RESEARCH ARTICLE

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

Vahid Ghavami[1], Mahmood Mahmoudi[1], Abbas Rahimi Foroushani[1]*, Hossein Baghishani[2], Fatemeh Homaei Shandiz[3], Mehdi Yaseri[1]

Abstract

Introduction: Survival modeling is a very important tool to detect risk factors and provide a basis for health care planning. However, cancer data may have properties leading to distorted results with routine methods. Therefore, this study aimed to cover specific factors (competing risk, cure fraction and heterogeneity) with a real dataset of Iranian breast cancer patients using a competing risk-cure-frailty model. **Materials and methods:** For this historical cohort study, information for 550 Iranian breast cancer patients who underwent surgery for tumor removal from 2001 to 2007 and were followed up to March 2017, was analyzed using R 3.2 software. **Results:** In contrast to T-stage and N-stage, hormone receptor status did not have any significant effect on the cure fraction (long-term disease-free survival). However, T-stage, N-stage and hormone receptor status all had a significant effect on short-term disease-free survival so that the hazard of loco-regional relapse or distant metastasis in cases positive for a hormone receptor was only 0.3 times that for their negative hormone receptor counterparts. The likelihood of locoregional relapse in the first quartile of follow up was nearly twice that of other quartiles. The least cumulative incidence of time to locoregional relapse was for cases with a positive hormone receptor, low N stage and low T stage. The effect of frailty term was significant in this study and a model with frailty appeared more appropriate than a model without, based on the Akaike information criterion (AIC); values for the frailty model and one without the frailty parameter were 1370.39 and 1381.46, respectively. **Conclusions:** The data from this study indicate ae necessity to consider competing risk, cure fraction and heterogeneity in survival modeling. The competing risk-cure-frailty model can cover complex situations with survival data.

Keywords: Breast neoplasms- competing risks- cure model- frailty model- survival analysis

Asian Pac J Cancer Prev, **18 (10)**, 2825-2832

Introduction

Breast cancer, with a 14-16% annual death rate, includes 23% of total female cancers worldwide, and is the most common cause of cancer mortality in women (Parkin et al., 2005; Anderson and Jakesz, 2008). It is the most prevalent cancer in Iran (Mousavi et al., 2009). Iranian patients feature being younger than their developed countries counterparts (Harirchi et al., 2004; Mousavi et al., 2007).

In cancer studies, survival modeling is a very important tool to detect risk factors, which may be the basis of health care planning. There are three important types of models in survival modelling as follows: 1) competing risks models 2) cure models and 3) frailty models. The usage of these models, from a practical point of view, can be recommended instead of popular ones in certain situations. Below, the situations in which the above said models can be used will briefly be explained and then a unique model will be introduced and utilized to analyze a real dataset of Iranian breast cancer patients.

Competing risks in breast cancer

One of the important aspects in survival modeling in breast cancer is to survey the disease-free survival (Nguyen et al., 2008a), meaning no patient experience no relapse (e.g., locoregional relapse) or no metastasis. The evaluation of the locoregional relapse free survival in breast cancer may be changed if patients first experience another event such as distant metastasis; therefore, assessing the cause specific failure rate is of great importance (Arriagadal et al., 1992). In this situation, the approach of competing risks should be applied. A competing risk, in survival analysis, is an alternative outcome (e.g., metastasis) that alters the probability of the event of interest (e.g., locoregional relapse) (Gooley et al., 1999). Several approaches for modeling competing risks such as cause-specific model, Fine and Gray's model

*[1]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, [2]Department of Applied Mathematics, Shahrood University of Technology, Shahrood, [3]Cancer Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. *For Correspondence: rahimifo@tums.ac.ir*

and vertical model have been presented (Prentice et al., 1978; Fine and Gray, 1999; Nicolaie et al., 2010). The use of popular methods in presence of competing risks may lead to biased results (Pepe and Mori, 1993).

Long-term survivors in breast cancer

Considerably decreased cancer mortality rates over the past 2 decades for the four major cancers (lung, breast, prostate, and colorectal) is the outcome of advances in early detection and treatment of them (Siegel et al., 2017). Some studies shows the high rate of disease-free survival in early breast cancer in a long term follow up, suggesting considerable fraction of patients experience neither relapse nor distant metastasis after sufficient follow up (Group, 2011; Cheng et al., 2016; van Maaren et al., 2016). On the other hand, models for survival analysis are based on this assumption that all people in the studied population are susceptible to the event of interest (say locoregional relapse in breast cancer) and will finally experience the event if there is a sufficiently long follow-up. However, as stated above, this assumption may not always be true as a fraction of patients do not experience the event of interest after sufficient follow-up; consequently, the use of popular survival models and ignoring such long term survivors (cured subjects) would lead to overestimation of the survival of the susceptible subjects (Corbière et al., 2009). Mixture models, known as cure models, are approaches for modeling long-term survival studies (Farewell, 1982). Cure models make it possible to estimate the probable susceptibility of a subject (incidence) and the time the event may occur, given that a failure is occurred (latency). Some studies on breast cancer have used cure models and have confirmed their suitability (Peng and Dear, 2000; Zhang and Peng, 2009; Rama et al., 2010; Rondeau et al., 2013).

Heterogeneity in breast cancer data

In medical and epidemiological studies, there may be some important information existed among individuals which cannot be explicated in a survival model because they have not been registered for any reason. For instance, microarray analysis has identified breast cancer subtypes with distinct gene expression profiles and the impact of this breast cancer subtypes on locoregional relapse has already been shown (Perou et al., 2000; Nguyen et al., 2008a); however, although there are identical biomarker profiles and stages, difference could be seen in the clinical outcome: a recurrence could be seen in 20% of patients who suffer from node-negative breast cancer disease and a disease-free status may remain in over 30% of patients who have lymph node metastases (Joensuu et al., 1998; Group, 2005). Thus, there may be other unobserved factors that affect the survival of patients. Lack of information on these factors in a study could result in heterogeneity and failing to consider heterogeneity may lead to distorted results (Price and Manatunga, 2001). Frailty models are recommended assessing this heterogeneity in such situations (Peng and Zhang, 2008).

The goal of this study is to evaluate free locoregional relapse survival in Iranian breast cancer patients in the presence of competing risk (distant metastasis) and the existence of long-term survivors (cure fraction),and the assessment of the heterogeneity in susceptible (uncured) patients. Precisely, a model has been used to cover four objectives as follows:

To estimate the effect of covariates on time to disease due to locoregional relapse or distant metastasis.

To estimate the cause-specific cumulative incidences for locoregional relapse in the population of susceptible patients.

To estimate the effect of covariates on the probability of being cured (long-term survivors).

To evaluate the individual heterogeneity.

To assess that the relative risk of locoregional relapse changes during the time of follow-up.

Materials and Methods

This is a historical cohort study that its data resource was from female breast cancer patients who underwent surgery for tumor removal at Qaem Hospital or Omid Hospital of Mashhad University of Medical Sciences from 2001 to 2007 and were followed up to March 2017. Follow-up time was from the date of operation to the date of the locoregional relapse or distant metastasis, or to the last confirmed date of breast cancer disease-free status. Patients who received neo-chemotherapy and patients in stage IV were excluded from the study. Finally, 550 patients entered in to the study. Patients were visited every 3 months in the first post-surgery year, every 6 months between the 2nd and 5th post-surgery years and every year from then on. Some visits was done at the own request of the patients. All patients treated with MRM or BCS surgical methods and according to the conditions received adjuvant therapies including radiotherapy, chemotherapy and hormone therapy. This study was approved by Ethics Committee affiliated to the Deputy of Research, Tehran University of medical sciences, Iran (Ethic No: IR.TUMS.SPH.REC.1395.1970). The variables in this study were age at diagnosis, clinical stage at diagnosis (based on TNM system), N-stage, T-stage (based on the American Joint Committee on Cancer classification), type of surgery, radiotherapy, chemotherapy, BMI and hormone receptor status. Endpoint of interest is the time from operation to locoregional relapse (cause 1), in the presence of a competing cause, that is, distant metastasis (cause 2).

Statistical analyses

Analysis of survival was done in a framework of a mixture cure competing risk model with individual frailty so that a vertical approach was used for modeling competing risks (Nicolaie et al., 2010; Nicolaie et al., 2015a; Nicolaie et al., 2015b).

Vertical competing risks' modeling utilizes new decomposition of the joint distribution of time to failure (T) and cause of failure (D) as:

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

This decomposition consists of two components: (1) the time of failure and (2) the cause of failure condition on the time of failure. Observable quantities, the total

hazard, and the relative cause-specific hazards form the basis of both components of the model. If λj (t) and λ.(t) refer to cause specific hazard and total hazard (hazard of any failure irrespective of its cause) the relative cause specific hazards were defined as:

$$P(D=j|T=t)=\pi_j(t)=\frac{\lambda_j(t)}{\lambda_.(t)} \qquad j=1,\ldots,J. \qquad (1)$$

The aim of vertical modeling in competing risk is to estimate the joint distribution of (T, D), given by the cumulative incidence functions:

$$F_j(t)=P(T\le t, D=j)=\int_0^t \lambda_j(s)S(s)ds=\int_0^t \pi_j(s)\lambda_.(s)S(s)ds. \qquad (2)$$

In equation 2, S(s) is the total survival function (survival function of any failure irrespective of its cause). Relating to modeling aspect, two models necessitate, a model for the overall failure time (regardless of its cause) and a model for the cause of failure, given the failure time (Nicolaie et al., 2010). For modelling relative cause-specific hazards, a time dependent logistic regression model can be used as:

$$\pi_j(t|\boldsymbol{X})=\frac{\exp(\alpha_j^T\boldsymbol{B}(t)+\beta_j^T\boldsymbol{X})}{\sum_{j=1}^{J}\exp(\alpha_j^T\boldsymbol{B}(t)+\beta_j^T\boldsymbol{X})} \qquad (3)$$

where B(t) is an r-vector of pre-specified time functions, X is a vector of independent variables, αj^T and βj^T stand for m-vectors of unknown regression parameters and j=1,…J is cause of failure. In this study, we utilized piecewise constant time functions, with cut-off points at the quartiles of the failure time distribution where:

$$\boldsymbol{B}(t)=(\mathbf{1}\{t\in(0,1.54]\},\mathbf{1}\in\{t\in(1.54,2.5]\},\mathbf{1}\{t\in(2.5,4.07]\},\mathbf{1}\{t\in(4.07,16]\}).$$

For modelling total hazard, all failures are considered as events, regardless of the cause of failure. Several choices including semiparametric and parametric models can be used.

In this study, we used a mixture cure model for the total hazard. In mixture cure models, it is supposed that the community falls into two groups: people who are at risk (susceptible group) and those who are long-term survivor (cure fraction). The equation mentioned below shows relation between population survival time (S(t)) and the survival time of susceptible group $S_u(t)$ via cure model:

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

where π is the probability of being uncured. In this model π and $S_u(t)$ can be modeled by logistic regression and weibull-gamma frailty model respectively (Farewell, 1982). In a frailty model, the intensity of an event (locoregional relapse or distant metastasis) for a given person i at time t is:

$$\lambda_{i.}(t)=w_i\lambda_{0.}(t)\exp(\gamma^T\boldsymbol{X})$$

In this formula $\lambda_0.(t)$ is the baseline hazard, γ^T stands for m-vector of unknown regression parameters, X is a vector of independent variables and w_i is the frailty for i.th patient that follows a distribution with mean equal one. Patients with a high w_i value tend to have a high rate of event; thus, variance of this distribution is a measure of the heterogeneity of the patients (Wienke, 2010a). In this study, the frailty followed a gamma distribution.

The two basic assumptions of cure models, i.e. the existence of patients with long-term survival and the sufficiency of follow up were tested by Maller and Zhou tests (Maller and Zhou, 1996). The presence of unobserved heterogeneity was evaluated by log likelihood ratio test. Selection of variables was done via the use of backward method. Cumulative incidence of locoregional relapse-free with bootstrapping confidence intervals (with 200 iterations) plotted based on model-estimated parameters. All analysis was performed using packages stat, survival and self-writing programs in R3.2 software (Team, 2008).

Results

Out of 550 patients who entered in the study, 312 patients were in early stage (stages I and II according to TNM system) of breast cancer. The patients mean age was 47.88±10.86 years at the time of diagnosis. Most of the patients (92.7%) underwent modified radical mastectomy. 63.6% of patients were with positive hormone receptor status. The details of patients and tumor characteristics are listed in Table 1. The maximum of follow-up was 15.6 years and the mean disease-free survival was 11.43 years. 49 patients (8.9%) had locoregional relapse and 122 patients (22 %) experienced distant metastasis. The Kaplan Meier diagram for disease-free survival has been shown in Figure 1. It appears that the disease-free survival curve reaches a plateau after 10 years from the time of surgery, indicating the presence of a sub-population, who survives event-free by the end of the follow-up. By considering these data and via the use of Maller and Zhou table, we accepted the hypothesis of long-term survivors or the existence of cured people and the sufficiency of follow up time (Maller and Zhou, 1996). Because of the existence of competing risk (distant metastasis) and long-term survivors (cure fraction), and with the goal of the evaluation of heterogeneity, a competing risks mixture cure frailty model was used.

Disease-free survival analysis

To evaluate the effect of covariates on disease-free

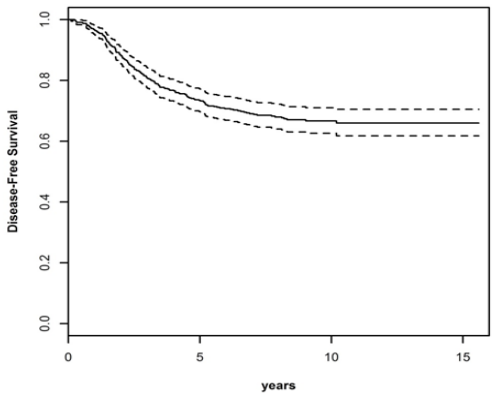

Figure 1. The Kaplan-Meier Disease- Free Survival with the 95% Confidence Interval

Table 1. Patient and Tumor Characteristics (N= 549)

Characteristics	Count (%)
Age, years	
≤ 45	242 (44)
> 45	308 (56)
BMI ,kg/m2	
< 30	410 (74.5)
≥ 30	140 (25.5)
Tumor stage	
T1	69 (12.5)
T2	319 (58.0)
T3	131 (23.8)
T4	31 (5.6)
N stage	
N0	204 (37.1)
N1	192 (34.9)
N2	109 (19.8)
N3	45 (8.2)
Clinical stage	
I	42 (7.7)
II	290 (52.7)
III	218 (39.6)
Hormone receptor	
negative	200 (36.4)
positive	350 (63.6)
Operation	
BCS	40 (7.3)
MRM	510 (92.7)
Radiotherapy	
No	96 (17.5)
Yes	454 (82.5)
Chemotherapy	
No	21 (3.8)
Yes	529 (96.2)
Hormone therapy	
No	187 (34)
Yes	363 (66)

survival time (disease due to locoregional relapse or distant metastasis) a weibull-gamma frailty cure model was used. Backward method was used for the selection of variables in a multiple regression model. As Clinical stage consists of two variables, N-stage and T-stage, it was excluded from multivariate analysis, avoiding collinearity. Incidence and latency parts of Table 2 show the final Weibull-gamma cure frailty model with their coefficients (odds ratio for incidence and hazards ratio for latency), their 95% confidence intervals and p-values. In incidence part of the model, the N-stage and T-stage had a significant effect, so that the odds of being susceptible (being uncured) for patients with N1 stage was 3.91 times more than N0 patients and odds ratio was 14 for patients with N2|N3 stage. In addition, the odds of being

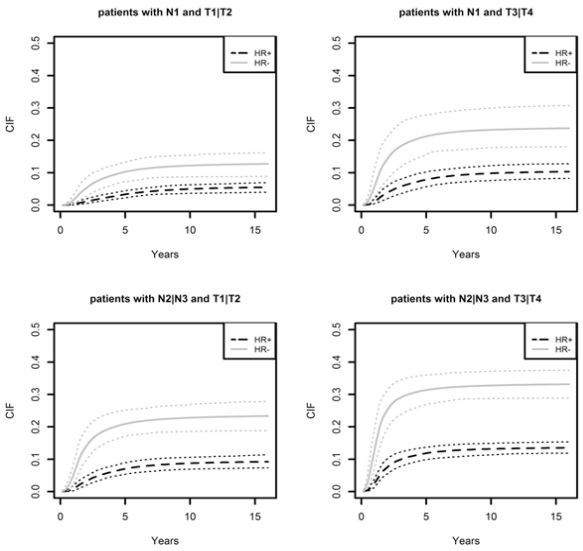

Figure 2. Comparison the Cumulative Incidence of Time to Locoregional Relapse with the 95% Confidence Intervals between Hormone Receptor (HR) Status at Different Levels of N Stage and T Stage.

susceptible for patients with tumor size more than 5 cm (T3|T4) was 2.53 times greater than patients with tumor size less than 5 cm (T1|T2). Based on the results of latency part, in susceptible patients, the hazard of relapse or distant metastasis in positive hormone receptor patients was less than other patients, given the same value of frailty (HR=0.3, p=0.001). Also, considering the same value of frailty, patients withN2|N3 or patients with T2|T3 had more hazards to experience failure (relapse or distant metastasis) earlier than N0 or T1|T2 patients respectively (Table2). The likelihood ratio test indicated that the model with frailty was better than the model without frailty term (χ^2=13.07, p<0.001); furthermore, the AIC in the frailty

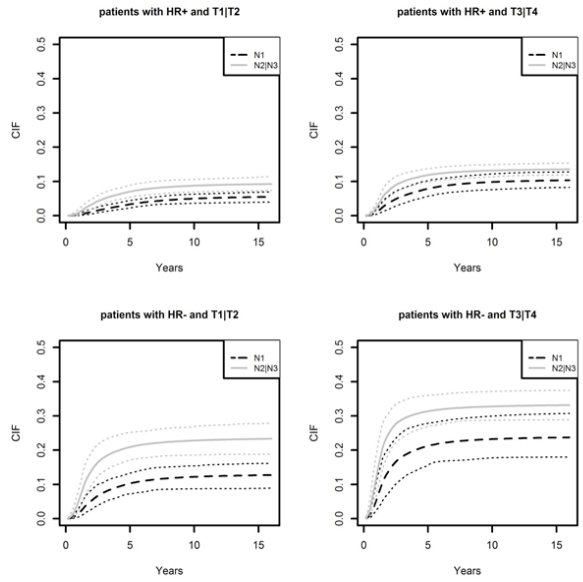

Figure 3. Comparison the Cumulative Incidence of Time to Locoregional Relapse with the 95% Confidence Intervals between N Stages at Different Levels of T Stage and Hormone Receptor (HR) Status.

Table 2. Results of Analyses Using the Weibull Cure Frailty Model

variables	Weibull Cure Frailty Model	
	OR* | HR$ (95%CI)	p-value
Incidence part		
Hormone receptor		
negative	-	-
positive	Not selected	-
N stage		
N0	-	-
N1	3.91 (2.04, 7.48)	< 0.001
N2|N3	14.00 (7.11,27.58)	< 0.001
Tumor stage		
T1|T2	-	-
T3|T4	2.53 (1.53, 4.17)	< 0.001
Latency part		
Hormone receptor		
negative	-	-
positive	0.30 (0.14, 0.61)	0.001
N stage		
N0	-	-
N1	1.99 (0.61, 6.44)	0.251
N2|N3	4.38 (1.43, 13.45)	0.014
Tumor stage		
T1|T2	-	-
T3|T4	2.32 (1.12, 4.79)	< 0.023
Frailty Variance (SE)	1.29 (0.610)	0.034
Scale parameter (SE)	0.04 (0.026)	0.08
Shape parameter (SE)	2.48 (0.335)	< 0.001

*, Odds Ratio; $, Hazards Ratio

model was 1370.39 and in the model without frailty term it was 1381.46, validating that the frailty model was more suitable than the other(details of result for the model without frailty term is available in the supplement file).

Cause specific relative hazards

Table 3 illustrates the assessment results of the relative hazard of locoregional relapse, and the effects of risk factors on it during the time of follow-up. As shown, the effects of hormone receptor, N-stage and time were significant, so that the odds of locoregional relapse in first quartile of follow up was more than other quartiles; specially, the positive hormone receptor patients had less odds for relapse compared to the negative hormone receptor patients. Table 4 gives the estimated cause specific relative hazards implied by fitted model with associated standard errors for positive hormone receptor patients with N1. It can be seen that the dominating cause of failure was distant metastasis and the probability of relapse in first quartile was nearly twice the probability in other quartiles.

Cause-specific cumulative incidences

After estimating parameters of the model, the cumulative incidence of time to locoregional relapse

Table 3. Results of Logistic Regression for Piecewise Constant Relative Hazards of Relapse

Cause specific relative hazards part	Regression parameter (SE)	p-value
Hormone receptor		
negative	-	-
positive	-1.05(0.38)	0.005
lymph node metastases		
N0	-	-
N1	-0.89(0.61)	0.141
N2|N3	-1.25(0.59)	0.034
$1\{t \in (0,1.54]\}$	1.27(0.64)	0.047
$1\{t \in (1.54,2.5]\}$	0.55(0.65)	0.403
$1\{t \in (2.5,4.07]\}$	0.26(0.57)	0.647
$1\{t \in (4.07,16]\}$	0.51(0.64)	0.432

was estimated based on equation 2. Figure 2 shows that the cumulative incidence of time to locoregional relapse in positive hormone receptor patients was lower than negative hormone receptor patients, meaning positive hormone receptor patients had better locoregional relapse free survival in the presence of distant metastasis. Since the 95% confidence bands did not cover each other, the difference was significant at α=0.05. Figure 3 illustrates that the cumulative incidence of time to relapse for patients with N1-stage tended to be less than N2|N3 patients in different levels of T stage and hormone receptor, but

Table 4. Estimated Piecewise Constant Relative Cause Specific Hazards of Relapse and Their Standard Error for a Hormone Receptor - Positive Patient with N1.

	(0, 1.54]	(1.54, 2.5]	(2.5, 4.07]	(4.07,16]
Relapse	0.34(0.11)	0.20(0.08)	0.16(0.07)	0.19(0.07)

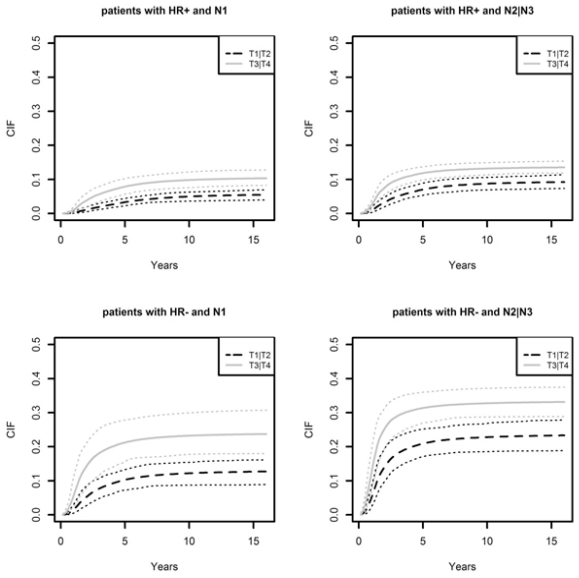

Figure 4. Comparison the Cumulative Incidence of Time to Locoregional Relapse with the 95% Confidence Intervals between T Stages at Different Levels of N Stage and Hormone Receptor (HR) Status

the difference was not significant for T3|T4 patients. According to figure 4, although patients with tumor size more than 5 cm had higher cumulative incidence of time to relapse than patients with tumor size less than 5 cm, the difference was greater in the negative hormone receptor patients. Based on figures 2, 3 and 4 the best locoregional relapse-free survival was for the patients with positive hormone receptor, low N stage and also low T stage.

Discussion

In this study, a competing risk-cure-frailty model was applied to determine the prognostic factors for: 1) long term disease-free survival, 2) short term disease free survival, 3) cause specific cumulative incidence and the evaluation of the unobserved heterogeneity. Although competing risk models and parametric cure models have previously been used for the assessment of free locoregional relapse survival in breast cancer patients(Nguyen et al., 2008b; Botteri et al., 2010) and Iranian patients breast cancer(Baghestani et al., 2015; Safe et al., 2016; Hoseini et al., 2017) respectively, this is the first study to present a unique model that combines the competing risks, cure and frailty models.

The mean age of patients in this study was 47.88 years indicating that the age of effect by breast cancer in Iranian women was one decade, at least, lower than their developed countries counterparts (Sant et al., 1998; Fredholm et al., 2009). The effect of age at diagnosis was not a significant effect on disease-free survival in our study that was in accord with some studies and in contrary to some others (Rondeau et al., 2013; Jafari-Koshki et al., 2014).

Ewertz et al. showed that BMI of 30 kg/m^2 or more had no influence on the risk of local relapse but it was an independent prognostic factor for developing distant metastases (Ewertz et al., 2011); in our study, obesity (BMI $\geq$ 30 kg/m2) had no influence on both disease-free survival and locoregional relapse.

In this study, we found that there was no significant difference between disease-free survival after MRM or BCS, being accord with some investigations (van Tienhoven et al., 1999) and contrary to some others (Shenouda et al., 2014). It should be noted that the proportion of patients to BCS was very low in our study (7.3%).

In the present study, the hazard of experiencing a post-surgery relapse, including locoregional relapse or distant metastasis, was found to be associated with T stage and N stage; consequently, increased number of positive lymph nodes raised the hazard level of relapse for susceptible patients and decreased the chance of cure. A study performed in Japan revealed that the number of positive lymph nodes significantly affected both the levels of disease-free survival for susceptible patients and being cured (Asano et al., 2013). Similar to our investigation, some other studies showed that T stage was another prognostic factor for disease-free survival (Forse et al., 2013; Rondeau et al., 2013).

In our study, positive hormone receptor susceptible patients had better disease-free survival than patients with negative hormone receptor; although the status of hormone receptor did not have any effect on long-term disease-free survival. This is in accord with some other studies that have found that the chance of breast cancer recurrence in positive hormone receptor tumors is a slightly lower than that of the negative hormone receptor tumors in the first five years after diagnosis; however, after this time, the difference begins to lower and finally disappear (Bentzon et al., 2008; Moffat, 2014).

Cong and Tsokos applied a parametric mixture cure model to analyze breast cancer relapse time with different treatments and concluded that patients who received both radiation and hormone therapy were more likely to be cured of breast cancer and less susceptible to recurrence (distant or locoregional) than those who received only hormone therapy (Cong and Tsokos, 2010). In our study, more patients received radiotherapy (82.5%) and there was a low number of patients with hormone therapy only in comparison to Cong and Tsokos's study; however, there was not any significant relation between radiotherapy or hormone therapy and relapse time for susceptible patients; radiotherapy and hormone therapy had not any significant effect on cure.

Botteri et al., (2010) reported that in breast cancer patients with a conservative surgery, size of tumor, ER, HER2 and ki-67 had a significant effect on local relapse (in the presence of metastatic relapse or death as competing risks). In this study, it was indicated that free locoregional relapse survival (in the presence of metastasis as competing risk) in the positive hormone receptor patients was significantly higher compared to the negative hormone receptor patients. Also, the size of tumor had a significant effect on free locoregional survival time.

Warren et al., (2016) utilized a competing risk model for evaluating the locoregional recurrence in women with early stage breast cancer. They indicated, similar to our results, that the hazard of locoregional recurrence was lower in patients with positive hormone receptor in comparison to patients with negative hormone receptor.

Similar to our findings, in a competing risk study (distant metastasis as a competing risk), the cumulative incidence of local recurrence was lower in the positive hormone receptor patients (Nguyen et al., 2008b).

Based on Akaike information criterion (AIC), the frailty model was better than the model without a frailty term; this expresses that there were other factors that have not been registered in our study. This finding showed that heterogeneity in data should be considered and expressed in some studies (Joensuu et al., 1998; Group, 2005; Gohari et al., 2006). It is interesting that in our study, the effect of tumor stage on the time of failure (locoregional relapse or metastasis) in susceptible patients was significant for the frailty model and not significant in the model without frailty term.

It can be seen that the probability of relapse in first 1.5 years after surgery was nearly twice the probability in other times, and the major cause of failure was distant metastasis that accords with other studies (Arriagadal et al., 1992; Nguyen et al., 2008b; Botteri et al., 2010).

Assessing the data in this study showed the necessity to consider competing risk, cure fraction and heterogeneity

in modeling and the competing risk-cure-frailty model can cover these complex situations in survival data. Weibull distribution, because of its convenience and flexibility, seems to be the most widely applied parametric lifetime models and used in a wide range of biostatistical problems (Wienke, 2010b). On the other hand, from a computational and analytical point of view, Gamma distribution, as a mixture distribution, fits very well with failure data Abbring and Van Den Berg, 2007); consequently, a weibull distribution with a gamma frailty was used for susceptible patients in a cure model in this study. Also, it is possible to use other parametric and semi parametric models in a vertical competing risk framework and compare them to each other in future studies.

Statement conflict of Interest

All authors reported no conflict of interest.

Acknowledgements

We are grateful to the physicians and staffs of the Qaem and Omid Hospitals of Mashhad University of Medical Sciences for their assistance with researchers in data gathering. This paper is a part of a thesis in department of epidemiology and biostatistics in faculty of health of Tehran University of Medical Sciences.

References

Abbring JH, Van Den Berg GJ (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika*, **94**, 87-99.

Anderson BO, Jakesz R (2008). Breast cancer issues in developing countries: an overview of the breast health global initiative. *World J Surg*, **32**, 2578-85.

Arriagadal R, Rutqvist L, Kramar A, et al (1992). Competing risks determining event-free survival in early breast cancer. Br. J. Cancer, 66, 957.

Asano J, Hirakawa A, Hamada C, et al (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.

Baghestani AR, Zayeri F, Akbari ME, et al (2015). Fitting cure rate model to breast cancer data of cancer research center. *Asian Pac J Cancer Prev*, **16**, 7923-7.

Bentzon N, During M, Rasmussen BB, et al (2008). Prognostic effect of estrogen receptor status across age in primary breast cancer. *Int J Cancer*, **122**, 1089-94.

Botteri E, Bagnardi V, Rotmensz N, et al (2010). Analysis of local and regional recurrences in breast cancer after conservative surgery. *Ann Oncol*, **21**, 723-8.

Cheng SH-C, Yu B-L, Horng C-F, et al (2016). Long-term survival and stage I breast cancer subtypes. *J Cancer Surviv*, **3**, 1-8.

Cong C, Tsokos C (2010). Statistical modeling of breast cancer relapse time with different treatments. *J Appl Sci*, **10**, 37-44.

Corbière F, Commenges D, Taylor JMG, et al (2009). A penalized likelihood approach for mixture cure models. *Stat Med*, **28**, 510-24.

Ewertz M, Jensen MB, Gunnarsdottir KA, et al (2011). Effect of obesity on prognosis after early-stage breast cancer. *J Clin Oncol*, **29**, 25-31.

Farewell VT (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041-6.

Fine JP, Gray RJ (1999). A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*, **94**, 496-509.

Forse CL, Yilmaz YE, Pinnaduwage D, et al (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**, 709-19.

Fredholm H, Eaker S, Frisell J, et al (2009). Breast cancer in young women: poor survival despite intensive treatment. *PLoS One*, **4**, e7695.

Gohari MR, Mahmoudi M, Mohammed K, et al (2006). recurrence in breast cancer. Analysis with frailty model. *Saudi Med J*, **27**, 1187-93.

Gooley TA, Leisenring W, Crowley J, et al (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med*, **18**, 695-706.

Group EBCTC (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*, **365**, 1687-717.

Group EBCTC (2011). Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10801 women in 17 randomised trials. *Lancet*, **378**, 1707-16.

Harirchi I, Karbakhsh M, Kashefi A, et al (2004). Breast cancer in Iran: results of a multi-center study. *Asian Pac J Cancer Prev*, **5**, 24-7.

Hoseini M, Bahrampour A, Mirzaee M (2017). Comparison of weibull and lognormal cure models with cox in the survival analysis of breast cancer patients in Rafsanjan. *J Res Health Sci*, **17**, 1-6.

Jafari-Koshki T, Mansourian M, Mokarian F (2014). Exploring factors related to metastasis free survival in breast cancer patients using Bayesian cure models. *Asian Pac J Cancer Prev*, **15**, 9673-8.

Joensuu H, Pylkkänen L, Toikkanen S (1998). Long-term survival in node-positive breast cancer treated by locoregional therapy alone. *Br J Cancer*, **78**, 795.

Maller RA, Zhou X (1996). Survival analysis with long-term survivors, Wiley New York, pp 29-39.

Moffat F (2014). Clinical and pathologic prognostic and predictive factors. In 'Diseases of the Breast: Fifth Edition', Eds Wolters Kluwer Health Adis (ESP), pp 439-51.

Mousavi SM, Gouya MM, Ramazani R, et al (2009). Cancer incidence and mortality in Iran. *Ann Oncol*, **20**, 556-63.

Mousavi SM, Montazeri A, Mohagheghi MA, et al (2007). Breast cancer in Iran: An epidemiological review. *Breast J*, **13**, 383-91.

Nguyen PL, Taghian AG, Katz MS, et al (2008a). Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. *J Clin Oncol*, **26**, 2373-8.

Nguyen PL, Taghian AG, Katz MS, et al (2008b). Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. *J Clin Oncol*, **26**, 2373-8.

Nicolaie M, Taylor J, Legrand C (2015a). Vertical modeling: analysis of competing risks data with a cure proportion. arXiv preprint arXiv:1508.03821.

Nicolaie M, van Houwelingen H, Putter H (2015b). Vertical modelling: Analysis of competing risks data with missing causes of failure. *Stat Methods Med Res*, **24**, 891-908.

Nicolaie M, van Houwelingen HC, Putter H (2010). Vertical modeling: A pattern mixture approach for competing risks modeling. *Stat Med*, **29**, 1190-205.
Parkin DM, Bray F, Ferlay J, et al (2005). Global cancer statistics, 2002. *CA Cancer J Clin*, **55**, 74-108.
Peng Y, Dear KB (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, **56**, 237-43.
Peng Y, Zhang J (2008). Estimation method of the semiparametric mixture cure gamma frailty model. *Stat Med*, **27**, 5177-94.
Pepe MS, Mori M (1993). Kaplan-meier, marginal or conditional probability curves in summarizing competing risks failure time data?. *Stat Med*, **12**, 737-51.
Perou CM, Sørlie T, Eisen MB, et al (2000). Molecular portraits of human breast tumours. *Nature*, **406**, 747-52.
Prentice RL, Kalbfleisch JD, Peterson Jr AV, et al (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541-54.
Price DL, Manatunga AK (2001). Modelling survival data with a cured fraction using frailty models. *Stat Med*, **20**, 1515-27.
Rama R, Swaminathan R, Venkatesan P (2010). Cure models for estimating hospital-based breast cancer survival. *Asian Pac J Cancer Prev*, **11**, 387-91.
Rondeau V, Schaffner E, Corbière F, et al (2013). Cure frailty models for survival data: Application to recurrences for breast cancer and to hospital readmissions for colorectal cancer. *Stat Methods Med Res*, **22**, 243-60.
Safe M, Faradmal J, Mahjub H (2016). A comparison between cure model and recursive partitioning: A retrospective cohort study of Iranian female with breast cancer. *Comput Math Methods Med*, **2016**, 9425629.
Sant M, Capocaccia R, Verdecchia A, et al (1998). Survival of women with breast cancer in Europe: variation with age, year of diagnosis and country. The Eurocare working group. *Int J Cancer*, **77**, 679-83.
Shenouda MN, Sadek BT, Goldberg SI, et al (2014). Clinical outcome of isolated locoregional recurrence in patients with breast cancer according to their primary local treatment. *Clin Breast Cancer*, **14**, 198-204.
Siegel RL, Miller KD, Jemal A (2017). Cancer statistics, 2017. *CA Cancer J Clin*, **67**, 7-30.
Team RDC (2008). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Retrieved from http://www.R-project.org.
Van Maaren MC, de Munck L, de Bock GH, et al (2016). 10 year survival after breast-conserving surgery plus radiotherapy compared with mastectomy in early breast cancer in the Netherlands: a population-based study. *Lancet Oncol*, **17**, 1158-70.
van Tienhoven G, Voogd A, Peterse J, et al (1999). Prognosis after treatment for loco-regional recurrence after mastectomy or breast conserving therapy in two randomised trials (EORTC 10801 and DBCG-82TM). *Eur J Cancer*, **35**, 32-8.
Warren LE, Ligibel JA, Chen YH, et al (2016). Body mass index and locoregional recurrence in women with early-stage breast cancer. *Ann Surg Oncol*, **23**, 3870-9.
Wienke A (2010a). Frailty models in survival analysis. In Eds CRC Press, pp 55-64.
Wienke A (2010b). Frailty models in survival analysis. In Eds CRC Press, pp 30-1.
Zhang J, Peng Y (2009). Accelerated hazards mixture cure model. *Lifetime Data Anal*, **15**, 455.