Pakistan Journal of Statistics and Operation Research

A competing risks cure frailty model: An application to relapse-free survival of breast cancer patients



Vahid Ghavami¹, Mahmood Mahmoudi², Abbas Rahimi Foroushani^{*3}, Hossein Baghishani⁴, Mehdi Yaseri⁵, Fatemeh Homaei Shandiz⁶

* Corresponding Author

1. Department of Epidemiology and Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran, Email: Ghavamiv@mums.ac.ir

2. Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, Email: mahmoodim@tums.ac.ir

3. Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, Email: rahimifo@tums.ac.ir

4. Department of Applied Mathematics, Shahrood University of Technology, Shahrood, Iran, Email: hbaghishani@shahroodut.ac.ir

5. Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, Email: myaseri@tums.ac.ir

6. Cancer Research Center, Mashhad University of Medical Sciences, Mashhad, Iran, Email: homaeef@mums.ac.ir

Abstract

If there is unobserved heterogeneity among susceptible patients in competing risks cure models, applying the methods that do not consider this heterogeneity may lead to invalid results. Therefore, this study aimed to introduce a model to cover the above properties of survival studies. We introduced a unified model by combining a parametric mixture cure gamma frailty model and vertical modeling of competing risks. We obtained estimates of parameters by an iterative method and Laplace transform technique. Then, we calculated the cumulative incidence functions (CIFs) and related confidence bounds using a bootstrap approach. We conducted an extensive simulation study to evaluate the performance of the proposed model. The results of the simulation study showed the superior performance of our proposed competing risks cure frailty model. Finally, we applied the proposed method to analyze a real dataset of breast cancer patients.

Key Words: Survival analysis; Competing risks; Cumulative incidences; Frailty model; Mixture cure model.

1. Introduction

Survival modeling is a tool in biostatistics and epidemiology that can help one to predict the risk of disease. In medical studies, identifying the risk of failure for a given disease is remarkable for patients as well as physicians; this would benefit patients to choose the lifestyle, and it would be guidelines to physicians selecting a specific treatment approach. The need to know these risk factors requires the analysis of outcomes to be specific to that disease. Still, the prior occurrence of different types of events may change the probability of observing the event of interest. From the statistical perspective, the analysis is incorporated in the competing risks framework in this situation, and the usual techniques may give distorted results (Pepe and Mori, 1993). For modeling competing risks, several approaches such as the cause-specific model by Prentice et al. (1978), the Larson and Dinse's model by Larson and Dinse (1985), the Fine and Gray's model by Fine and Gray (1999), and the vertical modeling by Nicolaie, van Houwelingen, and Putter (2010) have been introduced. Competing risks models usually assume that all patients are susceptible; if there is a sufficient follow-up, all patients will experience any of the possible outcomes (Pintilie, 2011). However, due to advances in the early detection of diseases and their treatment, some patients may be long-term event-free survivors (cured fraction). When it is confirmed that there is a cured fraction in the study population, typical survival models

lead to overestimating the survival of the susceptible subjects, and deployment cure models are recommended (Corbière et al., 2009). A popular and well-known class of cure models is the mixture cure model (Berkson and Gage, 1952; Boag, 1949; Farewell, 1982; Kuk and Chen, 1992). In a mixture cure model, the patients consist of two parts, including susceptible patients (uncured patients) and long-term survivor patients (cured patients). The relation between population survival time (S(t)) and the survival time of susceptible (uncured) patients (S_u(t)) is defined as follows

$$S(t) = \pi . S_u(t) + 1 - \pi.$$
(1)

In the above equation, $1-\pi$ is the probability of being cured, and it is evident that the survival probability of cured patients is embedded as one. In equation (1), $S_u(t)$ is referred to as the latency and π is also referred to the incidence part of the mixture model.

Modeling the competing risks in the presence of cured fractions has not yet developed very well. Gee (2004) extended a cause-specific competing risks model that a subpopulation of patients could be cured of the event of interest. He introduced two models based on exponential distribution and proportional hazards and employed expectation– maximization (EM) algorithm and weighted risk set approaches to estimate parameters in a semi-parametric model. Basu and Tiwari (2010) unified the mixture cure and competing risks models in a Bayesian framework. Several authors extended the competing risks model of Larson and Dinse (1985) to a competing risks cure model. It is noteworthy that in the approach of Larson and Dinse, the joint distribution of time of failure (T) and failure type (D) decomposes as: P(T, D) = P(T|D) P(D)

$$P(T,D) = P(T|D).P(D).$$
 (2)

For example, Cai (2013) presented a new estimation method for a semi-parametric mixture cure model with competing risks data. A disadvantage of the proposed approach of Larson and Dines (1985) is from an interpretational point of view. In this approach, the cause of failure is determined from the beginning, but its estimated distribution will depend on the length of follow-up, which is hard to adapt with the implied existence of such a distribution from the outset (Nicolaie, van Houwelingen, and Putter, 2010). Nicolaie, Taylor, and Legrand (2018) combined competing risks and cured fractions in vertical modeling in which the cured fraction meant the individuals who are immune to all causes of failures. In the idea of vertical modeling, a new decomposition of the joint distribution of time (T) and cause of failure (D) is utilized as:

$$P(T,D) = P(T).P(D|T).$$
(3)

From an interpretation point of view, the decomposition (3) does not have the disadvantage of the decomposition (2). Interestingly, the decomposition (3) consists of two observable quantities: the total hazard and the relative cause-specific hazard. In this approach, the marginal event time distribution, P(T), can be estimated considering covariate effects on the total hazard rate using a semiparametric or a parametric regression model for time-to-event data, treating events of any type as failures.

On the other hand, some unobserved information exists among individuals in the medical and epidemiological studies, which cannot be explained via observed covariates in survival analysis; so frailty models are recommended in this situation. In a univariate frailty model, the hazard at time t for a person with frailty w is:

$$\lambda(t) = W\lambda_0(t)\exp\left(\gamma^T X\right),\tag{4}$$

where W follows a probability distribution with mean equal to one and variance σ^2 , γ^T is a vector of coefficients, X is a vector of covariates with the same dimension of coefficients and $\lambda_0(t)$ is the baseline hazard function. Patients with a high W value tend to have a high rate of the event; thus, the variance of this distribution is interpretable as a measure of heterogeneity across the population (Wienke, 2010). Hence, developing a competing risks model to include cured fraction and frailty is critical for analyzing a dataset with such properties. In this study, we intend to extend the modeling of competing risks when there are cured fractions as well as heterogeneity due to unobserved covariates. The paper is organized as follows: In Sections 2 to 6, we give details of our method and an extensive simulation scenario. Section 7 presents the application of the proposed model on the real dataset of breast cancer patients. Finally, in Sections 8 and 9, some points for discussion and concluding remarks are made, respectively.

2. Competing risks-cure-frailty model

In a survival study, consider n patients who may experience one of the K competing events or be right-censored. Furthermore, suppose that a fraction of individuals are long-term survivors (cured fraction) so that they do not experience any of the competing risks events by the end of sufficient follow-up. In these circumstances, patients can belong to any of the two subpopulations, susceptible and non-susceptible groups. Moreover, there may be heterogeneity due to unobserved covariates between susceptible patients so that patients with similar characteristics experience the failure at different times. To handle the above conditions, we want to develop the vertical modeling of competing risks with cured fraction, earlier proposed by Nicolaie, Taylor, and Legrand (2018). The vertical modeling of competing risks with cured fraction includes two parts, namely the incidence and latency part. In the latency part, we determine whether an individual is susceptible, in the sense that she experiences an event finally; in the latency

(6)

part, conditional on an individual being susceptible, we determine when and which event might occur. We want to extend the vertical modeling approaches to the mixture cure frailty framework for the latency part. Let Y be a binary variable so that Y=1 refers to a susceptible patient and Y=0 refers to a non-susceptible patient. If \tilde{T} denotes the time of the event, D is the type of the event where D=1,2,...,k and C is the right-censoring time, the observed data for an individual i is $O_i = (T_i, \Delta_i, Q_i)$ where $T_i = \min(\tilde{T}_i, C_i)$ is the observed time, $\Delta_i = 1{\{\tilde{T}_i < C_i\}}D_i$ is an indicator for the type of event which is equal to zero in the case of censoring and Q_i is a vector of covariates measured at baseline for i=1,2,...,n. We assume that data from different patients are independent; also, the time of events and censoring times are independent given Q. Similar to mixture cure models, equation (1), the incidence part is specified by P(Y) so that P(Y = 1) = p, and 1 - p represents the fraction of patients who are cured. Following the vertical modeling approach to competing risks, when W is a frailty term, the conditional joint distribution P(T, D|Y = 1, W = w) is decomposed as

$$P(T,D|Y = 1, W = w) = P(T|Y = 1, W = w). P(D|T, Y = 1, W = w),$$
(5)

for the latency part (Nicolaie, van Houwelingen, and Putter, 2010). If the survival time is continuous, we can define the conditional total or overall caused hazard by $P(t < T < t + \Delta t | T > t | V = 1) | W = w)$

$$h_{\star}(t|Y=1, W=w) = \lim_{\Delta t \to 0} \frac{P(t \le t \le t + \Delta t) | t \ge t, Y=1, W=w)}{\Delta t},$$

rm by

and its cumulative form by

$$H_{*}(t|Y = 1, W = w) = \int_{0}^{t} h_{*}(u|Y = 1, W = w) du.$$

Equation (6) specifies the conditional failure distribution P(T|Y = 1, W = w) in equation (5), and the conditional survival function of susceptible individuals is given by:

$$S_*(t|Y = 1, W = w) = \exp(-H_*(t|Y = 1, W = w)).$$

To estimate P(D|T, Y = 1, W = w), if $h_j(t|Y = 1, W = w)$ is the conditional cause-specific hazard of cause j, we define the conditional relative cause-specific hazard as h(t|Y = 1, W = w)

$$\pi_{j}(t|Y = 1, W = w) = \frac{h_{j}(t|Y = 1, W = w)}{h_{*}(t|Y = 1, W = w)}$$

$$= \frac{Wh_{j}(t|Y = 1)}{Wh_{*}(t|Y = 1)} = \frac{h_{j}(t|Y = 1)}{h_{*}(t|Y = 1)} = \pi_{j}(t|Y = 1), \quad j = 1, \dots, K.$$
(7)

So we assume that W is conditionally independent of D|T,Y=1 and Y; therefore, equation (5) can be written as P(T,D|Y = 1, W = w) = P(T|Y = 1, W = w). P(D|T,Y = 1). (8)

The relative cause-specific hazard deals with cause and failure time. Hence, its estimation involves only observed susceptible individuals as

$$\pi_j(t|Y=1) = P(D=j|\tilde{T}=t, Y=1) = P(D=j|\tilde{T}=t) = \pi_j(t), \qquad j=1,2,\dots,k.$$

new part is completely specified by $\pi_i(t)$ and $h_i(t|Y=1, W=w)$. The conditional cumul

Thus, the latency part is completely specified by $\pi_j(t)$ and $h_*(t|Y = 1, W = w)$. The conditional cumulative incidence function of cause j can be estimated as

$$F_{j}(t|Y = 1, W = w) = \int_{0}^{t} h_{j}(u|Y = 1, W = w). S(u|Y = 1, W = w)du = \int_{0}^{t} h_{*}(u|Y = 1, W = w). \pi_{j}(u). S(u|Y = 1, W = w)du, \qquad j = 1, ..., K.$$
(9)

The population survival function can also be estimated as

$$S_{pop}(t|W = w) = P(Y = 1). P(T \ge t|Y = 1, W = w) + P(Y = 0). P(T \ge t|Y = 0, W = w)$$
$$= p.S(t|Y = 1, W = w) + 1 - p.$$

It is obvious that the population survival function is improper in the sense that

$$\lim_{t \to \infty} S_{pop}(t|W=w) = 1 - p$$

3. Assessing covariates effects on latency and incidence

For assessing the effects of covariates on the incidence, p(X) = P(Y = 1|X), where $X \subseteq Q$ we postulate a logistic model as

$$\ln \frac{p(X)}{1 - p(X)} = \beta^T \boldsymbol{X}^*, \tag{10}$$

where $X^* = (1, X)$ and β^T denotes a vector of unknown regression parameters.

For assessing the effects of covariates on the latency, we need two models; a model for conditional total hazard and a model for relative cause-specific hazard. For modeling the conditional total hazard, we postulate a Weibull Gamma frailty model as:

$$h_*(t|Y=1, W=w) = h_{0*}(t|Y=1). w. \exp(\boldsymbol{\gamma}^T. \boldsymbol{Z}),$$
 (11)

 $h_*(t|Y = 1, W = w) = h_{0*}(t|Y = 1). w. \exp(\gamma^t \cdot Z),$ (11) where, $Z \subseteq Q, \gamma^T$ stands for a vector of unknown regression parameters, $h_{0*}(t|Y = 1)$ follows a Weibull distribution with shape parameter k and scale parameter λ and W has a gamma distribution with mean equal to one and variance equal to σ^2 . Furthermore, for the relative cause-specific hazard we specify

$$\pi_j(t|\boldsymbol{U}) = \frac{\exp\left(\eta_j^T \boldsymbol{B}(t) + \nu_j^T \boldsymbol{U}\right)}{\sum_{i=1}^K \exp\left(\eta_i^T \boldsymbol{B}(t) + \nu_i^T \boldsymbol{U}\right)}$$
(12)

where, B(t) is an r-vector of pre-specified time functions, $U \subseteq Q$ is a vector of independent variables, η_i^T and v_i^T stand for m-vectors of unknown regression parameters and j=1,...,K is jth cause of failure. We denote all regression parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_K, \boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_K)$ by $\boldsymbol{\theta}$.

4. Likelihood

For convenience, we omit covariates for a moment. It is worth noting that we assumed W is conditionally independent of D|T, Y = 1 and Y. The observed likelihood (conditional on frailty term, W_i) is the product of contributions of patients from two categories: a patient who experiences an event with type i at time t_i contributes

$$P(\tilde{T}_{i} = t_{i}, D_{i} = j, Y_{i} = 1 | w_{i}) = P(\tilde{T}_{i} = t_{i}, Y_{i} = 1 | w_{i}). P(D_{i} = j | \tilde{T}_{i} = t_{i}, Y_{i} = 1, w_{i})$$

= $P(\tilde{T}_{i} = t_{i}, Y_{i} = 1 | w_{i}). P(D_{i} = j | \tilde{T}_{i} = t_{i}, w_{i}) = P(\tilde{T}_{i} = t_{i}, Y_{i} = 1 | w_{i}). P(D_{i} = j | \tilde{T}_{i} = t_{i})$
= $P(\tilde{T}_{i} = t_{i} | Y_{i} = 1, w_{i}). P(Y_{i} = 1 | w_{i}). P(D_{i} = j | \tilde{T}_{i} = t_{i})$
= $P(\tilde{T}_{i} = t_{i} | Y_{i} = 1, w_{i}). P(Y_{i} = 1). P(D_{i} = j | \tilde{T}_{i} = t_{i}),$

and a patient who is censored at time t_i contributes

$$P(\tilde{T}_i > t_i | w_i) = P(Y_i = 1 | w_i) . P(\tilde{T}_i > t_i | Y_i = 1, w_i) + P(Y_i = 0 | w_i)$$

= $P(Y_i = 1) . P(\tilde{T}_i > t_i | Y_i = 1, w_i) + P(Y_i = 0).$

Therefore, based on the non-informative censoring assumption, the observed likelihood is proportional to

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

In terms of cause-specific hazard and conditional total hazard, the observed likelihood can be written as

$$L = \prod_{i=1}^{n} \left[p_i \cdot h_*(t_i | Y_i = 1, w_i) \cdot \exp(-H_*(t_i | Y_i = 1, w_i)) \prod_{j=1}^{K} \pi_j(t_i)^{1\{D_i = j\}} \right]^{I_1 \cup I_2 \cup J_3}$$

$$\cdot \prod_{i=1}^{n} \left[p_i \cdot \exp(-H_*(t_i | Y_i = 1, w_i)) + (1 - p_i) \right]^{1\{D_i = 0\}}$$
(13)

which can be separated in the following way:

$$L(\boldsymbol{\theta}|w) = L_1(\boldsymbol{\beta}, \boldsymbol{\gamma}|w).L_2(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_K, \boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_K),$$
(14)

where

and

$$L_{1}(\boldsymbol{\beta}, \boldsymbol{\gamma}|w) = \prod_{i=1}^{n} \left[p_{i} \cdot h_{*}(t_{i}|Y_{i} = 1, w_{i}) \cdot \exp(-H_{*}(t_{i}|Y_{i} = 1, w_{i})) \right]^{1\{D_{i} > 0\}}$$
$$\cdot \prod_{i=1}^{n} \left[p_{i} \cdot \exp(-H_{*}(t_{i}|Y_{i} = 1, w_{i})) + (1 - p_{i}) \right]^{1\{D_{i} = 0\}}$$
$$L_{2} = \prod_{i=1}^{n} \prod_{j=1}^{K} \pi_{j}(t_{i})^{1\{D_{i} = j\}} \cdot$$

5. Estimation

We wish to estimate the regression parameters in the incidence, latency, and relative cause-specific hazard parts of the model. In cure models, the EM algorithm is usually used to estimate the parameters (Peng and Dear, 2000; Sy and Taylor, 2000); when the baseline hazard is specified, the Newton-Raphson iterative procedure can be applied (Rondeau et al., 2013). We obtained the asymptotic variance of regression parameters by inverting the observed information matrix, and we used the asymptotic properties of maximum likelihood estimators for inferences. An interesting fact in the likelihood (14) is that the maximum likelihood estimations of the parameters can be obtained by maximizing the $L_1(\beta, \gamma, \sigma^2)$ and $L_2(\eta_1, ..., \eta_K, \nu_1, ..., \nu_K)$ independently. Therefore, to estimate the corresponding parameters in $L_2(\eta_1, ..., \eta_K, \nu_1, ..., \nu_K)$, standard softwares such as the glm function in R can be used. We employed the optim function in R and utilized the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method to estimate the CIFs and obtained their confidence intervals using the bootstrapping method. In the bootstrapping method, we re-sampled the original data 200 times and estimated the CIFs in each iteration. We used a quantile-based bootstrap method to compute 95% confidence bands for CIFs.

6. Simulation

We conducted a simulation study of the competing risks-cure-frailty model to assess the performance of the estimators when the proposed model is correct and compare the competing risks-cure-frailty model with a competing risks cure one (without frailty term). We assumed that there are 500 patients, each of whom can experience event 1 (the event of interest) or event 2 (competing event) or can be subjected to right-censoring. We also assumed a fraction of patients could be cured (which means they are immune to two causes of failure). The study cut-off time is 16 years after inclusion in the study. For simplicity, we only considered two covariates, X and Z, for the incidence and latency parts, respectively. The covariates X and Z were both generated from a Bernoulli distribution with the probability of success equal to 0.5. For each patient (i), the survival data based on the competing risks-cure-frailty model were generated as follows:

The censoring times C_i were generated for each patient from a uniform distribution with a maximum of 16 years.

- 1. The frailty term (W_i) was generated from a gamma distribution with mean equals one and variance equals σ^2 .
- 2. The cure status was generated from a Bernoulli distribution with the probability

$$1 - \pi_i = 1 - P(Y_i = 1 | X_i = x_i) = 1 - \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

3. The overall failure time (\tilde{T}_i) was generated from a Weibull-Gamma frailty model as

$$h_*(t_i|Y_i = 1, W_i = w_i, Z_i = z_i) = h_{0*}(t_i|Y_i = 1).w_i.\exp(\gamma z_i),$$

where

$$W_i \sim \text{Gamma}\left(\text{shape} = \frac{1}{\sigma^2}, \text{scale} = \sigma^2\right)$$

and

$$h_{0*}(t_i|Y_i = 1) = \lambda k t_i^{k-1}$$
, $k > 0$, $\lambda > 0$.

	cure		true	competing	g risks cure	competing r	isks cure model
σ^2	rate for	parameter	value	frailty	model	competing i	
	X=1			bias	MSE	bias	MSE
		β_0	-0.9	0.0023	0.0199	-0.0018	0.0200
		β_1	2.0	0.0041	0.0406	-0.0004	0.0402
	25%	γ	0.5	0.0107	0.0350	-0.1343	0.0439
	2370	k	2.0	0.0205	0.0325	-0.5001	0.2574
		λ	0.5	0.0013	0.0057	-0.0538	0.0053
		σ^2	0.5	0.0050	0.0346	-	-
		β_0	1.0	0.0146	0.0199	0.0061	0.0194
		β_1	-1.0	-0.0186	0.0357	-0.0145	0.0352
0.5	500/	γ	0.5	0.0122	0.0304	-0.1343	0.0389
0.5	50%	k	2.0	0.0186	0.0284	-0.5074	0.2639
		λ	0.5	0.0012	0.0050	-0.0532	0.0048
		σ^2	0.5	0.0074	0.0328	-	-
		β_0	1.0	0.0052	0.0211	-0.0012	0.0207
	000/	β_1	-2.4	-0.0127	0.0483	-0.0050	0.0480
		γ	0.5	0.0137	0.0388	-0.1379	0.0453
	80%	k	2.0	0.0232	0.0427	-0.5007	0.2655
		λ	0.5	0.0021	0.0060	-0.0515	0.0051
		σ^2	0.5	0.0067	0.0430	-	-
		β_0	-0.9	-0.0082	0.0194	-0.0326	0.0197
		β_1	2.0	0.0248	0.0476	-0.0196	0.0427
	200/	γ	0.5	0.0107	0.0544	-0.2233	0.0763
	20%	k	2.0	0.0350	0.0499	-0.7443	0.5587
		λ	0.5	0.0125	0.0097	-0.0855	0.0094
		σ^2	1	0.0388	0.1302	-	-
		β_0	1.0	0.0078	0.0241	-0.0559	0.0233
		β_1	-1.0	-0.0144	0.0418	0.0154	0.0376
1	500/	γ	0.5	0.0221	0.0486	-0.2261	0.0731
1	50%	k	2.0	0.0370	0.0440	-0.7486	0.5643
		λ	0.5	0.0093	0.0082	-0.0849	0.0090
		σ^2	1	0.0475	0.1107	-	-
		β_0	1.0	0.0209	0.0260	-0.0456	0.0234
		β_1	-2.4	-0.0303	0.0530	0.0138	0.0486
	000/	γ	0.5	0.0152	0.0587	-0.2207	0.0751
	80%	k	2.0	0.0411	0.0579	-0.7463	0.5624
		λ	0.5	0.0129	0.0108	-0.0856	0.0095
		σ^2	1	0.0524	0.1443	-	-

 Table 1: Estimated bias and mean squared error (MSE) of regression parameters based on simulated data in the competing risks cure frailty model

	cure		true	competing	g risks cure	competing risks cure model		
σ^2	rate for	parameter	value	frailty	model	competing i	isks cure model	
	X=1	-		bias	MSE	bias	MSE	
		β_0	-0.9	-0.0197	0.0222	-0.0681	0.0259	
		β_1	2.0	0.0301	0.0554	-0.0249	0.0394	
	2004	γ	0.5	0.0136	0.0769	-0.2772	0.1161	
	20%	k	2.0	0.0325	0.0696	-0.8460	0.7206	
		λ	0.5	0.0374	0.0180	-0.1072	0.0142	
		σ^2	1.5	0.0299	0.3053	-	-	
	500/	β_0	1.0	-0.0013	0.0370	-0.1493	0.0420	
		β_1	-1.0	-0.0089	0.0470	0.0602	0.0399	
15		γ	0.5	-0.0045	0.0613	-0.2717	0.0963	
1.5	50%	k	2.0	0.0310	0.0565	0.8520	0.7289	
		λ	0.5	0.0453	0.0142	-0.1080	0.0133	
		σ^2	1.5	0.0186	0.2742	-	-	
		β_0	1.0	0.0123	0.0431	-0.1488	0.0446	
		β_1	-2.4	-0.0331	0.0663	0.0793	0.0577	
	800/	γ	0.5	0.0199	0.0850	-0.2616	0.0999	
	80%	k	2.0	0.0431	0.0832	-0.8502	0.7268	
		λ	0.5	0.0309	0.0239	-0.1103	0.0144	
		σ^2	1.5	0.0861	0.4357	-	-	

Table 1 (continued): Estimated bias and mean squared error (MSE) of regression parameters based on simulated data in the competing risks cure frailty model

- 4. If a patient is cured, then the overall failure time is $\tilde{T}_i = 200$, else the overall time is the value of \tilde{T}_i in step (4).
- 5. The overall observed time for patient *i* is $T_i = \min(\tilde{T}_i, C_i)$.
- 6. After generating overall failure times, the relative cause-specific hazards are constant in the intervals: (0,1], (1,2], and $(2,\infty)$ such that the favoring cause 2 over cause 1 was $\pi_1(t)=0.50$, 0.30, 0.20 in the intervals. In other words, based on the equation (12), the values of cause-specific relative hazards can be specified by the following equations:

$$\pi_{1}(t) = \frac{\exp\left(0.01 \times \mathbf{1}_{(0,1]}(t) - 0.85 \times \mathbf{1}_{(1,2]}(t) - 1.39 \times \mathbf{1}_{(2,16]}(t)\right)}{1 + \exp\left(0.01 \times \mathbf{1}_{(0,1]}(t) - 0.85 \times \mathbf{1}_{(1,2]}(t) - 1.39 \times \mathbf{1}_{(2,16]}(t)\right)}$$
$$\pi_{2}(t) = 1 - \pi_{1}(t).$$

The parameter values for incidence were specified as $(\beta_0, \beta_1) \in \{(-0.9, 2), (1, -1), (1, -2.4)\}$, leading to three scenarios with various amounts of cure proportions, 25, 50 and 80% for X = 1, respectively. The parameter values for latency were specified as $\gamma = 0.5$, $\lambda = 0.5$, k = 2, $\sigma^2 \in \{0.5, 1, 1.5\}$, leading to three scenarios with various heterogeneity amounts. Since the parameters in relative cause-specific hazard were independently estimated from other parameters and we used the standard software (glm function in R) to estimate these parameters, we did not include these parameters in the simulation. We fitted a competing risks-cure-frailty model and a competing risks-cure model to the simulated data. In the competing risks-cure-frailty model, the incidence and latency components of the model were assumed to have the same form as the correct model. In the competing risks-cure model, the incidence part and the relative cause-specific hazard had the same form as the correct model, but for the conditional (on Y = 1) total hazard, we employed a Weibull model (without frailty term).

2	cure		cure with a	frailty Model	cure without	t a Frailty Model
σ^2	rate	time	bias	RMSE	bias	RMSE
		0.25	0.0000	0.0045	0.0105	0.0117
		1.00	0.0002	0.0229	-0.0098	0.0236
	20%	2.00	0.0013	0.0259	-0.0083	0.0273
		5.00	0.0012	0.0261	-0.0038	0.0261
		10.00	0.0013	0.0261	-0.0047	0.0262
		0.25	0.0000	0.0045	0.0072	0.0080
		1.00	0.0003	0.0164	-0.0063	0.0167
0.5	50%	2.00	0.0002	0.0194	-0.0063	0.0202
		5.00	0.0002	0.0196	-0.0033	0.0197
		10.00	0.0010	0.0197	-0.0039	0.0198
		0.25	0.0000	0.0014	0.0105	0.0117
		1.00	0.0004	0.0090	-0.0021	0.0087
	80%	2.00	0.0003	0.0114	-0.0021	0.0113
		5.00	0.0002	0.0119	-0.0009	0.0117
		10.00	0.0003	0.0119	-0.0011	0.0117
		0.25	0.0002	0.0050	0.0107	0.0119
		1.00	0.0018	0.0232	-0.0133	0.0245
	20%	2.00	0.0011	0.0253	-0.0171	0.0300
		5.00	0.0010	0.0254	-0.0073	0.0262
		10.00	0.0010	0.0256	-0.0097	0.0268
		0.25	0.0000	0.0030	0.0103	0.0110
		1.00	0.0003	0.0154	-0.0100	0.0170
1	50%	2.00	0.0001	0.0175	-0.0124	0.0210
		5.00	-0.0002	0.0181	-0.0058	0.0187
		10.00	-0.0002	0.0183	-0.0073	0.0192
		0.25	0.0000	0.0015	0.0040	0.0045
		1.00	-0.0004	0.0085	-0.0043	0.0086
	80%	2.00	-0.0005	0.0106	-0.0052	0.0111
		5.00	-0.0006	0.0114	-0.0027	0.0113
		10.00	-0.0006	0.0116	-0.0033	0.0116
		0.25	0.0009	0.0053	0.0111	0.0123
		1.00	0.0040	0.0235	-0.0118	0.0204
	20%	2.00	0.0034	0.0257	-0.0148	0.0250
		5.00	0.0024	0.0257	-0.0070	0.0222
		10.00	0.0019	0.0260	-0.0085	0.0227
		0.25	0.0005	0.0032	0.0108	0.0114
		1.00	0.0024	0.0148	-0.0110	0.0171
1.5	50%	2.00	0.0021	0.0169	-0.0138	0.0212
		5.00	0.0013	0.0173	-0.0066	0.0184
		10.00	0.0009	0.0176	-0.0083	0.0191
		0.25	0.0000	0.0016	0.0042	0.0047
		1.00	-0.0002	0.0082	-0.0046	0.0085
	80%	2.00	-0.0004	0.0099	-0.0056	0.0109
		5.00	-0.0006	0.0107	-0.0029	0.0109
		10.00	-0.0008	0.0109	-0.0036	0.0113

Table 2: Estimated bias and root mean squared error (RMSE) of CIF (for cause 2) based on simulated data in the competing risks cure frailty model

After estimating the regression parameters, the cumulative incidences of j = 2 for the susceptible patient *i* with Z = 1 and X = 1 were estimated at time points {0.25, 1, 2, 5, 10}. The bias and MSE of the regression parameters as well

as the bias and root of MSE (RMSE) of the CIFs were reported under various heterogeneity and cured fractions. We ran each scenario 1000 times. The results of the simulation study showed the efficiency of the competing risks-cure-frailty model. To check the subject with more precision, we present the simulation results based on different heterogeneity and cured fraction values.

6.1. Low heterogeneity

In the low level of heterogeneity ($\sigma^2 = 0.5$) and various levels of cured fraction, the bias and MSEs of regression parameters (γ) in the latency component for competing risks cure-frailty model were less than the model without frailty term. Further, the bias of the regression parameter in the incidence part (β_1) for the model without frailty term was a little less than the competing risks–cure-frailty model, although its MSE was almost the same (Table 1). The biases and RMSEs of CIFs in the competing risks-cure-frailty model were less than the model without frailty term in all of the time points, but the difference decreased in the last follow-up times (Table 2).

6.2. Moderate heterogeneity

In the moderate heterogeneity ($\sigma^2 = 1$) and the low level of cured fraction, the biases and MSEs of regression parameters in the latency and incidence parts of the competing risks-cure-frailty model were less than the model without frailty. However, in the moderate and high levels of cured fraction, the biases and MSEs of the parameters in the incidence part of the competing risks-cure-frailty model were a little more than the model without frailty (Table 1). The biases and MSEs of regression parameters in the latency part of the competing risks-cure-frailty model were smaller compared to the model without frailty (Table 1). Comparison of the biases and RMSEs of CIFs in the competing risks-cure-frailty model and the model without frailty term indicated that the model with frailty term was more efficient than the model without frailty term in all of the time points. However, the differences were more evident in the early time of follow-up (Table 2).

6.3. High heterogeneity

In the high heterogeneity ($\sigma^2 = 1.5$) and all levels of cured fraction, the biases and MSEs of competing risks-curefrailty model were less than the model without frailty term (Table 1). The biases and RMSEs of CIFs in the competing risks-cure-frailty model were less than the model without frailty term in all of the time points, and the difference was more evident in the early time of follow-up (Table 2).

7. Analysis of breast cancer data

The data resource of this paper was part of a historical cohort study that included 550 breast cancer patients who had been referred to Qaem or Omid hospitals of Mashhad University of Medical Sciences from 2001 to 2007 and followed up till March 2017 (Ghavami et al., 2017). The inclusion criteria of the study were being female and had received therapy through Breast Conservation Surgery (BCS) or Modified Radical Mastectomy (MRM) surgical methods. Metastatic breast cancer patients were excluded from the study. According to the conditions, all patients received adjuvant therapies, including radiotherapy, chemotherapy, and hormone therapy. The start time of the study was the date of operation and the end time was the loco-regional relapse or distant metastasis or the last confirmed date of breast cancer disease-free status. The endpoint of interest was the time from operation to loco-regional relapse (cause 1) in the presence of a competing cause, that is, distant metastasis (cause 2). Additionally, some prognostic medical factors, including hormone receptor status, tumor size, and lymph node status (the number and location of lymph nodes with cancer), were gathered. The tumor size and lymph node status were reported based on the American Joint Committee on Cancer classification (Cancer, 2002).

Table 3: Prognostic factors of patients

Characteristics	Count (%)
Hormone Receptor	
negative	200 (36.4)
positive	350 (63.6)
Tumor stage	
T1	69 (12.5)
Т2	319 (58.0)
ТЗ	131 (23.8)
T4	31 (5.6)
N stage	
NO	204 (37.1)
N1	192 (34.9)
N2	109 (19.8)
N3	45 (8.2)



Figure1: The Kaplan-Meier disease-free survival for Iranian breast cancer patients.

These prognostic factors are reported in Table 3. The maximum follow-up time was 15.6 years. Out of 550 patients, 171 individuals (30.9%) experienced loco-regional or distant metastasis. In the Kaplan-Meier plot of disease-free survival, the curve reaches a plateau after ten years, indicating the presence of cured fraction (Figure 1). Also, the central assumptions of cure models, that is sufficient follow-up time and the presence of cured fraction, were tested and confirmed by Maller and Zhou methods (Maller and Zhou, 1996). Owing to the properties of the data, the presence of cured fraction and competing risk (distant metastasis), and for evaluation of heterogeneity due to unobserved covariates, we used the introduced model in section 2.2, and a similar model without frailty term. We fitted these models without selecting variables; the results of fitting these models are presented in Table 4.

Table 4: Comparison o	of the results	of analysis	the 1	Iranian	breast	cancer	data	by	competing	risks	cure	model	and
competing risks cure fra	ailty model												

variables	competing risks cure r	nodel	competing risks cure frailty model			
	OR [*] HR ^{\$} (95%CI)	p-value	OR [*] HR ^{\$} (95%CI)	p-value		
Incidence part						
Intercept (se)	-2.38(0.28)	< 0.001	-2.32 (0.32)	< 0.001		
Hormone receptor						
negative	-	-				
positive	1.11 (0.70, 1.77)	0.652	1.21 (0.73, 2.01)	0.460		
N stage						
NO	-	-	-			
N1	3.98 (2.10, 7.56)	< 0.001	3.79 (1.95, 7.36)	< 0.001		
N2 N3	13.29 (7.02, 25.19)	< 0.001	13.81 (6.94, 27.49)	< 0.001		
Tumor stage						
T1 T2	-	-	-	-		
T3 T4	2.66 (1.66, 4.28)	< 0.001	2.58 (1.55, 4.30)	< 0.001		
Latency part						
Hormone receptor						
negative	-	-	-	-		
positive	0.55 (0.38, 0.80)	0.002	0.28 (0.13, 0.60)	0.001		
N stage						
NO	-	-	-	-		
N1	1.22 (0.56, 2.67)	0.611	2.05 (0.62, 6.84)	0.241		
N2 N3	2.28 (1.10, 4.71)	0.027	4.56 (1.44, 14.44)	0.010		
Tumor stage						
T1 T2	-	-	-	-		
T3 T4	1.23 (0.84, 1.80)	0.286	2.32 (1.12, 4.83)	0.024		
Variance of frailty (se)	-	-	1.34 (0.63)	0.035		
Scale parameter (se)	0.10 (0.036)	0.007	0.05 (0.030)	0.083		
Shape parameter (se)	1.64 (0.105)	< 0.001	2.49 (0.34)	< 0.001		
Relative cause-specific	Parameter estimation (se)	n-value	Parameter estimation (se)	n-value		
hazard part	Tarameter estimation (se)	p value	Tarameter estimation (se)	p value		
Hormone receptor						
negative	-	-	-	-		
positive	-1.04 (0.38)	0.006	-1.04 (0.38)	0.006		
N stage						
NO	-	-	-	-		
N1	-0.90 (0.62)	0.148	-0.90 (0.62)	0.148		
N2 N3	-1.25 (0.60)	0.038	-1.25 (0.60)	0.038		
Tumor stage						
T1 T2	-	-	-	-		
T3 T4	0.01 (0.40)	0.979	0.01 (0.40)	0.979		
$1{t \in (0,1.54]}$	1.27(0.64)	0.047	1.27(0.64)	0.047		
1 { $t \in (1.54, 2.5]$ }	0.55(0.65)	0.403	0.55(0.65)	0.403		
$1{t \in (2.5, 4.07]}$	0.26(0.57)	0.647	0.26(0.57)	0.647		
$1{t \in (4.07, 16]}$	0.51(0.64)	0.432	0.51(0.64)	0.432		
AIC	1385.258		1373.832			

* Odds Ratio \$ Hazards Ratio

Based on a likelihood ratio test, the model with a frailty term was better than the model without frailty ($\chi^2 = 13.43$, p < 0.001). Further, the Akaike information criterion (AIC) in the frailty model and the model without frailty were 1373.83 and 1385.26, respectively, which confirmed that the frailty model was more suitable than the other. An interesting point in the compared results of these models is that the effect of tumor stage in the latency part was significant in the model with frailty but was not significant in the model without frailty (Table 4). According to the

competing risks cure-frailty model in the incidence part of the model, the N-stage and T-stage significantly affected being susceptible (uncured). However, the hormone receptor did not have any significant effect (Table 4). After estimating the parameters of the competing risks-cure-frailty model, the CIFs with bootstrapping confidence intervals were plotted for evaluating the cause-specific cumulative incidence of time to loco-regional relapse. For example, in Figure 2, we compared the cumulative incidence of time to loco-regional relapse with the cumulative incidence of time to distant relapse in patients with positive hormone receptors, tumor size more than 5 cm, and more than 4 lymph nodes metastases. Table 5 shows the estimated relative cause-specific hazard implied by the fitted model with associated standard errors for patients with tumor size more than 5 cm, more than 4 lymph nodes metastases, and positive hormone receptors. The dominating cause of failure was distant metastasis, and the probability of relapse in the first quartile was near twice the probability in other quartiles.

Table 5: Estimated piecewise constant relative cause-specific hazards of relapse and their standard errors for patients with the tumor size more than 5 cm, more than three lymph nodes metastases and positive hormone receptors.

Time	(0, 1.54]	(1.54, 2.5]	(2.5, 4.07]	(4.07,16]
Relapse	0.26(0.08)	0.15(0.06)	0.12(0.06)	0.14(0.07)



Figure 2: The estimated cumulative incidences of time to loco-regional relapse and time to distant relapse based on competing risks cure frailty model for patients with the tumor size more than 5 cm, more than three lymph nodes metastases and positive hormone receptors, with the 95% confidence intervals.

8. Discussion

Based on our search, no study has combined competing risks models with cure models and univariate frailty models; although few articles have been published regarding the combination of competing risks models with cure models (Choi, Huang, and Cormier, 2015; Eloranta et al., 2014; Nicolaie, Taylor, and Legrand, 2018). The competing risks-cure-frailty model is suitable for conditions in which there is clinical evidence of a cured fraction (non-susceptible patients) in a cohort. Also, there is heterogeneity among susceptible patients, which causes some individuals to

experience failure faster than other patients with similar prognostic factors. Furthermore, the introduced model can estimate the cumulative incidence of time to a specific cause of failure in the presence of competing risks. Therefore, in the competing risks-cure-frailty model, besides the test of heterogeneity within susceptible patients, we can also infer about the effect of covariates on the chance of cure and the risk of failure, regardless of the cause of failure. In addition, we can set parameters on the relative position of each failure type among all failure types. An appealing technical characteristic of our model that has been adapted from the vertical modeling of competing risks is that the parameters of the two components in likelihood (14) can be estimated independently (Nicolaie, van Houwelingen, and Putter, 2015; Nicolaie, van Houwelingen and Putter, 2010). So, the standard packages such as glm in R can be used for relative cause-specific hazards. For maximizing the other part of the likelihood function (14), the optim function in R can be employed (R Team, 2008). In this study, we used individual frailty, but it is possible to extend this model by using shared frailty for recurrent events or cluster situations. Because vertical modeling can be naturally employed for complex competing risks data such as competing risks with the missing cause of failure, the competing risks-curefrailty model can be extended similarly for masked competing risks (Nicolaie, van Houwelingen, and Putter, 2015). In this study, we used a Weibull Gamma frailty model, but other parametric or semi-parametric models can be employed for different real data. Assessment of the real data of Iranian female breast cancer patients in this study showed the necessity to consider competing risk, cured fraction, and heterogeneity in modeling. The introduced competing risks-cure-frailty model covered these complex situations as well. Based on AIC, it was necessary to include a frailty term in the model. Compared to the study of Ghavami et al., 2017, we reported the results of both the competing risks cure model and competing risks cure frailty model to emphasize the importance of examining heterogeneity. We can see the effect of the tumor stage on the time of failure (loco-regional relapse or metastasis) in susceptible patients was not significant in the competing risks cure model while it is significant in the competing risks cure frailty model. It is closer to many studies that have demonstrated that the tumor stage is a prognostic factor for disease-free survival (Forse et al., 2013; Rondeau et al., 2013). In our study, the increased positive lymph nodes raised the hazard of failure for susceptible patients and decreased the chance of cure, which agrees with another study in Japan (Asano et al., 2013). An impressive result in our study was that the patients with positive hormone receptors had a significant effect on short-term disease-free survival but did not have any significant effect on long-term diseasefree survival. It is in accordance with medical findings in some studies that have shown that the hazard of recurrence in breast cancer patients with negative hormone receptor is more than patients with positive hormone receptor tumors in the first years after diagnosis, but this difference begins to lower after this time and finally disappears (Bentzon et al., 2008; Moffat, 2014).

9. Conclusion

In a survival study with competing risks and cured fraction, heterogeneity due to unobserved covariates between susceptible patients can destroy the estimates of parameters in the latency part of the cure model and the CIFs. In this complex situation, the competing risks-cure-frailty model introduced in this study can be recommended for assessing homogeneity in the susceptible population. This model can estimate the rate of cure and determine the prognostic factors that influence short-term and long-term survival. Additionally, the proposed model can estimate the cause-specific cumulative incidences.

Acknowledgements

This study is a part of the result of a Ph.D. thesis in the department of epidemiology and biostatistics in the school of public health of Tehran University of Medical Sciences that the Ethics Committee approved affiliated to the Deputy of Research, Tehran University of Medical Sciences, Iran (Ethic No: IR.TUMS.SPH.REC.1395.1970). We thank the physicians and staff at Qaem and Omid Hospitals of Mashhad University of Medical Sciences for cooperating during data collection.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

- Asano, J., Hirakawa, A., Hamada, C., Yonemori, K., Hirata, T., Shimizu, C., ... Fujiwara, Y. (2013). Use of Cox's Cure Model to Establish Clinical Determinants of Long-Term Disease-Free Survival in Neoadjuvant-Chemotherapy-Treated Breast Cancer Patients without Pathologic Complete Response. *Int J Breast Cancer*, 2013, 354579. doi: 10.1155/2013/354579
- 2. Basu, S., and Tiwari, R. C. (2010). Breast cancer survival, competing risks and mixture cure model: a Bayesian analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 173*(2), 307-329.
- Bentzon, N., During, M., Rasmussen, B. B., Mouridsen, H., and Kroman, N. (2008). Prognostic effect of estrogen receptor status across age in primary breast cancer. *Int J Cancer*, 122(5), 1089-1094. doi: 10.1002/ijc.22892
- 4. Berkson, J., and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American statistical association*, 47(259), 501-515.
- 5. Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological), 11*(1), 15-53.
- 6. Cai, C. (2013). Advanced methodology developments in mixture cure models (Order No. 3565728). Retrieved from https://search.proquest.com/docview/1417778036?accountid=41308 Available from Biological Science Database; ProQuest Dissertations and Theses Global
- 7. Cancer, A. J. C. o. (2002). Breast AJCC cancer staging manual (pp. 223-240): Springer.
- 8. Choi, S., Huang, X., and Cormier, J. N. (2015). Efficient semiparametric mixture inferences on cure rate models for competing risks. *Canadian Journal of Statistics*, 43(3), 420-435.
- 9. Corbière, F., Commenges, D., Taylor, J. M. G., and Joly, P. (2009). A penalized likelihood approach for mixture cure models. *Statistics in Medicine*, 28(3), 510-524. doi: 10.1002/sim.3481
- 10. Eloranta, S., Lambert, P. C., Andersson, T. M.-L., Björkholm, M., and Dickman, P. W. (2014). The application of cure models in the presence of competing risks: a tool for improved risk communication in population-based cancer patient survival. *Epidemiology*, 25(5), 742-748.
- 11. Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 1041-1046.
- 12. Fine, J. P., and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, *94*(446), 496-509.
- Forse, C. L., Yilmaz, Y. E., Pinnaduwage, D., O'Malley, F. P., Mulligan, A. M., Bull, S. B., and Andrulis, I. L. (2013). Elevated expression of podocalyxin is associated with lymphatic invasion, basal-like phenotype, and clinical outcome in axillary lymph node-negative breast cancer. *Breast Cancer Res Treat*, 137(3), 709-719. doi: 10.1007/s10549-012-2392-y
- 14. Gee, C. (2004). *Cure rate estimation in the analysis of survival data with competing risks* (Ph.D), university of southern California Retrieved from https://search.proquest.com/docview/305126122?accountid=41308 Available from ProQuest Dissertations and Theses Global
- Ghavami, V., Mahmoudi, M., Foroushani, A. R., Baghishani, H., Shandiz, F. H., and Yaseri, M. (2017). Long-Term Disease-Free Survival of Non-Metastatic Breast Cancer Patients in Iran: A Survival Model with Competing Risks Taking Cure Fraction and Frailty into Account. *Asian Pacific journal of cancer prevention: APJCP*, 18(10), 2825.
- 16. Kuk, A. Y., and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79(3), 531-541.
- 17. Larson, M. G., and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Applied statistics*, 201-211.
- 18. Maller, R. A., and Zhou, X. (1996). Survival analysis with long-term survivors (pp. 29-39): Wiley New York.
- 19. Moffat, F. (2014). Clinical and pathologic prognostic and predictive factors *Diseases of the Breast: Fifth Edition*: Wolters Kluwer Health Adis (ESP).
- 20. Nicolaie, M., van Houwelingen, H., and Putter, H. (2015). Vertical modelling: Analysis of competing risks data with missing causes of failure. *Statistical methods in medical research*, 24(6), 891-908.
- 21. Nicolaie, M., van Houwelingen, H., and Putter, H. (2010). Vertical modeling: A pattern mixture approach for competing risks modeling. *Statistics in Medicine*, 29(11), 1190-1205.
- 22. Nicolaie, M., Taylor, J. M. G., and Legrand, C. (2018). Vertical modeling: analysis of competing risks data with a cure fraction. *Lifetime Data Anal*. doi: 10.1007/s10985-018-9417-8
- 23. Peng, Y., and Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56(1), 237-243.

- 24. Pepe, M. S., and Mori, M. (1993). Kaplan—meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, *12*(8), 737-751.
- 25. Pintilie, M. (2011). An introduction to competing risks analysis. *Revista Española de Cardiología (English Edition)*, 64(7), 599-605.
- 26. Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V., and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 541-554.
- 27. Rondeau, V., Schaffner, E., Corbière, F., Gonzalez, J. R., and Mathoulin-Pélissier, S. (2013). Cure frailty models for survival data: Application to recurrences for breast cancer and to hospital readmissions for colorectal cancer. *Statistical methods in medical research*, 22(3), 243-260.
- 28. Sy, J. P., and Taylor, J. M. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, 56(1), 227-236.
- 29. Team, R. D. C. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org.
- 30. Wienke, A. (2010). Frailty models in survival analysis (pp. 55-56): CRC Press.