boldsymbol{\gamma}$, one cannot eliminate $\lambda_0(t|Y=1)$ in the Cox proportional hazards model embedded in the VMCF without losing information about $\boldsymbol{\beta}$. The main reason is that the hazard rate at the population level is no longer proportional [see (17) and (18)] and arguments similar to those leading to the Cox partial likelihood do not hold true anymore. Peng and Dear (2000), Sy and Taylor (2000) proposed a partial likelihood type method to estimate $\boldsymbol{\gamma}$ without specifying the nuisance parameter $\Lambda_0(t|Y=1)$. This involves the updating of the Aalen-Nelson estimator of $\Lambda_0(t|Y=1)$ at the mth iteration as given by

$$\widehat{\Lambda}_0(t|Y=1) = \sum_{i:t_i\le t} \frac{d_i}{\sum_{l\in R_i} w_l^{(m)} \exp\left(\widehat{\boldsymbol{\gamma}}^\top \mathbf{Z}_l\right)},$$

where d_i is the number of events at time t_i, irrespective of the cause, and R_i is the risk set at t_i. By substituting $\widehat{\Lambda}_0(t|Y=1)$ into $\log L_{32}(\boldsymbol{\gamma}, \lambda_0(t|Y=1); w^{(m)})$ we get the weighted partial likelihood of $\boldsymbol{\gamma}$, that is

$$\prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{\gamma}^\top \mathbf{Z}_i)}{\sum_{l\in R_i} w_l^{(m)} \exp(\boldsymbol{\gamma}^\top \mathbf{Z}_l)}\right]^{\mathbf{1}\{\Delta_i>0\}}.$$

To assure identifiability, we impose the zero-tail constraint (see Sy and Taylor 2000), that is, $\widehat{S}_0(t|Y=1)=0$ for $t>t_K$.

Appendix B: The standard errors of conditional (on $Y=1$) cumulative hazards from vertical modeling with a cure fraction

In this appendix we derive the formula for the standard error of the conditional (on $Y=1$) cause-specific cumulative hazard from vertical modeling with a cure fraction, when we omit the vector $\mathbf{Z}$ of covariates from the model of the latency component.

As a consequence, the vector of parameter describing VMCF is $(\boldsymbol{\beta}, \boldsymbol{\eta}, \lambda_\bullet(t_1|Y = 1), \ldots, \lambda_\bullet(t_K|Y = 1))$.

The Nelson–Aalen estimator of $\Lambda_\bullet(t|Y = 1)$, denoted by $\widehat{\Lambda}_\bullet(t|Y = 1)$, makes jumps of size $d\widehat{\Lambda}_\bullet(t|Y = 1)$ at event time points $0 = t_0 \leq t_1 < t_2 < \ldots < t_K < \infty$. An approximation of the covariance matrix of $(\boldsymbol{\beta}, d\widehat{\Lambda}_\bullet(t_1|Y = 1), \ldots, d\widehat{\Lambda}_\bullet(t_K|Y = 1))$ is given by the inverse of the observed full information matrix $\Im_{\boldsymbol{\beta}, d\widehat{\Lambda}_\bullet(\cdot|Y=1)}$.

Remark The matrix $\Im_{\boldsymbol{\beta}, d\widehat{\Lambda}_\bullet(\cdot|Y=1)}$ is not diagonal. Let $\boldsymbol{\Xi}_{\boldsymbol{\beta}, d\widehat{\Lambda}_\bullet(\cdot|Y=1)}$ denote an inverse of $\Im_{\boldsymbol{\beta}, d\widehat{\Lambda}_\bullet(\cdot|Y=1)}$ and let $\boldsymbol{\Xi}_{d\widehat{\Lambda}_\bullet(\cdot|Y=1)}$ be the sub-matrix of $\boldsymbol{\Xi}_{\boldsymbol{\beta}, d\widehat{\Lambda}_\bullet(\cdot|Y=1)}$ corresponding to $d\widehat{\Lambda}_\bullet(\cdot|Y = 1)$.

Relevant quantities for our purposes are the relative hazards $\pi_j(t)$; we model them as in formula (14).

Remark Often it will convenient to retain the system (14) and to work with the $r \times J$ Fisher information matrix of $\boldsymbol{\eta}^\top = \left(\eta_1, \ldots, \eta_J\right)^\top$, denoted by $\Im_{\boldsymbol{\eta}}$ which has rank $r(J-1)$ and, in particular, is not invertible. Let $\boldsymbol{\Xi}_{\boldsymbol{\eta}}$ denote a Moore-Penrose generalized inverse of $\Im_{\boldsymbol{\eta}}$.

We are interested to develop a formula for the $JK \times JK$ covariance matrix $\text{var}(\widehat{\boldsymbol{\Lambda}}(\cdot|Y = 1)) =: \boldsymbol{\Xi}_{\boldsymbol{\Lambda}}$ of the estimator $\widehat{\Lambda}_j(t|Y = 1) = \sum_{t_s \leq t} \widehat{\pi}_j(t_s)\widehat{\lambda}_\bullet(t_s|Y = 1) = \sum_{t_s \leq t} \widehat{\lambda}_{js,Y=1}$, where $\widehat{\lambda}_{js,Y=1} = \widehat{\pi}_j(t_s)\widehat{\lambda}_\bullet(t_s|Y = 1)$. First, we introduce some notation, as follows:

$$\boldsymbol{\tau} = (\boldsymbol{\eta}, \lambda_\bullet(t_1|Y = 1), \ldots, \lambda_\bullet(t_K|Y = 1))^\top,$$
$$\boldsymbol{\Lambda} = \left((\Lambda_j(t_1|Y=1))_{j=1,\ldots,J}, (\Lambda_j(t_2|Y=1))_{j=1,\ldots,J}, \ldots, (\Lambda_j(t_K|Y=1))_{j=1,\ldots,J}\right)^\top$$

and

$$\boldsymbol{\lambda} = \left(\lambda_{11,Y=1}, \lambda_{21,Y=1}, \ldots, \lambda_{J1,Y=1}, \lambda_{12,Y=1}, \ldots, \lambda_{J2,Y=1}, \lambda_{1K,Y=1}, \ldots, \lambda_{JK,Y=1}\right).$$

According to the Delta-method, we get

$$\boldsymbol{\Xi}_{\boldsymbol{\Lambda}} = \frac{\partial \boldsymbol{\Lambda}(\cdot|Y = 1)}{\partial \boldsymbol{\lambda}(\cdot|Y = 1)} \cdot \frac{\partial \boldsymbol{\lambda}(\cdot|Y = 1)}{\partial \boldsymbol{\tau}} \cdot \text{var}(\widehat{\boldsymbol{\tau}}) \cdot \left(\frac{\partial \boldsymbol{\Lambda}(\cdot|Y = 1)}{\partial \boldsymbol{\lambda}(\cdot|Y = 1)} \cdot \frac{\partial \boldsymbol{\lambda}(\cdot|Y = 1)}{\partial \boldsymbol{\tau}}\right)^\top. \quad (23)$$

It is straightforward to see that the matrix $\frac{\partial \boldsymbol{\Lambda}}{\partial \boldsymbol{\lambda}}$ of order JK is given by

$$\frac{\partial \boldsymbol{\Lambda}}{\partial \boldsymbol{\lambda}} = \begin{pmatrix} I_{J\times J} & 0 & 0 & & 0 \\ I_{J\times J} & I_{J\times J} & 0 & & 0 \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots & \\ & & & & \ddots & 0 \\ I_{J\times J} & I_{J\times J} & & \ldots & & I_{J\times J} \end{pmatrix} \quad (24)$$

where $I_{J\times J}$ is the identity matrix of order J. Also, we have

$$\frac{\partial \lambda_{js,Y=1}}{\partial \eta_{lu}} = -\pi_j(t_s)\pi_l(t_s)\lambda_\bullet(t_s|Y=1)B_u(t_s),$$

where $j, l \in \{1,\ldots,J\}$, $j \neq l$, $s \in \{1,\ldots,K\}$, $u \in \{1,\ldots,r\}$, and

$$\frac{\partial \lambda_{js,Y=1}}{\partial \eta_{ju}} = \pi_j(t_s)\big[1-\pi_j(t_s)\big]\lambda_\bullet(t_s|Y=1)B_u(t_s),$$

where $j \in \{1,\ldots,J\}$, $s \in \{1,\ldots,K\}$, $u \in \{1,\ldots,r\}$.

Moreover,

$$\frac{\partial \lambda_{js,Y=1}}{\partial \lambda_\bullet(t_v|Y=1)} = \pi_j(t_s)\delta_{s,v},$$

where $j \in \{1,\ldots,J\}$, $s, v \in \{1,\ldots,K\}$ and δ stands for the Kronecker delta.

We shall define, for $t \geq 0$, the $J \times J$ matrix $\boldsymbol{\Omega}(t)$ as follows:

$$\boldsymbol{\Omega}(t) = \begin{pmatrix} \pi_1(t) & 0 & \ldots & 0 \\ 0 & \pi_2(t) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \pi_J(t) \end{pmatrix} - \big(\pi_1(t), \pi_2(t), \ldots, \pi_J(t)\big)^\top \big(\pi_1(t), \pi_2(t), \ldots, \pi_J(t)\big),$$

and the r-vector

$$\boldsymbol{\alpha}(t) = \big(\lambda_\bullet(t|Y=1)B_1(t), \lambda_\bullet(t|Y=1)B_2(t), \ldots, \lambda_\bullet(t|Y=1)B_r(t)\big)^\top.$$

Setting

$$\boldsymbol{\Pi}(t_s) = \big(\pi_1(t_s), \pi_2(t_s), \ldots, \pi_J(t_s)\big)^\top,$$

a column vector of length J, for $s \in \{1,\ldots,K\}$, we get

$$\frac{\partial(\lambda_{1s,Y=1}, \lambda_{2s,Y=1}, \ldots, \lambda_{Js,Y=1})}{\partial(\eta_1, \eta_2, \ldots, \eta_J)} = \boldsymbol{\Omega}(t_s) \otimes (\boldsymbol{\alpha}(t_s))^\top, \quad s \in \{1,\ldots,K\}, \tag{25}$$

where $\otimes$ stands for the Kronecker product, and finally

$$\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\tau}} = \begin{pmatrix} \boldsymbol{\Omega}(t_1) \otimes (\boldsymbol{\alpha}(t_1))^\top & \boldsymbol{\Pi}(t_1) & 0 & 0 & & & 0 \\ \boldsymbol{\Omega}(t_2) \otimes (\boldsymbol{\alpha}(t_2))^\top & 0 & \boldsymbol{\Pi}(t_2) & 0 & & & 0 \\ & & & 0 & \ddots & \ddots & \\ \vdots & & \vdots & & \ddots & \ddots & \ddots \\ & & & & & \ddots & \ddots & 0 \\ \boldsymbol{\Omega}(t_K) \otimes (\boldsymbol{\alpha}(t_K))^\top & 0 & 0 & 0 & \ldots & 0 & \boldsymbol{\Pi}(t_K) \end{pmatrix} \tag{26}$$

which is a matrix of order $JK \times (Jr + K)$. As a result, we obtain

$$\boldsymbol{\Xi}_{\boldsymbol{\lambda}} = \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\tau}} \begin{pmatrix} \boldsymbol{\Xi}_{\boldsymbol{\eta}} & | & 0 \\ - & | & - \\ 0 & | & \boldsymbol{\Xi}_{d\widehat{\Lambda}_{\bullet}(\cdot|Y=1)} \end{pmatrix} \left(\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\tau}}\right)^{\top}. \tag{27}$$

In conclusion, using (23), (24) and (27), we have that

$$\boldsymbol{\Xi}_{\boldsymbol{\Lambda}} = \begin{pmatrix} \mathbf{W}_1 \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_1 + \widetilde{\boldsymbol{\Pi}}_1 & \mathbf{W}_1 \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_2 + \widetilde{\boldsymbol{\Pi}}_1 & \dots & \mathbf{W}_1 \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_K + \widetilde{\boldsymbol{\Pi}}_1 \\ \mathbf{W}_2 \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_1 + \widetilde{\boldsymbol{\Pi}}_1 & \mathbf{W}_2 \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_2 + \widetilde{\boldsymbol{\Pi}}_2 & \dots & \mathbf{W}_2 \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_K + \widetilde{\boldsymbol{\Pi}}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_K \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_1 + \widetilde{\boldsymbol{\Pi}}_1 & \mathbf{W}_K \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_2 + \widetilde{\boldsymbol{\Pi}}_2 & \dots & \mathbf{W}_K \boldsymbol{\Xi}_{\boldsymbol{\eta}} \mathbf{W}_K + \widetilde{\boldsymbol{\Pi}}_K \end{pmatrix}, \tag{28}$$

where

$$\mathbf{W}_k = \sum_{s=1}^{k} \boldsymbol{\Omega}(t_s) \otimes (\boldsymbol{\alpha}(t_s))^{\top}, \quad k \in \{1, \dots, K\},$$

and

$$\widetilde{\boldsymbol{\Pi}}_{kl} = \sum_{l=1}^{k} \boldsymbol{\Xi}_{d\widehat{\Lambda}_{\bullet}(\cdot|Y=1)} \otimes \left(\sum_{s=1}^{k} \boldsymbol{\Pi}(t_s)(\boldsymbol{\Pi}(t_l))^{\top} \right), \quad k, l \in \{1, \dots, K\}.$$

References

Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer, Berlin

Andersson TML, Dickman PW, Eloranta S, Lambert PC (2011) Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. BMC Med Res Methodol 11:1–11

Andrae B, Andersson TML, Lambert PC, Kemetli L, Silverdal L, Strander B, Ryd W, Dillner J, Tolnber S, Sparen P (2012) Screening and cervical cancer cure: population based cohort study. Br Med J 344:e900

Basu S, Tiwari TC (2010) Breast cancer survival, competing risks and mixture cure model: a Bayesian analysis. J R Stat Soc Ser A 173:307–329

Canty A, Ripley B (2008) Boot: Bootstrap r (s-plus) functions. R package version 1.2-34

Chao EC (1998) Gibbs sampling for long-term survival data with competing risks. Commun Stat Theory Methods 54:350–366

Chen CH, Tsay Y, Wu Y, Horng C (2013) Logistic aft location-scale mixture regression models with nonsusceptibility for left-truncated and general interval-censored data. Stat Med 32:4285–4305

Choi KC, Zhou X (2002) Large sample properties of mixture models with covariates competing risks. J Multivar Anal 82:331–366

Corbiere F, Commenges D, Taylor JMG, Joly P (2009) A penalized likelihood approach for mixture cure models. Stat Med 28:510–524

Farewell VT (1986) Mixture models in survival analysis: are they worth the risk? Can J Stat 14:257–262

Jia X, Sima CS, Brennan MF, Panageas KS (2013) Cure models for the analysis of time-to-event data in cancer studies. J Surg Oncol 108:342–347

Kim S, Zeng D, Li Y, Spiegelman D (2013) Joint modeling of longitudinal and cure survival data. J Stat Theory Pract 7:324–344

Klebanov L, Yakolev A (2007) A new approach to testing for sufficient follow-up in cure-rate analysis. J Stat Plan Inference 137:3557–3569
Kuk AYC, Chen C (1992) A mixture model combining logistic regression with proportional hazards regression. Biometrika 79:531–541
Larson MG, Dinse GE (1985) A mixture model for the regression analysis of competing risks data. Appl Stat 34:201–211
Laska EG, Meisner MJ (1992) Nonparametric estimation and testing in a cure model. Biometrics 48:1223–1234
Li CS, Taylor JMG (2002) A semiparametric accelerated failure time cure model. Stat Med 21:3235–3247
Maller R, Zhou S (1994) Testing for sufficient follow-up and outliers in survival data. J Am Stat Assoc 89(428):1499–1506
Ng SK, McLachlan GJ (1998) On modifications to the long-term survival mixture model in the presence of competing risks. J Stat Comput Simul 61:77–96
Nicolaie MA, Putter H, van Houwelingen JC (2010) Vertical modeling: a pattern mixture approach for competing risks data. Stat Med 29:1190–1205
Nicolaie MA, Putter H, van Houwelingen JC (2015) Vertical modeling: analysis of competing risks data with missing causes of failure. Stat Methods Med Res 24(6):891–908
Peng Y (2003) Estimating baseline distribution in proportional hazards cure models. Comput Stat Data Anal 42:187–201
Peng Y, Dear KBG (2000) A nonparametric mixture model for cure rate estimation. Biometrics 56:2236–243
Peng Y, Taylor JMG (2013) Cure models. In: Klein JP, van Houwelingen HC, Ibrahim JG, Scheike TH (eds) Handbook of survival analysis. Chapman and Hall CRC, Boca Raton, pp 113–134
Peng Y, Taylor JMG (2017) Residual-based model diagnosis methods for mixture cure models. Biometrics 73:495–505
R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Scolas S, Ghouch AE, Legrand C, Oulhaj A (2016a) Diagnostic checks in mixture cure models with interval-censoring. Stat Methods Med Res. https://doi.org/10.1177/0962280216676502
Scolas S, Ghouch AE, Legrand C, Oulhaj A (2016b) Variable selection in a flexible parametric mixture cure model with interval-censored data. Stat Med 35:1210–1225
Sy JP, Taylor JMG (2000) Estimation in a Cox proportional hazards cure model. Biometrics 56:227–236
Taylor JMG (1995) Semi-parametric estimation in failure time mixture models. Biometrics 51:899–907
Taylor JMG, Liu N (2007) Statistical issues involved with extending standard models. In: Nair V (ed) Advances in statistical modeling and inferences: essays in honor of Kjell A. World Scientific Publishing Company, Singapore, Doksum, pp 299–311
Wei L (1992) The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. Stat Med 11:1871–1879
Yamaguchi K (1992) Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of permanent employment in Japan. J Am Stat Assoc 87:284–292
Yu M, Law NJ, Taylor JMG, Sandler HM (2004) Joint longitudinal survival-cure models and their application to prostate cancer. Stat Sin 14:835–862
Yu XQ, De Angelis R, Andersson TML, Lambert PC, Dickman P (2013) Estimating the proportion cured of cancer: some practical advice for users. Cancer Epidemiol 37:836–842