

RESEARCH COMMUNICATION

Modeling of Breast Cancer Prognostic Factors Using a Parametric Log-Logistic Model in Fars Province, Southern Iran

Najaf Zare¹, Marzieh Doostfateme¹, Abass Rezaianzadeh^{2*}

Abstract

In general, breast cancer is the most common malignancy among women in developed as well as some developing countries, often being the second leading cause of cancer mortality after lung cancer. Using a parametric log-logistic model to consider the effects of prognostic factors, the present study focused on the 5-year survival of women with the diagnosis of breast cancer in Southern Iran. A total of 1,148 women who were diagnosed with primary invasive breast cancer from January 2001 to January 2005 were included and divided into three prognosis groups: poor, medium, and good. The survival times as well as the hazard rates of the three different groups were compared. The log-logistic model was employed as the best parametric model which could explain survival times. The hazard rates of the poor and the medium prognosis groups were respectively 13 and 3 times greater than in the good prognosis group. Also, the difference between the overall survival rates of the poor and the medium prognosis groups was highly significant in comparison to the good prognosis group. Use of the parametric log-logistic model - also a proportional odds model - allowed assessment of the natural process of the disease based on hazard and identification of trends.

Keywords: Survival analysis - log-logistic model - proportional odds - prognostic factors - breast cancer - Iran

Asian Pacific J Cancer Prev, **13**, 1533-1537

Introduction

In general, breast cancer is the most common malignancy among women in developed as well as some developing countries. Breast cancer is the second leading cause of cancer mortality after lung cancer (Nagel et al., 2004; Fisch et al., 2005; Grau et al., 2005). Based on the report by the Disease Control and Prevention Center of Iran Ministry of Health and Medical Education, breast cancer is the most common type of cancer among Iranian women and allocates 21% of all malignancies. The prevalence of breast cancer is estimated between 8 – 10% in Europe and the U.S. Moreover, the lowest prevalence of the disease is reported as 1% in Asia. The prevalence of breast cancer in Iran is estimated 6.7 in 1000 people, which is lower than its global measure. Although a great number of studies have been conducted on the risk factors as well as the prognostic factors of breast cancer around the world, just a few studies have investigated the issue in Iran (Al-Moundhri et al., 2004; Foo et al., 2005; Kim et al., 2005).

Iran has a population of 75 million people, and most of the studies on breast cancer are performed in the capital, Tehran, which has a population of 14 million people. However, the focus of the most of those studies has not been the natural course and prognostic factors. Also,

not many professional studies have been conducted on breast cancer in Southern Iran which has a population of approximately 4 million people. Just one study was done on the survival rate of breast cancer and its relationship with the socioeconomic, demographic and clinicopathological factors.

In general, medical texts which are related to survival are full of Kaplan-Meier survival curves, and the hazard functions are seen in a few of them and in a few of published articles regarding survival. For example, the Kaplan-Meier graph (Figure 1) neither reveals any idea of the hazard function, nor does it show how the hazard ratio changes over time.

In addition, the success of the Cox regression model has inevitably caused the researchers to limitedly study the basic hazard function. On the other hand, parametric models lead to more exact estimations of people's survival as well as hazard probabilities over time, and simultaneously result in a better understanding of the phenomena under study – especially in a medical study with a long follow-up period (Royeston et al., 2001)

Recently, as breast cancer has grown and most countries have developed screening programs, a lot of studies have been conducted on the classification of the patients with the diagnosis of breast cancer into prognosis groups in order to perform more effective treatments

¹Department of Biostatistics, ²Research Center for Health Sciences, Department of Epidemiology, Shiraz University of Medical Sciences, Shiraz, Iran *For correspondence: rezaiana@sums.ac.ir

(Blamey et al., 2007). In some of these studies, due to screening programs, timely diagnosis of the disease, and a long follow-up period, the patients benefited in experiencing a much higher survival rate in comparison to previous studies. Therefore, sometimes the patients are categorized even into five prognosis groups and are compared regarding the 5, 10, and sometimes 15-year survival (Sundquist et al., 1999). However, none of these studies have investigated the natural course of the disease based on the hazard function and the hazard rates of the patients in each moment of time. Besides, all analyses have been done using the Cox regression and the Kaplan-Meier methods; and none of them have used advanced statistical approaches for modeling the different levels of prognosis groups.

The present study aims to investigate different parametric odds models and proportional hazards by classifying the patients into three prognosis groups based on the NPI in order to choose the best model which can identify the natural course as well as the survival rates of these three groups. The study, also, examines the natural course of breast cancer as well as the 5-year survival rates of the three prognosis groups under study in Southern part of Iran by using the final model of the hazard function which is the log-logistic model in the present study.

Materials and Methods

From January 2001 to January 2005 a total of 6253 patients with the diagnosis of 10 most common cancers were registered at Cancer Registry Center of Namazi Hospital, Shiraz, Iran, which 1192 of them were breast cancer patients. Forty four of them were excluded, 23 patients due to a previous breast cancer, 14 patients due to bilateral tumors, 2 patients because of clinical reasons, and 5 patients because of a previous cancer. Therefore, the present study includes 1148 women with primary invasive breast cancer who underwent breast surgery with axillary lymph node dissection and radiotherapy. Except for a small number of patients, the rest underwent chemotherapy.

Since no regular screening program is performed in Iran, most of the patients were referred to this center with a palpable mass, and mammography was only performed in order to confirm the presence of the mass.

At the end of the follow-up period (January 2005), from 1148 patients, 859 ones were alive, while 269 ones died because of the cancer. Therefore, 20 patients were lost from the follow-up or were censored. The average of the follow-up time, which was considered as the period from the initial pathological diagnosis to either the time of death or the end of the study, was 34 months.

Statistical methods

At first, the Cox regression was performed in order to classify the patients into the prognosis groups and identify the prognostic factors. Then, based on the coefficients of the five variables of age, menopausal status, tumor grade, tumor size, and the number of lymph nodes involved, just the variables of tumor grade, tumor size, and the number of lymph nodes involved were considered as the prognostic variables in classifying the patients. The classification was done based on the NPI which is mainly used in clinical

studies in order to categorize the patients. In this way, based on the progression of the disease, the patients were classified into three groups of good prognosis, medium prognosis, and poor prognosis.

Then, the parametric log-logistic model (the proportional odds with one degree of freedom) was fitted to the data by the following hazard function (Royston et al., 2002-2003):

In this function, t stands for the survival time, and $o(t; z)$ stands for the chance of the event for each person with z explanatory variables vector. In this function, the z vector is a dual variable (Z_1, Z_2) with a reference level for the good condition of the disease.

Results

Mean age at diagnosis was 47 years, ranged from 19 to 86 years. Distribution of factors involved in the study and result of Cox's regression are presented by Table 1 and 2, respectively. Figure 1 shows the Kaplan-Meier diagram which compares the patients' survival experience based on their classification according to the prognostic index and without taking any particular distribution of the survival times into account. As the diagram depicts, the good and the medium prognosis groups seem to be close to each other regarding the survival experience, while the poor prognosis group shows a great distance with the two.

Moreover, in the present study, the 5-year survival of the patients in the good prognosis group, medium prognosis group, and poor prognosis group with advanced disease, were measured as 74% (about 20% of the patients), 72% (about 44% of the patients), and 30% (about 36% of the patients), respectively. Besides, the overall survival was calculated as 56%.

The hazard functions of all the prognosis groups were measured using the parametric log-logistic model. Figure 2 shows the patients' hazards based on their prognosis groups. In addition, the hazard ratio, based on the log-logistic model, in the medium and the poor prognosis groups is depicted in Figure 3. Observing the two diagrams, it is quite obvious that the poor prognosis group is at a high risk of death in the first 3-4 years after the diagnosis and may even experience death. After that, the hazard quickly decreases and reaches a fixed rate. The medium and the good prognosis groups, also, show an increasing trend of the hazard, which gets close to a fixed trend toward the end of the study. This implies that if a patient in the advanced status of the disease does not

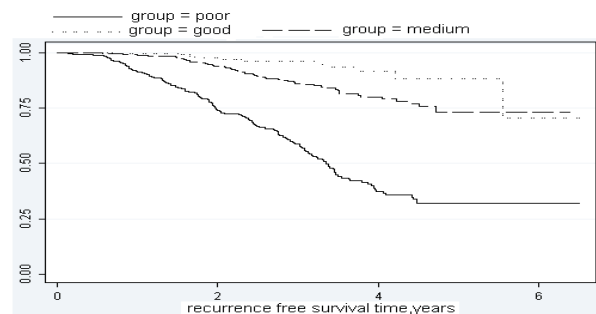


Figure 1. Kaplan-Meier Diagram Classifying the Three Prognosis Groups

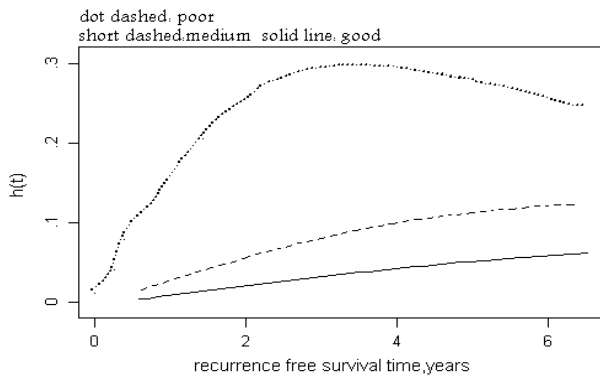


Figure 2. The Hazard Function Diagram, the Log-Logistic Model.

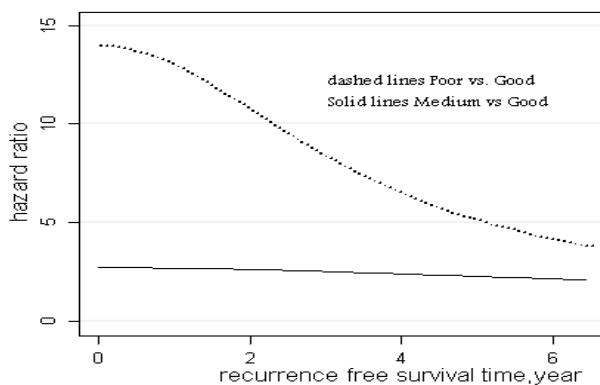


Figure 3. The Diagram of the Comparison of the Hazard Ratio in the Medium and the Good Prognosis Groups, the Log-Logistic Model.

Table 1. Distribution of 5 Factors Involved in the Study

		Percent	Number
Age, year	<36	15.0	172
	36 - 49	45.2	519
	50 - 64	29.8	342
	65+	7.4	85
Nodal Status(N):	0	30.4	349
	1-3	30.1	346
	4+	12.7	146
Tumor size ,cm:	<3	10.7	123
	>3 & <57.1	82.0	
	5+	26.0	298
Grade:	I	20.4	234
	II	45.6	523
	III	15.0	250
Menopause status:	Yes	55.5	637
	No	21.8	172
Prognostic Index:	Group1(PI< 3.4)	42.0	482.0
	Group2(3.4< PI <5.4)	58.0	666
	Group3 (PI >5.4)	25	286

Table 2. Result of Cox's Regression Model

Prognostic factor	Hazard ratio	SE	Coefficient	95% CI
Tumor size	1.41	0.10	0.34	1.17-1.71
Tumor grade	1.76	0.11	0.57	1.42-2.19
Nodal Status(N)	2.71	0.11	0.99	2.18-3.36
Menopause status	1.07	0.15	0.07	0.80-1.43

experience death in the first 3-4 years after the diagnosis, s/he will have more chance to live in the following years since the hazard decreases. On the other hand, the good and the poor prognosis groups face a lower hazard as well as an increasing death hazard with a slow slope in the follow-up period. In addition, the hazard ratio for both groups is highly significant in comparison to the good prognosis group. It is quite clear that the parametric model shows the natural process of the disease during the study. This issue will be completely explained in the discussion section.

Discussion

The present study aimed to propagate an approach which is flexible, clinically interpretable, robust, clear, and at the same time reasonably available in the standard softwares. Parametric models are utilized in modeling the continuous variables in medicine and epidemiology. Besides, the interpretation of regression parameters in our model is just the same as that of the simple Cox regression model. However, for further studies, it is quite beneficial to compare the estimations of the hazard function and the hazard ratio to the simulation studies and other estimations.

We conducted our statistical models based on the data gathered from 1148 women with primary invasive breast cancer who were registered at Cancer Registry Center of Namazi Hospital, Shiraz, Iran, and had underwent the breast surgery.

Based on the results obtained from the sample, the significance of the three variables of tumor grade, tumor size, and the number of involved lymph nodes in the patients' survival experience, and NPI, the patients were classified into three prognosis groups regarding their survival experience. Also, various methods and models were employed in order to compare the survival in the three groups under study. In the end, we came to the conclusion that based on the Akaike statistics, which was the base of selection among the existing models, we had achieved the best model with odds scale which could clearly compare the three groups' survival experience based on the hazard function.

Moreover, based on this model, we found out that the prognosis of the three groups of patients is significantly different from each other. Also, it was revealed that at the beginning of the study, the hazard rate of death in the poor and the medium prognosis groups has increased respectively 13 and 3 times in comparison to the good prognosis group. In other words, at the beginning of the study, the patients of the poor and the medium prognosis groups were respectively 13 and 3 times more at a higher risk of death in comparison to the patients of the good prognosis group. Of course, this amount decreased respectively to 4 and 2 at the end of the study. Moreover, the log-logistic proportional odds model was chosen as the best model with the odds scale since it could explain the natural process of the disease and had the lowest amount of Akaike among the existing models.

Observing the two diagrams of the hazard function and the hazard ratio, it is quite clear that the poor prognosis

group is at a high risk of death in the first 3-4 years after the diagnosis and may even experience death (Figure 2, b). Then, the hazard quickly decreases and reaches a fixed rate after several years. This implies that if a patient in the advanced status of the disease does not die in the first 3-4 years after the diagnosis, s/he will have more chance to survive in the following years since the hazard decreases. On the other hand, the good and the medium prognosis groups face a lower hazard. Actually, the hazard increases with a slow slope and reaches a fixed rate at the end.

One of the aims of the present study is to statistically and clinically evaluate the data and compare these two aspects with other studies. A great number of studies have been conducted in Iran as well as other countries which clinically investigate breast cancer, survival of the patients suffering from breast cancer, and factors affecting the disease using the simple Cox regression model. However, regarding the statistical aspect, classification of the patients into three prognosis groups based on the significant factors obtained from the Cox regression model, comparison of the survival experience of the three groups with the advanced statistical topics such as the parametric regressions, parametric hazard function, and the explicit investigation of the hazard function in order to study the historical process of the breast cancer, this study has been performed for the first time in Iran.

Considering the sample size under study on breast cancer, Rezaianzadeh et al. (2009), conducted the most wide-spread study on the investigation of the survival and its affecting factor in Southern part of Iran utilizing the Cox regression model. In that study – based on which the present study is conducted – the three main factors of our study were revealed to be the most effective pathological factors in the patients' survival. However, no attempt was made in order to separately classify the patients in that study. On the other hand, just their overall survival was investigated based on life tables and simple Kaplan-Meier diagrams.

Todd et al., 1987, conducted a study on 379 patients with breast cancer in Nottingham. They classified these patients into three prognosis groups based on the NPI, and used the simple Kaplan-Meier method in order to compare the three groups' survivals. Moreover, using life tables, the 5-year survival of the good, medium, and poor prognosis groups were computed as 88% (about 33% of the patients), 69% (about 52% of the patients), and 22% (about 15% of the patients), respectively. Besides, the overall survival rate was calculated as 64%. Although that study had a 20-year time interval, it shows a significant difference with the overall survival in the present study. Also, the good and the medium prognosis groups included more patients in comparison to the poor prognosis group (85% vs. 68%) which might be due to the lack of screening programs as well as the undesirable level of public awareness.

In addition, Galea et al. (1992), conducted a study based on the same index in. The 15-year survival of the patients under study was measured as 83%. Also, it was measured as 80%, 42%, and 13% (about 29%, 54%, and 17% of the patients, respectively) for the three prognosis groups. Although a 5-year survival as well as the time

interval existed in that study, a significant difference revealed to be there regarding both the percentage of the patients in each group and their survival rates.

Finally, regarding the statistical aspects, Royeston et al. (2001), in the first statistical approach, proposed a flexible model in order to classify the groups. In this way, not only they could compare the groups' survivals, they could also investigate the process of the disease under study as well as the process of changes in the treatment and the variables in a period of time. Actually, they presented a flexible parametric model which could clearly estimate the hazard function. In addition, they used the practical instances of breast cancer in the data which was gathered from German Registry Center in 1984 in order to fit their model. Moreover, in 2002, they introduced the parametric odds model using the same data. They continued their work to the relative survival model in 2009. The results of the present study were statistically different from the two studies mentioned above. In Royeston's study, the parametric flexible odds model with two degree of freedom was considered as the best model in explaining the natural process of the study. In our study, on the other hand, the proportional odds form of the log-logistic model revealed the best fitness to the data. The researchers of the present study utilized the parametric log-logistic model – which is a proportional odds model, as well, and managed to explicitly investigate the natural course of the disease, based on the hazard function and the hazard ratio function, and identify the increasing as well as the decreasing trend of the hazard during the study.

Acknowledgements

This article is adopted from the proposal number 88-4778 and Ms. Doostfateme's M.S. thesis. We would like to thank the Research Deputy of Shiraz University of Medical Sciences for their financial support. The Cancer Registry Center of Namazi Hospital, Shiraz, Iran is, also, appreciated for their cooperation in gathering the data.

References

- Abadi A , Saadat S, Yavari P, Bajdik C, Jalili P (2011). Comparison of Aalen's additive and Cox proportional hazards models for breast cancer survival: analysis of population-based data from British Columbia, Canada. *Asian Pac J Cancer Prev*, **12**, 3113-16
- Al-Moundhri M, Al-Bahrani B, Pervez I, et al (2004). The outcome of treatment of breast cancer in a developing country – Oman. *The Breast*, **13**, 139-45.
- Ardoino I, Ambrogi F, Bajdik C, et al (2012). Flexible parametric modelling of the hazard function in breast cancer studies. *Applied Statistics*, Taylor & France.
- Blamey RW, Ellis IO, Pinder SE, et al (2007). Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990–1999. *Eur J Cancer*, **43**, 1548-55.
- Collett D(2003). *Modelling Survival Data in Medical Research*. Chapman and Hall.
- D'Eredita G, Martellotta M, Natale T, Ferrarese F(2001). Prognostic factors in breast cancer: the predictive value of the Nottingham Prognostic Index in patients with a long

- term follow-up and treated in a single institution. *Eur J Cancer*, **37**, 591-6.
- Fisch T, Pury P, Probst N, et al (2005). Variation in survival after diagnosis of breast cancer in Switzerland. *Ann Oncol*, **16**, 1882-8.
- Foo CS, Su D, Chong CK, et al (2005), Breast cancer in young Asian women: Study on survival. *ANZ J Surg*, **75**, 566-72.
- Galea MH, Blamey RW, Elston CE, Ellis IO (1992). The Nottingham Prognostic Index in primary breast cancer. *Br Cancer Res Treat*, **2**, 207-21.
- Grau AM, Ata A, Foster L, et al (2005). Effect of race on long-term survival of breast cancer patients: Transinstitutional analysis from an inner city hospital and university medical center. *Am Surg*, **71**, 164-70.
- Kalbfleish JD, Prentice RL(2002). *The Statistical Analysis of Failure Time Data*. Wiley-Interscience.
- Kim KJ, Huh SJ, Yang JH, et al (2005). Treatment results and prognostic factors of early breast cancer treated with a breast conserving operation and radiotherapy. *Jpn J Clin Oncol*, **35**, 126-33.
- Lambert PC, Royston P (2009). Further development of flexible parametric models for survival analysis. *The Stata J*, **9**, 265-90
- Nagel G, Wedding U, Hoyer H, Röhrig B, Katenkamp D(2004). The impact of comorbidity on the survival of postmenopausal women with breast cancer. *J Cancer Res & Clin Oncol*, **130**, 664-70.
- Rezaianzadeh A , Peacock J, Reidpath D, et al (2009). Survival analysis of 1148 women diagnosed with breast cancer in Southern Iran. *BMC Cancer*, **9**, 168-78
- Royston P (2001). Flexible parametric alternatives to the Cox model, and more. *Stata J*, **1**, 1-28.
- Royston P, Parmar MK(2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Med*, **21**, 2175-97.
- Royston P, Sauerbrei W (2003). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials source. *J Royal Stat Soc*, **162**, 71-94.
- Sundquist M, Thorstenson S, Brudin L, Nordenskjöld B (1999). Applying the Nottingham prognostic index to a Swedish population. *Breast Cancer Res Treat*, **53**, 1-8.
- Todd JH, Dowle C, Williams MR, et al (1987). Confirmation of a Prognostic Index in primary breast cancer. *Br J Cancer*, **56**, 489-92.
- Venkatesan P, Raman TT, Ponnuraja C(2011). Survival analysis of women with breast cancer under adjuvant therapy in South India. *Asian Pac J Cancer Prev*, **12**, 1533-5.
- Yu K, Lee CH, Tan PH, Hong GS, Wee SB, Wong CY, Tan P(2004). A molecular signature of the Nottingham Prognostic Index in breast cancer. *Cancer Res*, **64**, 2962-68.