RESEARCH ARTICLE

Application of a Non-Mixture Cure Rate Model for Analyzing Survival of Patients with Breast Cancer

Ahmad Reza Baghestani[1], Sahar Saeedi Moghaddam[2, 1]*, Hamid Alavi Majd[1], Mohammad Esmaeil Akbari[2], Nahid Nafissi[2], Kimiya Gohari[1]

Abstract

<u>Background</u>: As a result of significant progress made in treatment of many types of cancers during the last few decades, there have been an increased number of patients who do not experience mortality. We refer to these observations as cure or immune and models for survival data which include cure fraction are known as cure rate models or long-term survival models. <u>Materials and Methods</u>: In this study we used the data collected from 438 female patients with breast cancer registered in the Cancer Research Center in Shahid Beheshti University of Medical Sciences, Tehran, Iran. The patients had been diagnosed from 1992 to 2012 and were followed up until October 2014. We had to exclude some because of incomplete information. Phone calls were made to confirm whether the patients were still alive or not. Deaths due to breast cancer were regarded as failure. To identify clinical, pathological, and biological characteristics of patients that might have had an effect on survival of the patients we used a non-mixture cure rate model; in addition, a Weibull distribution was proposed for the survival time. Analyses were performed using STATA version 14. The significance level was set at $P \leq 0.05$. <u>Results</u>: A total of 75 patients (17.1%) died due to breast cancer during the study, up to the last follow-up. Numbers of metastatic lymph nodes and histologic grade were significant factors. The cure fraction was estimated to be 58%. <u>Conclusions</u>: When a cure fraction is not available, the analysis will be changed to standard approaches of survival analysis; however when the data indicate that the cure fraction is available, we suggest analysis of survival data via cure models.

Keywords: Breast cancer - long-term survival analysis - non-mixture cure rate model

Introduction

Medical researchers are largely interested in studying the effect of multi-explanatory variables on the time to occurrence of a specific event; the event can be death, relapse of a specific event, divorce, recurrence of a disease, and so forth. When the analysis of survival data is mainly aimed to estimate the effect of a covariate, the Cox's proportional hazard model as a semi-parametric approach is the most frequently-used method (Yu, 2008). Typically when using models for survival data, if the period of follow-up is sufficiently long, all patients will be eventually susceptible to experience the event under the study (Cox, 1972). One of the most important characteristics of survival data is the presence of censored observation, that are actually either lost during the follow-up or are the cases who do not experience the event under the study during the period of study. Also, for some outcomes, there may be some cases who never face the event of interest (Lambert, 2007).

As a result of significant progresses made in treatment of many types of cancers during the last few decades, there

have been an increased number of patients who do not experience the event under study (Lambert et al., 2007; Rondeau et al., 2013). The cases that never experience the event of interest are referred to as cured, immune, or non-susceptible. For example, researchers may be interested in analyzing the rejection of an organ transplanted. Many cases may never reject transplanted organ; therefore, a cured fraction of the population exists. With present of cure fraction in data, standard survival analysis methods such as Cox's proportional hazard model or log-rank test may not be appropriate because they do not account for the possibility of cure (Arano et al., 2010). Cure rate models (also known as long-term survival models) which are a special type of survival analysis models were developed to be used in such a situation.

Cure fraction can be of importance for patients, clinicians, and policy makers, and it may give more valuable insights into time trends in survival analysis of cancer patients (Andersson et al., 2011). Due to significant progress in therapies in recent years, there has been an increased focus on analyzing survival data via these type of models in all fields (Corbière and Joly, 2007; Ortega

[1]*Department of Biostatistics, Faculty of Paramedical Sciences*, [2]*Cancer Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran *For correspondence: s.saeedimoghaddam@gmail.com*

Ahmad Reza Baghestani et al

et al., 2014). Moreover, these survival models are known as split population models in the field of economists and other social scientists (Schmidt and Witte, 1989). Cure rate models are applicable for various types of cancers such as prostate, breast, melanoma, non-Hodgkin's lymphoma, leukemia, and head and neck cancer, in which a significant proportion of cases are immune of the event under study (Chen et al., 1999; Ortega et al., 2008). When the survival curve has a stable plateau at the end of the study it is recommended to use cure rate models for analyzing the survival data (Kim et al., 2013). Further, in the absence of cure fraction, cure rate models will be changed to standard survival analysis models (Yu, 2008).

The aim of the current study was to investigate the effects of clinical, pathological, and biological characteristics of patients on survival of patients with breast cancer using non-mixture cure rate model.

Materials and Methods

We studied 438 female patients with breast cancer who were visited and treated at Cancer Research Center in Shahid Beheshti University of Medical Sciences, Tehran, Iran. These patients were diagnosed from 1992 to 2012 and followed up until October 2014. We had to exclude some patients because of their incomplete information. The patients or patients' family members were contacted via phone calls to confirm whether the patients are still alive or not. Deaths due to breast cancer were regarded as failure and survival time was calculated as the time interval between date of cancer diagnosis and date of death due to breast cancer (for those who died) and date of the last follow-up (for those who were alive).

There are two major approaches in cure rate models. The oldest one which was introduced by Boag in 1949 is the mixture cure rate model, also known as standard cure rate model (Boag, 1949). In this model we assume that a certain fraction of the population are cured and the remaining are not cured (Maller and Zhou, 1996). In order to model cure fraction, parametric and non-parametric methods are used in this approach (Achcar et al., 2012). Second approach which was introduced by Chen et al in 1999 is the non-mixture cure rate model (Chen et al., 1999), also known as bounded cumulative hazard model and promotion time cure model (Tsodikov et al., 2003; Tournoud and Ecochard, 2008). As one of the advantages of the non-mixture cure rate model, it has a proportional hazards model (Lambert et al., 2007). In this approach it is assumed that an individual is left with N_i metastatic-component cancer cells after treatment (Yakovlev et al., 1996; Tsodikov et al., 2003). In this study we assumed that N_i had a Poisson distribution.

For data analysis, clinical, pathological, and biological characteristics of patients were assessed in the cure model. A Weibull distribution was proposed for the survival time. The data were analyzed using STATA version 14 and the figure was prepared by R version 3.1.2. Quantitative results were expressed as mean ± standard deviation. The significance level was set at $P \leq 0.05$.

Results

Patients' characteristics are presented in Table 1. A total of 438 women with breast cancer were enrolled into this analysis. The mean of age at the time of diagnosis was 48.36 ± 10.90 years. Patients' age ranged from 22

Table 1. Clinical, Pathological, and Biological Characteristics in Patients with Breast Cancer

Factors	No. of patients	Percentage (%)
Age at diagnosis		
Under 40 yrs.	88	20.09
40 to 70 yrs.	336	76.71
Over 70 yrs.	14	3.2
Education		
Illiterate	37	8.45
Elementary school	51	11.64
Secondary school	53	12.1
Diploma	161	36.76
Academic	136	31.05
Marital		
Single	24	5.48
Married	414	94.52
Abortion		
Yes	153	34.93
No	285	65.07
Breastfeeding		
Yes	381	86.99
No	57	13.01
Family history		
Yes	124	28.31
No	314	71.69
Type of surgery		
BCS	274	62.56
MRM	164	37.44
Lymph node status		
N0	198	45.21
N1	121	27.63
N2	39	8.9
N3	80	18.26
Tumor size		
T1	111	25.34
T2	242	55.25
T3	85	19.41
Stage		
I	96	21.92
II	193	44.06
III	135	30.82
IV	14	3.2
Histologic grade		
Well differentiated	47	10.73
Moderately differentiated	235	53.65
Poorly differentiated	156	35.62
Estrogen Receptor		
Positive	312	71.23
Negative	126	28.77
Progesterone Receptor		
Positive	288	65.75
Negative	150	34.25
Lymphovascular Invasion		
Positive	172	39.27
Negative	266	60.73
Recurrence		
Yes	70	15.98
No	368	84.02
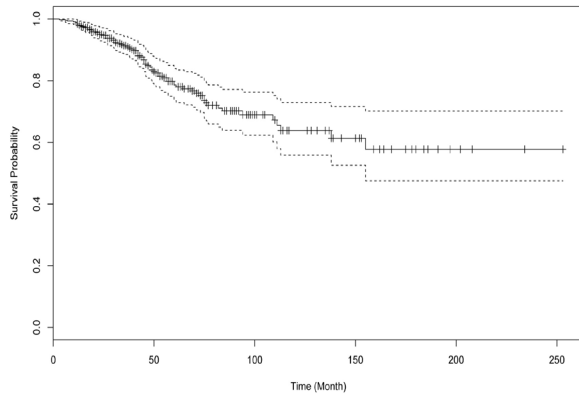
Figure 1. Kaplan-Meier Survival Curve of the Entire Study Population

Table 2. Estimation Based on Non-mixture Cure Rate Model

Factors	HR	SD	P-value
Age at diagnosis			
Under 40 yrs.	0.7	1.504	0.642
40 to 70 yrs.	1.951	1.466	0.183
Over 70 yrs. **	1		
Marital			
Single	0.898	1.371	0.512
Married **	1		
Abortion			
Yes	0.666	0.498	0.181
No**	1		
Breastfeeding			
Yes	0.614	0.782	0.432
No **	1		
Family history			
Yes	0.18	0.53	0.735
No **	1		
Type of surgery			
BCS	0.952	0.49	0.052
MRM **	1		
Lymph node status			
N0	3.099	1.022	0.002 *
N1	1.737	0.78	0.026 *
N2	0.514	0.791	0.516
N3 **	1		
Stage			
I	1.662	1.905	0.383
II	2.272	1.804	0.208
III	2.434	1.689	0.149
IV **	1		
Histologic grade			
Well differentiated	4.067	1.74	0.019 *
Moderately differentiated	1.508	0.561	0.007 *
Poorly differentiated **	1		
Estrogen Receptor			
Positive	0.423	0.727	0.56
Negative **	1		
Progesterone Receptor			
Positive **	1		
Negative	0.1	0.706	0.887
Lymphovascular Invasion			
Positive **	1		
Negative	0.319	0.498	0.522

* Significant at the 5% level; ** Stands for a control group

to 84 years. The longest duration of follow-up was 253 months. A total of 75 patients (17.08%) died due to breast cancer during the study up to October 2014. The Kaplan-Meier observed curve, with 95% confidence interval, is illustrated in Figure 1. As the Figure shows, the curve reached a plateau at the right tail, so using the cure rate model led to more accurate results. The prognostic variables included in the model were age at diagnosis, marital status, history of abortion, history of breastfeeding, family history of breast cancer, type of surgery, number of metastatic lymph nodes, tumor size, stage, histologic grade, estrogen receptor, progesterone receptor, and lymphovascular invasion. According to the results which are presented in Table 2, number of metastatic lymph nodes and histologic grade were significant. In addition cure fraction was estimated to be 58%.

Discussion

One of the appropriate and useful ways to investigate any progresses in cancer therapy is to monitor trends in survival of patients over time (Bakar et al., 2008). The time from diagnosis until occurrence of a specific event under study is the most valuable measure of care for patients, clinicians, and even policy makers (Andersson et al., 2011). Therapies for many types of diseases especially cancers have been improved significantly in the recent years, so we are facing an increase in the proportion of patients who do not experience the event under study, i.e. those who are not susceptible to the occurrence of the event under study (Asano et al., 2014). The proportion of such observations is known as cure fraction. In addition, survival analysis via considering cure fraction is called cure rate models or long-term survival models and they perform a considerable function in survival analysis (Ortega et al., 2014).

For the past 20 years or more, several studies have been conducted on cure rate models and they have been well developed in the statistical literature (Rama et al., 2010; Achcar et al., 2012; Othus et al., 2012; Akhlaghi et al., 2013; Yu et al., 2013; Rahimzadeh et al., 2014), but the models are not common in all fields. More specifically, most of them are focused on cancers-related subjects. For cancers in which some cases may have long-term survival, cure rate models can be an interesting approach to characterize and study the survival of patients. It is worth mentioning that cure rate models are studied because of their optimal properties which may better explain the responses to the variables at different times (Akhlaghi et al., 2013). There are two general types of cure models, mixture and non-mixture. Mixture cure rate models may fit some data better than non-mixture and vice versa (Othus et al., 2012), however in this report we only study the non-mixture cure rate model. Non-mixture cure rate model can explain remaining cancer cells after therapy by using latent variable (Rahimzadeh et al., 2014).

Breast cancer is the first common cancer and the third leading cause of death among Iranian women (Sadjadi et al., 2005); in addition, it's mortality is still lower than that in Western industrialized countries (Taghavi et al., 2012). The aim of this study was to find factors that may have

an effect on the survival of patients with breast cancer including clinical, pathological, and biological variables. Based on our findings, the number of metastatic lymph nodes and histologic grade were significant. In our study a Weibull distribution was proposed for survival time but it was not restricted to Weibull and some other lifetime distributions could be also considered (Borges et al., 2012); However, Jafari-Koshki's study indicated that cure model with Weibull distribution had the smallest DIC in their report and fitted the data best (Jafari-Koshki et al., 2014). In our study we used logistic link function; as a matter of fact, determining the link function is important because it leads to different assumptions about the joint effect of covariates (Lambert et al., 2007).

Most of the studies on cure models are conducted on prostate or breast cancers, but Sposto used mixture and non-mixture cure models to analyze children's cancer group data (Sposto, 2002). The majority of studies investigating the factors affecting the survival of patients with breast cancer mainly used Log-rank test and Cox analysis. Rahimzadeh's study indicated that in the survival analysis of breast cancer patients with cure rate model, metastasis and stage had a significant effect (Rahimzadeh et al., 2014). In this study data for some clinical, pathological, and biological variables were missing and some were recorded in a wrong way so that they had become useless. In addition, our study was carried out based on the data collected from a single institution; large sample size is one of the desirable properties required for cure models, and can reduce the problems related to convergence and fit of the model (Jafari-Koshki et al., 2014). Nevertheless, in our study the model converged after a few iterations. It is recommended to conduct further studies by assuming other distributions rather than Poisson for Ni and Weibull for lifetime.

Acknowledgements

References

Abu Bakar MR, Salah KA, Ibrahim NA, et al (2008). Cure fraction, modelling and estimating in a population-based cancer survival analysis. *Malaysian J Mathematical Sciences*, **2**, 113-34.

Achcar JA, Coelho-Barros EA, Mazucheli J (2012). Cure fraction models using mixture and non-mixture models. *Tatra Mountains Mathematical Publications*, **51**, 1-9.

Akhlaghi AA, Najafi I, Mahmoodi M, et al (2013). Survival analysis of iranian patients undergoing continuous ambulatory peritoneal dialysis using cure model. *J Research in Health Sci*, **13**, 32-6.

Andersson TM, Dickman PW, Eloranta S, et al (2011). Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Med Res Methodol*, **11**, 96.

Arano I, Sugimoto T, Hamasaki T, et al (2010). Practical application of cure mixture model for long-term censored survivor data from a withdrawal clinical trial of patients with major depressive disorder. *BMC Med Res Methodol*, **10**, 33.

Asano J, Hirakawa A, Hamada C (2014). Assessing the prediction accuracy of cure in the Cox proportional hazards cure model: an application to breast cancer data. *Pharmaceutical statistics*, **13**, 357-63.

Boag JW (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society. *Series B (Methodological)*, **11**, 15-53.

Borges P, Rodrigues J, Louzada F, et al (2012). A cure rate survival model under a hybrid latent activation scheme. *Statistical methods in medical research*, 0962280212469682.

Chen M-H, Ibrahim JG, Sinha D (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**, 909-19.

Corbière F, Joly P (2007). A SAS macro for parametric and semiparametric mixture cure models. *Computer Methods And Programs In Biomedicine*, **85**, 173-80.

Cox DR (1972). Regression models and life tables (with discussion). journal of the royal statistical society, series b, **34**, 187-220.

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.

Kim S, Zeng D, Li Y, et al (2013). Joint Modeling of longitudinal and cure-survival data. *J Statistical theory and Practice*, **7**, 324-44.

Lambert PC (2007). Modeling of the cure fraction in survival studies. *Stata Journal*, **7**, 351.

Lambert PC, Thompson JR, Weston CL, et al (2007). Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, **8**, 576-94.

Maller RA, Zhou X 1996. Survival Analysis with Long-Term Survivors, Wiley.

Ortega EM, Barriga GD, Hashimoto EM, et al (2014). A new class of survival regression models with cure fraction. *J Data Science*, **12**, 107-36.

Ortega EM, Cancho VG, Lachos VH (2008). Assessing influence in survival data with a cure fraction and covariates.

Othus M, Barlogie B, LeBlanc ML, et al (2012). Cure models as a useful statistical tool for analyzing survival. *Clinical Cancer Research*, **18**, 3731-6.

Rahimzadeh M, Baghestani AR, Gohari MR, et al (2014). Estimation of the cure rate in Iranian breast cancer patients. *Asian Pac J Cancer Prev*, **15**, 4839-42.

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. *Statistical Methods in Medical Research*, **22**, 243-60.

Sadjadi A, Nouraie M, Mohagheghi Mohammad A, et al (2005). Cancer occurrence in Iran in 2002, an international perspective. *Asian Pac J Cancer Prev*, **6**, 359.

Schmidt P, Witte AD (1989). Predicting criminal recidivism using 'split population'survival time models. *J Econometrics*, **40**, 141-59.

Sposto R (2002). Cure model analysis in cancer: an application

to data from the Children's Cancer Group. *Statistics in Medicine*, **21**, 293-312.

Taghavi A, Fazeli Z, Vahedi M, et al (2012). Increased trend of breast cancer mortality in Iran. *Asian Pac J Cancer Prev*, **13**, 367-70.

Tournoud M, Ecochard R (2008). Promotion time models with time changing exposure and heterogeneity: application to infectious diseases. *Biometrical Journal*, **50**, 395-407.

Tsodikov A, Ibrahim J, Yakovlev A (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *J Am Statistical Association*, **98**.

Yakovlev AY, Tsodikov AD, Asselain B 1996. Stochastic models of tumor latency and their biostatistical applications, OECD Publishing.

Yu B (2008). A frailty mixture cure model with application to hospital readmission cata. *Biometrical J*, **50**, 386-94.

Yu X, De Angelis R, Andersson TM, et al (2013). Estimating the proportion cured of cancer: some practical advice for users. *Cancer Epidemiol*, **37**, 836-42.