

RESEARCH ARTICLE

Survival Analysis of Patients with Breast Cancer using Weibull Parametric Model

Ahmad Reza Baghestani¹, Sahar Saeedi Moghaddam^{1,2*}, Hamid Alavi Majd¹,
Mohammad Esmaeil Akbari², Nahid Nafissi², Kimiya Gohari¹

Abstract

Background: The Cox model is known as one of the most frequently-used methods for analyzing survival data. However, in some situations parametric methods may provide better estimates. In this study, a Weibull parametric model was employed to assess possible prognostic factors that may affect the survival of patients with breast cancer. **Materials and Methods:** We studied 438 patients with breast cancer who visited and were treated at the Cancer Research Center in Shahid Beheshti University of Medical Sciences during 1992 to 2012; the patients were followed up until October 2014. Patients or family members were contacted via telephone calls to confirm whether they were still alive. Clinical, pathological, and biological variables as potential prognostic factors were entered in univariate and multivariate analyses. The log-rank test and the Weibull parametric model with a forward approach, respectively, were used for univariate and multivariate analyses. All analyses were performed using STATA version 11. A P-value lower than 0.05 was defined as significant. **Results:** On univariate analysis, age at diagnosis, level of education, type of surgery, lymph node status, tumor size, stage, histologic grade, estrogen receptor, progesterone receptor, and lymphovascular invasion had a statistically significant effect on survival time. On multivariate analysis, lymph node status, stage, histologic grade, and lymphovascular invasion were statistically significant. The one-year overall survival rate was 98%. **Conclusions:** Based on these data and using Weibull parametric model with a forward approach, we found out that patients with lymphovascular invasion were at 2.13 times greater risk of death due to breast cancer.

Keywords: Breast cancer - survival analysis - Weibull parametric model - log-rank test

Asian Pac J Cancer Prev, 16 (18), 8567-8571

Introduction

Cancer is a general term used for a group of diseases that cause abnormal cells to divide without control and overpass other tissues. In addition, if they expand out of control, cancer can result in death (American Cancer Society, 2014). Based on GLOBOCAN 2012, an estimated 14.1 million new cases of cancer and 8.2 million deaths from cancer occurred in 2012 in both sexes. Estimation of 5-year prevalent cases in 2012 showed that there were 32.5 million people (adult population) alive from both sexes who had a cancer diagnosed during the previous five years (Ferlay et al., 2014). According to the Global Burden of Disease (GBD) study, the estimated rate of Disability-Adjusted Life Years (DALY) attributed to all neoplasms in both sexes worldwide was 2,793 (95% UI: 2,580-2,985) in 1990, which decreased over the time and finally reached 2,736 (95% UI: 2,532-2,889) in 2010 (IHME, 2013). Every one DALY demonstrates the loss of one year of healthy life (May et al., 2015). In addition, 7.6% of global DALYs are assigned to the neoplasms (Murray et al., 2013).

In 2012, the three most leading cancers in both sexes worldwide were lung cancer (13% of the total), breast cancer (11.9%) and colorectal cancer (9.7%); the most common types of cancer in men, respectively, are lung cancer (16.8%), prostate cancer (14.8%) and colorectal cancer (10.1%) while in women they are ordered as breast cancer (25.1%), colorectal cancer (9.2%) and lung cancer (8.8%) (Ferlay et al., 2014). In spite of the reduction of DALYs attributed to neoplasms, in patients with breast cancer DALYs increased from 167 (95% UI: 162-175) in 1990 to 174 (95% UI: 167-184) per 100,000 in 2010 (IHME, 2013). According to the most recent estimates for breast cancer, there are 1.67 million new cases and 0.52 million deaths globally (Ferlay et al., 2014). Similar to the global, breast cancer is the first common cancer and the third leading cause of death among Iranian women (Sadjadi et al., 2005). Also, the age standardized mortality rate of breast cancer among Iranian women increased from 1.40 to 3.52 per 100,000 from 1995 to 2004 (Taghavi et al., 2012), which indicates the need for making further progress in care and services which are necessary for patients who suffer from cancer (Bray et al., 2013).

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

In order to evaluate the effectiveness of cares provided for patients with cancer, especially when the event of death is under the study, we can use survival analysis. Survival analysis is a set of methods used for analyzing a non-negative random variable which measures time to occurrence of a specific event; the event can be death, response to a particular treatment, occurrence, or relapse of a specific disease, marriage, divorce, birth of children, and so forth. These methods can describe the relation between survival and covariates. There are three approaches for survival analysis; nonparametric, semi-parametric, and parametric. Cox semi-parametric method is the most common approach which is used by many researchers because of its fewer assumptions. Among parametric approach, Weibull model is popular and more flexible than Cox semi-parametric model.

The Weibull, which was developed by Waloddi Weibull in 1951, comes originally from engineering issues to analyze the survival data (Weibull, 1951); actually, it has been used to predict the proportion of future failures after it has observed a failed process in a given point of time (Nelson, 2000). It has a hazard rate which is either increasing, decreasing, or constant (Klein and Moeschberger, 2003). If the hazard rate is constant it will become exponential. Weibull is the only parametric model which has both proportional hazards and an accelerated failure-time representation (Klein and Moeschberger, 2003). In addition, acceptance of Weibull model can be checked via graphical assessment (Kleinbaum and Klein, 2011).

Aim of the present study was to apply Weibull model to investigate the possible prognostic factors that may have an effect on survival of patients with breast cancer. Therefore, the only outcome considered here is survival.

Materials and Methods

The present study incorporates data from patients who were visited and treated at Cancer Research Center in Shahid Beheshti University of Medical Sciences, Tehran, Iran during 1992 to 2012; the patients were followed up until October 2014. The patients or patients' family members were contacted via phone calls to confirm their health status (i.e. whether they are still alive or not) and to fill any gaps in their medical records. We had to exclude some patients because of some reasons, first their medical records had incomplete information, second they were related to male and third, their cause of death was not breast cancer. Eventually a total number of 438 women with breast cancer were included in this analysis.

Univariate analysis was conducted using log-rank test in order to find the factors that had a significant effect on survival time. The log-rank test is the most commonly-used statistical test in order to compare the survival distributions of two or more groups (ZAMAN and PFEIFFER, 2012). The null hypothesis in the log-rank test is that all survival curves are the same (Kleinbaum and Klein, 2011), more specifically for each time it calculates the observed number that faced to the event of interest in each group and the number expected if there were in reality no difference between the groups (Bland

and Altman, 2004).

Weibull distribution in parametric survival model was used for multivariate analysis. In this step we used forward approach. So we added the variable which had the lowest value of Akaike Information Criterion (AIC). The AIC, which was developed by Hirotugu Akaike in 1974, is a measure of the goodness of fit of an estimated statistical model (Akaike, 1974). When none of the remaining

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

Factors	Categories	No. of patients	(%)	P-value
Age at diagnosis				
	Under 40 yr.	88	20.1	0.005 *
	40 to 70 yr.	336	76.7	
	Over 70 yr.	14	3.2	
Education				
	Illiterate	37	8.45	0.004 *
	Elementary school	51	11.64	
	Secondary school	53	12.1	
	Diploma	161	36.76	
	Academic	136	31.05	
Marital				
	Single	24	5.48	0.2
	Married	414	94.52	
Abortion				
	Yes	153	34.93	0.72
	No	285	65.07	
Breastfeeding				
	Yes	381	86.99	0.194
	No	57	13.01	
Family history				
	Yes	124	28.31	0.409
	No	314	71.69	
Type of surgery				
	BCS	274	62.56	<0.001 *
	MRM	164	37.44	
Lymph node status				
	N0	198	45.21	<0.001 *
	N1	121	27.63	
	N2	39	8.9	
	N3	80	18.26	
Tumor size				
	T1	111	25.34	<0.001 *
	T2	242	55.25	
	T3	85	19.41	
Stage				
	I	96	21.92	<0.001 *
	II	193	44.06	
	III	135	30.82	
	IV	14	3.2	
Histologic grade				
	Well differentiated	47	10.73	<0.001 *
	Moderately differentiated	235	53.65	
	Poorly differentiated	156	35.62	
Estrogen Receptor				
	Positive	288	65.75	0.008 *
	Negative	150	34.25	
Progesterone Receptor				
	Positive	172	39.27	0.007 *
	Negative	266	60.73	
Lymphovascular Invasion				
	Positive	312	71.23	<0.001 *
	Negative	126	28.77	

* Significant at the 5% level

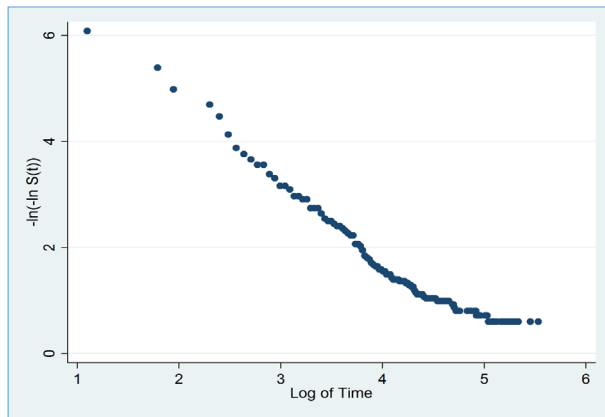


Figure 1. Negative Log of Negative Log Survivor Function Estimates

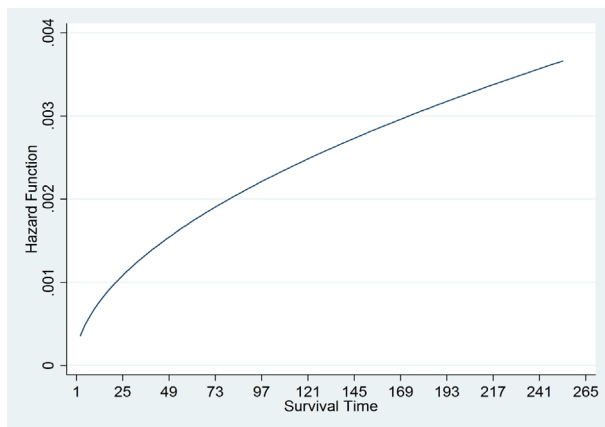


Figure 1. Hazard function against survival time

Table 2. Multivariate Analysis of Weibull Parametric Model with Prognostic Factors

Factors	Categories	HR (CI: 95%)	P-value
Lymph node status			
	N0	0.18 (0.06 - 0.52)	0.002 *
	N1	0.44 (0.21 - 0.91)	0.027 *
	N2	0.64 (0.33 - 1.25)	0.192
	N3 **	1	
Stage			
	I	0.42 (0.09 - 1.94)	0.268
	II	0.33 (0.12 - 0.90)	0.030 *
	III	0.45 (0.21 - 0.93)	0.030 *
	IV **	1	
Histologic grade			
	Well differentiated	0.14 (0.02 - 1.11)	0.063
	Moderately differentiated	0.51 (0.31 - 0.86)	0.011 *
	Poorly differentiated **	1	
Lymphovascular invasion			
	Positive	2.13 (1.26 - 3.58)	0.005 *
	Negative **	1	

* Significant at the 5% level; ** Stands for a control group; HR, Hazard Ratio; CI, Confidence Interval

variables had significant effect, adding variables was stopped. Univariate and multivariate analyses were carried out based on the clinical, pathological and biological characteristics of patients.

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). The proposal of the present study was approved by the ethical committee of Shahid Beheshti University of Medical Sciences. The privacy of participants upheld by the research organizer and personal information were not included in this study. The data were analyzed by STATA version 11. Quantitative results were expressed as mean \pm standard deviation. A P-value lower than 0.05 was defined as significant.

Results

A total number of 438 women with breast cancer were included in this analysis. The mean of age at the time of diagnosis was 48.37 ± 10.92 years and patients' age ranged from 22 to 84 years. Most of the patients were married women (94.52 %) and they didn't have history of breast cancer in their family (71.69 %). The longest duration of follow-up was 253 months. A total of 75 (17.12%) deaths due to breast cancer were observed (non-censored observations) up to October 2014. Using the life-table method, the one-year overall survival (OS) rate of this study population was 98% (95% CI: 97%-99%).

The graph of the $-\ln(-\ln(S(t)))$ against log of failure time followed a linear trend which indicated that Weibull model is appropriate for this data (Figure 1). The prognostic variables included in the model are age at diagnosis, level of education, 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.

In univariate analysis age at diagnosis ($p=0.005$), level of education ($p=0.004$), type of surgery ($p<0.001$), lymph node status ($p<0.001$), tumor size ($p<0.001$), stage ($p<0.001$), histologic grade ($p<0.001$), estrogen receptor ($p=0.008$), progesterone receptor ($p=0.007$), and lymphovascular invasion ($p<0.001$) had a statistically significant effect on survival time (Table 1).

Multivariate analysis with forward approach was used for total clinical, pathological, and biological variables, regardless of whether the variable was found to be significant or not based on the results obtained from the univariate analysis. According to the results of multivariate analysis which are presented in Table 2, lymph node status, stage, histologic grade, and lymphovascular invasion were statistically significant. Indeed, patients with lymphovascular invasion positive were 2.13 times more at the risk of death due to breast cancer. Moreover, the slope of Weibull model is 1.54 (95% CI: 1.30-1.83) which indicates the increasing hazard rate (Figure 2).

Discussion

To estimate the survival using multi-explanatory variables, researchers are largely interested in Cox proportional hazard model more than parametric models. In a review of cancer journals for survival analyses (Altman et al., 1995) it was found that only 5% of all studies had checked the underlying Cox's assumptions.

Furthermore, in many situations these assumptions do not hold (Cox and Oakes, 1984) and will result in unreliable outcomes. Therefore, it is recommended to use parametric survival models. These models provide an interpretation based on a particular distribution of the time to event irrespective of proportional hazard assumptions.

Several studies have been conducted on parametric survival models (Sirott et al., 1993; Wilson, 1994; Nardi and Schemper, 2003; Akram et al., 2007; Pourhoseingholi et al., 2007; Moghimi-Dehkordi et al., 2008; HAYAT et al., 2010; Pourhoseingholi et al., 2011; Ravangard et al., 2011; Zhu et al., 2011; Alizadeh et al., 2013; Binti Zulkifli et al., 2013; Hashemian et al., 2013; Zare et al., 2013; Vallinayagam et al., 2014). Also, Carroll's study indicated that in the analysis of survival data Weibull model can provide a useful, parametric alternative to Cox's regression modeling (Carroll, 2003).

Aim of this study was to assess the relationship between survival of patients with breast cancer and some clinical, pathological, and biological variables. Drawing on our findings, age at diagnosis, level of education, type of surgery, lymph node status, tumor size, stage, histologic grade, lymphovascular invasion, estrogen receptor, and progesterone receptor were significant. Also, based on the result of multivariate analysis lymph node status, stage, histologic grade, and lymphovascular invasion had a statistically significant effect on survival.

In our data, the mean of age at the time of diagnosis is about 48 years which is exactly similar to Eivazi-Ziaei's study (Eivazi-Ziaei et al., 2013), while this mean is not in line with other Iranian published reports (Vahdaninia and Montazeri, 2004; Akbari et al., 2011; Akbari et al., 2012; Faradmal et al., 2012; Movahedi et al., 2012; Nafissi et al., 2012). According to the results of this study, the one-year overall survival rate was found to be 98%, while Movahedi's study estimated this rate fewer than the present study (Movahedi et al., 2012). Our study found no evidence for a relationship between survival of patients and family history of breast cancer, this is in accordance with the results of several other studies (Harris et al., 2000; Eccles et al., 2001; Gonzalez Angulo et al., 2005; Rezaianzadeh et al., 2009). Our result is consistent with previous study except for estrogen receptor and progesterone receptor since in our study they were significant in univariate analysis (Zulkifli et al., 2013). Based on our findings there is an association between histological grade and survival which is not in common with other studies (Kuru et al., 2003; Arpino et al., 2004; Rezaianzadeh et al., 2009).

There are some limitations to this study. Data for clinical, pathological, and biological variables were missing and some were recorded in a wrong way that made them unused. Also, due to some changes in telephone registries especially for totally changes in some provinces or changing their addresses (for those who did not report their mobile phone), we couldn't communicate with them and their records were removed. Furthermore, we didn't have any information about the socioeconomic factors of patients but interest may be done either by investigating the effect of these factors.

In addition, our study was carried out based on data

collected from a single institution and therefore the findings could not be generalized. Overall, about 80% of our data were censored, therefore, for patients with breast cancer who did not experience death during the follow-up, cure rate models can be a good choice. Actually in the case of long-term survival, cure rate models are better than Cox model (Rahimzadeh et al., 2014). So we suggest to analyze these survival data using cure models.

Acknowledgements

It's notable that the paper is extracted from MSc thesis of Sahar Saeedi Moghaddam in Biostatistics and has been financial supported by Cancer Research Center in Shahid Beheshti University of Medical Sciences. In addition, it deserves that we thank employees of Cancer Research Center and appreciate the collaboration of Dr. Maryam Khayamzadeh for her kindly contribution.

References

- Akaike H (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, **19**, 716-23.
- Akbari A, Razzaghi Z, Homae F, et al (2011). Parity and breastfeeding are preventive measures against breast cancer in Iranian women. *Breast Cancer*, **18**, 51-5.
- Akbari ME, Khayamzadeh M, Khoshnevis S, et al (2012). Five and ten years survival in breast cancer patients mastectomies vs. breast conserving surgeries personal experience. *Iranian Journal of Cancer Prevention*, **1**, 53-6.
- Akram M, Ullah MA, Taj R (2007). Survival analysis of cancer patients using parametric and non-parametric approaches. *Pakistan Veterinary Journal*, **27**, 194.
- Alizadeh A, Mohammadpour RA, Barzegar MR, et al (2013). Comparing cox model and parametric models in estimating the survival rate of patients with prostate cancer on radiation therapy. *Journal of Mazandaran University of Medical Sciences (JMUMS)*, **23**.
- Altman D, De Stavola B, Love S, et al (1995). Review of survival analyses published in cancer journals. *British Journal of Cancer*, **72**, 511.
- American Cancer Society 2014. *Cancer Facts & Figures 2014*. American Cancer Society (ACS) Atlanta, GA: American Cancer Society, 2014. 72 p., pdf.
- Arpino G, Bardou VJ, Clark GM, et al (2004). Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast Cancer Res*, **6**, 149-56.
- Bland JM, Altman DG (2004). The logrank test. *Bmj*, **328**, 1073.
- Bray F, Ren JS, Masuyer E, et al (2013). Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International Journal of Cancer*, **132**, 1133-45.
- Carroll KJ (2003). On the use and utility of the Weibull model in the analysis of survival data. *Controlled clinical trials*, **24**, 682-701.
- Cox DR, Oakes D 1984. *Analysis of survival data*, CRC Press.
- Eccles D, Simmonds P, Goddard J, et al (2001). Familial breast cancer: an investigation into the outcome of treatment for early stage disease. *Familial cancer*, **1**, 65-72.
- Eivazi-Ziaei J, Sanaat Z, Asvadi I, et al (2013). Survival analysis of breast cancer patients in northwest Iran. *Asian Pac J Cancer Prev*, **14**, 39-42.
- Faradmal J, Talebi A, Rezaianzadeh A, et al (2012). Survival analysis of breast cancer patients using cox and frailty models. *Journal of Research in Health Sciences*, **12**, 127-30.

- Ferlay J, Soerjomataram I, Ervik M, et al (2014). GLOBOCAN 2012 v1. 0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013. Visit: <http://globocan.iarc.fr>.
- Gonzalez Angulo AM, Broglio K, Kau SW, et al (2005). Women age \leq 35 years with primary breast carcinoma. *Cancer*, **103**, 2466-72.
- Harris EE, Schultz DJ, Peters CA, et al (2000). Relationship of family history and outcome after breast conservation therapy in women with ductal carcinoma in situ of the breast. *International Journal of Radiation Oncology* Biology* Physics*, **48**, 933-41.
- Hashemian AH, Beiranvand B, Rezaei M, et al (2013). A comparison between cox regression and parametric methods in analyzing kidney transplant survival. *World Applied Sciences Journal*, **26**, 502-7.
- HAYAT EA, Suner A, Burak U, et al (2010). Comparison of five survival models: Breast cancer registry data from ege university cancer research center. *Turkiye Klinikleri Journal of Medical Sciences*, **30**, 1665-74.
- IHME 2013. The Global Burden of Disease: Generating Evidence, Guiding Policy, University of Washington Seattle, WA, USA.
- Klein JP, Moeschberger ML 2003. Survival analysis: techniques for censored and truncated data, Springer Science & Business Media.
- Kleinbaum DG, Klein M 2011. Survival Analysis: A Self-Learning Text, Third Edition, Springer.
- Kuru B, Camlibel M, Ali Gulcelik M, et al (2003). Prognostic factors affecting survival and disease free survival in lymph node negative breast carcinomas. *Journal of surgical oncology*, **83**, 167-72.
- May AM, Struijk EA, Franssen HP, et al (2015). The impact of a healthy lifestyle on Disability-Adjusted Life Years: a prospective cohort study. *BMC medicine*, **13**, 39.
- Moghimi-Dehkordi B, Safaee A, Pourhoseingholi MA, et al (2008). Statistical comparison of survival models for analysis of cancer data. *Asian Pacific J Cancer Prev*, **9**.
- Movahedi M, Haghghat S, Khayamzadeh M, et al (2012). Survival rate of breast cancer based on geographical variation in Iran, a national study. *Iranian Red Crescent Medical Journal*, **14**, 798.
- Murray CJ, Vos T, Lozano R, et al (2013). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet*, **380**, 2197-223.
- Nafissi N, Saghafinia M, Motamedi M, et al (2012). A survey of breast cancer knowledge and attitude in Iranian women. *Journal of cancer research and therapeutics*, **8**, 46.
- Nardi A, Schemper M (2003). Comparing Cox and parametric models in clinical studies. *Statistics in Medicine*, **22**, 3597-610.
- Nelson W (2000). Weibull prediction of a future number of failures. *Quality and Reliability Engineering International*, **16**, 23-6.
- Pourhoseingholi M, Pourhoseingholi A, Vahedi M, et al (2011). Alternative for the cox regression model: using parametric models to analyze the survival of cancer patients. *Iranian Journal of Cancer Prevention*, **4**, 1-9.
- Pourhoseingholi MA, Hajizadeh E, Moghimi Dehkordi B, et al (2007). Comparing Cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pac J Cancer Prev*, **8**, 412.
- 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.
- Ravangard R, Arab M, Rashidian A, et al (2011). Comparison of the results of Cox proportional hazards model and parametric models in the study of length of stay in a tertiary teaching hospital in Tehran, Iran. *Acta Medica Iranica*, **49**, 650-8.
- 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.
- 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.
- Sirott MN, Bajorin DF, Wong GY, et al (1993). Prognostic factors in patients with metastatic malignant melanoma: a multivariate analysis. *Cancer*, **72**, 3091-8.
- 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.
- Vahdaninia M, Montazeri A (2004). Breast cancer in Iran: a survival analysis. *Asian Pac J Cancer Prev*, **5**, 223-5.
- Vallinayagam V, Prathap S, Venkatesan P (2014). Parametric regression models in the analysis of breast cancer survival data. *International Journal of Science and Technology*, **3**, 163-7.
- Weibull W (1951). A Statistical Distribution Function of Wide Applicability. *Journal of applied mechanics*.
- Wilson DL (1994). The analysis of survival (mortality) data: fitting Gompertz, Weibull, and logistic functions. *Mechanisms of ageing and development*, **74**, 15-33.
- Zaman Q, Pfeiffer KP (2012). Does Log-rank test give satisfactory results? *Journal of Applied Quantitative Methods*, **3**.
- Zare A, Mahmoodi M, Mohammad K, et al (2013). Comparison between parametric and semi-parametric cox models in modeling transition rates of a multi-state model: application in patients with gastric cancer undergoing surgery at the Iran cancer institute. *Asian Pac J Cancer Prev*, **14**, 6751-5.
- Zhu HP, Xia X, Chuan HY, et al (2011). Application of Weibull model for survival of patients with gastric cancer. *BMC gastroenterology*, **11**, 1.