Determining Prognostic Factors of Disease-Free Survival in Breast Cancer Using Censored Quantile Regression

Akram Yazdani[1] (ID) and Shahpar Haghighat[2] (ID)

[1]Department of Biostatistics and Epidemiology, School of Public Health, Kashan University of Medical Sciences, Kashan, Iran. [2]Breast Cancer Research Center, Motamed Cancer Institute, ACECR, Tehran, Iran.

ABSTRACT

BACKGROUND: The analysis of disease-free survival and related factors leads to a better understanding of the patient's condition and recurrence-related characteristics and provides a basis for more appropriate treatment guidance. In this study, we aimed to investigate the role of prognostic factors on disease-free survival in breast cancer with a quantile regression model.

METHODS: This retrospective study was conducted by reviewing data obtained from 2056 breast cancer patients. Age at diagnosis and education status, tumor size, lymph node ratio, tumor grade, estrogen receptor and progesterone receptor, type of surgery, use of radiotherapy, chemotherapy, and hormone therapy were the prognosis factors considered in this study. A quantile regression model was used to investigate prognostic factors of disease-free survival in breast cancer.

RESULTS: Disease recurrence was verified in 251 (13.9%) women, and 39 (0.02%) women died before experience recurrence. The 10th percentile of disease-free survival for patients with the hormone therapy was 23.85 months greater than patients who did not receive this treatment (P value < .001). In the examination of the tumor size, the 10th and 20th percentiles of disease-free survival for patients with tumor size > 5 cm were 31.06 and 27 months less than patients with the tumor size < 2 cm, respectively (P value = .006 and .021, respectively). Compared with grade 1 tumors, the 10th and 20th percentiles of disease-free survival for patients with grade 3 tumors decreased 30.11 and 38.32 months, respectively (P value < .001 and .038, respectively). The 10th and 20th percentiles of disease-free survival decreased 28.16 and 45.32 months with a 1 unit increase in lymph node ratio, respectively (P value = .032 and .032, respectively).

CONCLUSIONS: Among the prognostic factors, tumor size, grade, and lymph node ratio showed a close relationship with disease-free survival in breast cancer. The findings indicated that developing public screening and educational programs through the health care system with more emphasis on low-educated women is needed among Iranian women.

KEYWORDS: Breast cancer, quantile regression, Laplace regression, disease-free survival

Introduction

Breast cancer is the most common cancer among women worldwide; in 2020, there were 2.3 million women diagnosed with breast cancer and 685 000 deaths globally.[1] As of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past 5 years, making it the most common cancer in the world. The incidence and mortality of breast cancer in different countries are determined by the country's economic development, environmental factors, and the ethnic origin of the population. When comparing data from developed and developing countries, one can see the difference between breast cancer incidence and mortality.[2] Breast cancer has a high incidence and low mortality in developed countries, and a low incidence and high mortality in developing countries.[3] According to the cancer registry in Iran, breast cancer is one of the most common types of cancer in women, which is most prevalent in the fourth and fifth decade of life. That is at least 10 years lower than the world.[4]

A major factor in reducing the survival of breast cancer patients is recurrence.[5] Cancer may come back in the same place as original cancer (local recurrence), or it may spread to other areas of your body (distant metastases).[6] In general, one-third of recurrence cases are local recurrence, and the rest are distant metastases. The average survival for patients with breast cancer metastases is 18 to 24 months, and 10% to 15% of patients have an early recurrence, that is, a recurrence within 2 years of treatment. In addition, 50% of patients have late relapse after 5 years.[7] The time between initial diagnosis and first any type of recurrence (invasive ipsilateral breast tumor recurrence, local invasive recurrence, regional invasive recurrence, invasive contra lateral breast cancer, distant recurrence, second primary invasive cancer [non-breast cancer]) or death from any cause is defined as disease-free survival (DFS).[8] Analysis of DFS and related factors can help better understand the patient's condition and recurrence-related characteristics, and provide a basis for more appropriate treatment guidance. However, it should be noted that there are very few national literature studies conducted to investigate disease recurrence and its related factors in the Iranian context.

In recent decades, many prognostic factors have been discovered that can be used to predict the risk of breast cancer recurrence and the possibility of death after the initial

surgery. Traditional prognostic factors, those that predict the risk of breast cancer recurrence or death, include the number of positive axillary lymph nodes, tumor size, tumor grade, lymphatic and vascular invasion, the estrogen receptor (ER), and the progesterone receptor (PR).[9,10]

Many studies have used parametric or nonparametric models to study prognostic factors for DFS. The Cox proportional hazards model is a widely used semiparametric survival model that requires proportional hazard assumption. The Cox method models the hazard ratio rather than the survival time directly, and the complexity of the hazard ratio estimate interpretation is considered as a major problem.[4,11] The censored quantile regression (CQR) model seeks to model the survival time separately for each given quantile directly, and investigates the covariate effects in different quantiles. In this model, the covariate effects can be changed for patients with different risks and is a flexible model for controlling the heterogeneity of covariate effects.[11] The CQR model does not need to assume homogeneous covariate effects and can directly explain the impact of the covariates on the time of the event.[12,13] The model was introduced by Koenker and Bassett[14] and can estimate the effect of explanatory variables on response variable for a set of properly selected quantiles; thus, it provides a complete picture and may show different covariate effects at different follow-up times.[13]

In this study, we analyzed DFS according to traditional prognostic factors, and we tried to identify indicators that could predict recurrence in patients with the CQR model.[15,16]

Methods

Patients

This retrospective study was conducted by reviewing data obtained from 2056 breast cancer patients (based on the diagnosis of breast cancer pathology) diagnosed at the Motamed Cancer Institute in Tehran, Iran. The participants of this study were referred to Breast Clinic between 1996 and 2011 and were followed up to 2016. If the previous 6 months status of patient had not been recorded, their recurrence status was actively completed by telephone interview and inviting them for attendance in clinic. All patients were followed up according to a defined protocol in different intervals from 3, 6, to 12 months (in 5 years) by different methods consisting of physical examination, breast ultrasonography, and mammography, and chest x-ray, abdominal and pelvic ultrasonography, and if needed full body scan. The exclusion criteria were patients with incomplete records, no access to patients, and lack of consent to participate in the study. According to the established selection criteria, the population of this study only included patients with stage I, II, or III who had undergone any type of primary breast conserving surgery (BCS) at the time of diagnosis (non-metastatic), and patients with distant metastases within 30 days after surgery were excluded. The median follow-up duration calculated by reverse Kaplan-Meier estimator was 57.46 months (95% confidence interval [CI], 53.73-61.19 months).

Disease-free survival was defined as the duration (months) from the initial diagnosis of breast cancer to first any type of recurrence (invasive ipsilateral breast tumor recurrence, local invasive recurrence, regional invasive recurrence, invasive contra lateral breast cancer, distant recurrence, second primary invasive cancer [non-breast cancer]) or death from any cause. Patients that remained without signs of recurrence by the end date of monitoring and those with loss of monitoring were regarded as censored observations.

The independent variables were distributed into the following blocks: (1) socio-demographic: age at diagnosis (year) and education status (categorized into 4 groups: illiterate, primary school, high school, and university); (2) tumor related: tumor size (categorized as ≤2 cm and 2-5 cm and >5 cm), lymph node ratio (the lymph node ratio is defined as the number of positive lymph nodes divided by the total number of lymph nodes excised), tumor grade (grades 1-3), ER and PR (a positive ER or PR test is defined as positive staining of greater than or equal to 1% of tumor cells; a negative test is defined as staining of less than 1% of tumor cells); and (3) treatment related: type of surgery (modified radical mastectomy [MRM] or BCS); use of radiotherapy, chemotherapy, and hormone therapy.

Statistical analysis

Categorical data were described in frequencies and percentages whereas continuous data were described by mean ± standard deviation. Survival estimates were computed using the Kaplan-Meier method. All analyses were performed using Stata, version 13.0.

For the assessment of prognostic variables, we used the Laplace quantile regression model and the Cox proportional hazards model. Quantile regression is a statistical technique intended to inference about conditional quantile functions. This method offers a mechanism for estimating models for the conditional median function, and the full range of other conditional quantile functions. The quantiles were chosen based on the proportion of censored patients.[15]

The selection of variables was carried out using the significance obtained from the univariate Laplace model, considering the significance level of $P \leqslant .2$, and relevance in the literature. The variables that met the criteria previously described were included in the multivariate Laplace quantile regression analysis and selected through the process of "backward elimination."

Laplace quantile regression model. Bottai and Zhang[17] proposed the Laplace regression model as a method to model the conditional quantiles of the response variable with random censoring in which the error term is assumed to follow the asymmetric Laplace distribution.

Let $T_i, i = 1, 2, \ldots, n,$ be the independent response variable and let x_i, k-dimensional vectors of observed covariates.[13]

Suppose there is a fixed *k*-dimensional parameter vector β(*p*) such that

$$T_i = x_i' \beta(p) + \varepsilon_i, \qquad (1)$$

where $p \in (0,1)$ is a fixed and given probability. We assume that ε_i are independent and identically distributed residuals and assume that under the condition of x_i, T_i follows the asymmetric Laplace distribution with probability density function,

$$f(t_i) = \exp\left[\left\{I(t_i \le \mu_i) - p\right\} \frac{t_i - \mu_i}{\sigma(p)}\right] \frac{p(1-p)}{\sigma(p)}, \qquad (2)$$

where $(p) \in \mathbb{R}^k$, $\sigma(p) \in (0,\infty)$, $\mu_i = x_i'\beta(p)$, and $P(T_i \le \mu_i \mid x_i) = p$.

Results

The demographic and clinicopathologic characteristics of the 2056 patients included in our study are summarized in Table 1. The mean (SD) age of the patients at the time of diagnosis was 46.97 (10.92) years, with most women (33.6%) in an age group with the highest incidence of the disease, that is, from 50 to 69 years, and 29.7% aged less than 40 years. The education level of the study population was low, with 12.4% of the women being illiterate and 35.3% of the women having primary school education only. The mean (SD) lymph node ratio was 0.26 (0.37); 76.9% of patients had tumor size more than 2 cm and a great number of patients had positive hormonal receptors.

Disease recurrence was verified in 251 (13.9%) women, and 39 (0.02%) women died before experience recurrence. The 20% DFS time for the population studied was 90.69 months (95% CI, 74.1-125.25). About the socio-demographic variables, the rates of recurrence or death presented the lowest percentages for women with a high university level of education (15.4%). The mean (SD) age of the patients with recurrent or death was 47.92 (11.95) years. Meanwhile, about tumor-related variables, the lowest rates of recurrence or death were those of women who had tumors ⩽2 cm (12.9%), grade 1 tumor (10.7%). Concerning treatment-related variables, the lowest rates of recurrence or death were identified in women who were undergoing conservative surgery 25.6%, those who did not receive chemotherapy 7.7%, and hormone therapy 25.7%. The mean (SD) lymph node ratio in the recurrent or death group was 0.39 (0.57) months.

Based on the Kaplan-Meier curve (Figure 1) and the censoring rate (84%), estimates of the conditional percentiles of DFS at any level higher than 25 would be infinite. Therefore, we focused on the percentile levels of 10 and 20.

The multivariate analysis of the Laplace regression model is shown in Table 2. The 10th percentile of DFS for patients with hormone therapy was 23.85 months greater than patients

**Table 1.** Profile of patient demographics and clinical characteristic.

PROGNOSTIC FACTORS	N (%) MEAN (SD)	RECURRENT OR DEATH (%)	P VALUE
Surgical procedure			
MRM	1149 (58.8)	198 (74.4)	<.001
BCS	805 (41.2)	68 (25.6)	
Radiotherapy			
Yes	1286 (83.6)	203 (81.9)	.311
No	253 (16.4)	45 (18.1)	
Chemotherapy			
Yes	1327 (88.6)	217 (92.3)	.111
No	171 (11.4)	18 (7.7)	
Hormone therapy			
Yes	1148 (84.5)	150 (74.3)	<.001
No	210 (15.5)	52 (25.7)	
Estrogen receptor			
Negative	469 (30.7)	90 (41.3)	<.001
Positive	1057 (69.3)	128 (58.7)	
Progesterone receptor			
Negative	527 (34.9)	103 (47.7)	<.001
Positive	984 (65.1)	113 (52.3)	
Tumor size			
<2 cm	336 (23)	25 (12.9)	<.001
2-5 cm	892 (61.1)	127 (65.5)	
>5 cm	231 (15.8)	42 (21.6)	
Tumor grade			
1	193 (13.5)	21 (10.7)	<.001
2	865 (60.5)	101 (51.3)	
3	372 (26.0)	75 (38.1)	
Education status			
Illiterate	244 (12.4)	49 (17.5)	.001
Primary school	692 (35.3)	112 (40.0)	
High school	622 (31.7)	76 (27.1)	
University	404 (20.6)	43 (15.4)	
Lymph node ratio	0.26 (0.37)	0.39 (0.57)	<.001
Age	46.97 (10.92)	47.92 (11.95)	.242

Abbreviations: BCS, breast conserving surgery; MRM, modified radical mastectomy.

Table 2. The results of multivariate Laplace regression and Cox model assessing the effect of prognostic factors on the DFS.

| PROGNOSTIC FACTORS	Q10		Q20		COX MODEL	
	COEF. (SE)	*P* VALUE	COEF. (SE)	*P* VALUE	HR (SE)	*P* VALUE
Hormone therapy (yes)	23.85 (6.51)	<.001	21.06 (10.37)	.252	0.67 (0.15)	.082
Tumor size						
<2 cm						
2-5 cm	−12.41 (10.79)	.250	−12.99 (17.27)	.452	1.20 (0.33)	.498
>5 cm	−31.06 (11.25)	.006	−33.27 (15.95)	.021	1.88 (0.61)	.049
Tumor grade						
1						
2	−15.34 (11.02)	.164	−7.11 (16.17)	.660	1.05 (0.34)	.871
3	−30.11 (9.21)	.001	−38.32 (18.44)	.038	1.85 (0.62)	.042
Lymph node ratio	−28.16 (9.92)	.032	−45.32 (17.01)	.032	1.38 (0.20)	.045
Education status						
Illiterate						
Primary school	10.66 (10.81)	.094	23.62 (19.49)	.226	0.76 (0.22)	.373
High school	20.58 (11.57)	.025	41.89 (20.46)	.041	0.64(0.20)	.152
University	10.00 (11.59)	.194	18.08 (20.00)	.366	0.76 (0.24)	.372

Abbreviations: Coef., estimated parameter; DFS, disease-free survival; HR, hazard ratio.
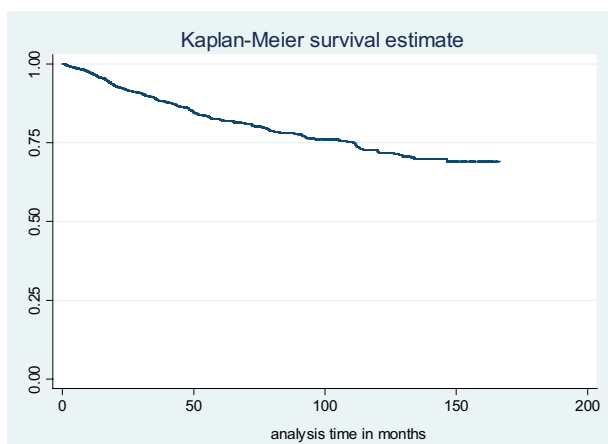
Figure 1. Kaplan-Meier curves in breast cancer patients.

who did not receive this treatment (*P* value < .001). On the contrary, the effect of hormone therapy was not significant at the 20th percentile. This result indicates that hormone therapy may have heterogeneous effects on DFS. In the examination of the tumor size, the 10th and 20th percentiles of DFS for patients with tumor size > 5 cm were 31.06 and 27 months less than patients with the tumor size < 2 cm, respectively (*P* value = .006 and .021, respectively).

Compared with grade 1 tumor, the 10th and 20th percentiles of DFS for patients with grade 3 tumor decreased 30.11 and 38.32 months, respectively (*P* value < .001 and .038, respectively). The 10th and 20th percentiles of DFS decreased 28.16 and 45.32 months with a 1 unit increase in lymph node ratio, respectively (*P* value = .032 and .032, respectively). The different effects of the lymph node ratio at the 10th and 20th percentiles of DFS indicate that the lymph node ratio may have heterogeneous effects, which could not be detected by the Cox model. In the examination of the education status, 10th percentiles of DFS increased 10.66 and 20.58 months for the primary and high level compared with the illiterate level, respectively (*P* value = .094 and .025, respectively). Compared with the CQR model, the result from the Cox model showed tumor size > 5 cm, the lymph node ratio, and grade 3 tumor are prognostic factors of DFS. As we show in Supplementary file, based on the Schoenfeld residuals, education status does not satisfy the non-proportionality. The asymmetric Laplace distribution assumption was confirmed.

The plots of CQR coefficients estimated and their 95% CIs for $p \in$ (0.05, 0.10, 0.15, 0.20, and 0.25) are displayed in Figure 2. They further compare the coefficients from CQR with a local quantile measure of the effects of covariates in a Cox model on conditional quantiles was proposed by Portnoy.[12]
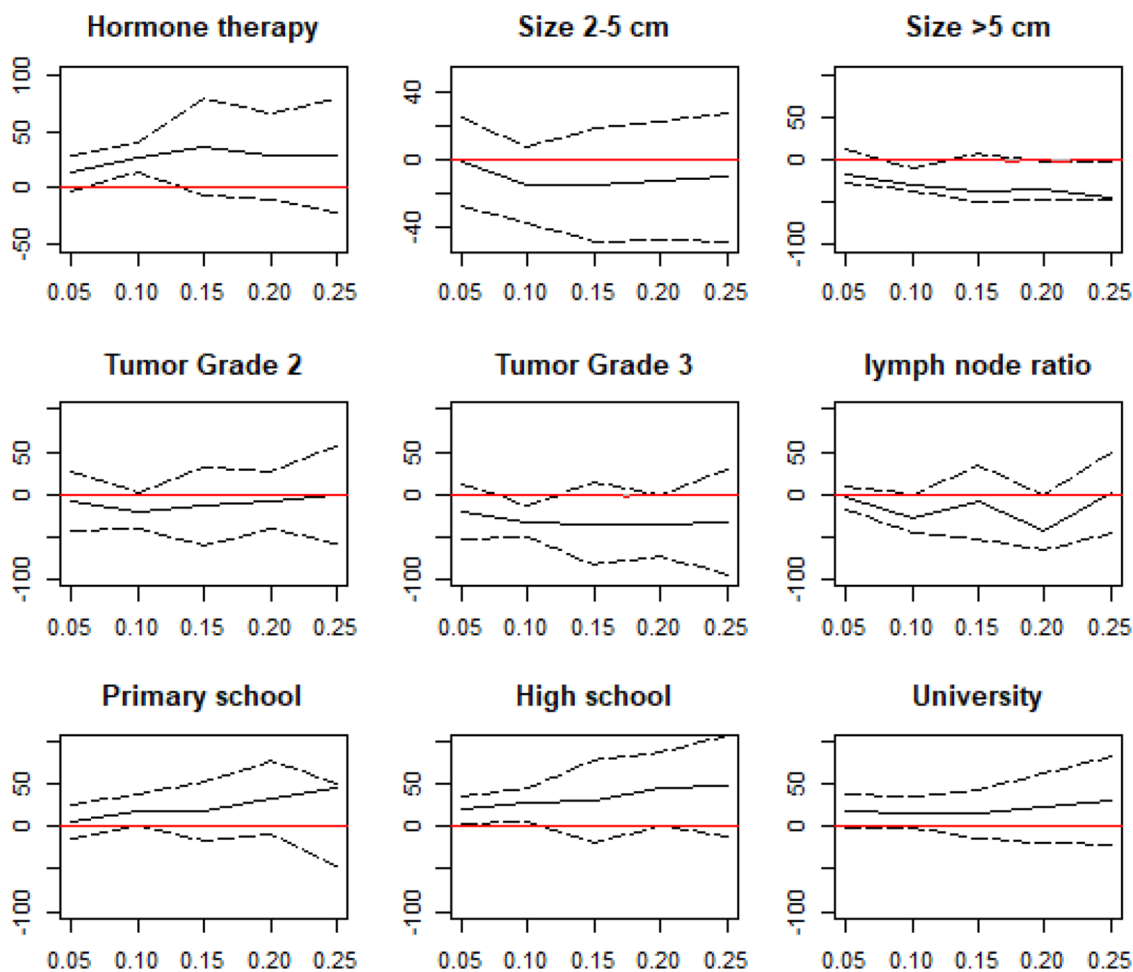
Figure 2. Results of censored quantile regression with point estimates (solid line) and 95% CI (longdash lines); the red lines are the estimated local quantile measures for the Cox proportional hazards model. CI indicates confidence interval.

Discussion

Patients with breast cancer have a risk of recurrence even years after receiving treatment. The highest risk is in 2 or 3 years after the primary tumor was discovered. Distant metastases are the most common form of recurrence in breast cancer patients and are the leading cause of death. The most common prognostic factors that cause recurrence and distant metastasis after treatment include pathologic of breast cancer, tumor grade, tumor size, involvement of lymph node, and hormone receptors' status of ER and PR.

This study is a retrospective study that analyzed 2056 patients suffering from breast cancer in Iran. The initial investigation of prognosis factors of DFS using Laplace regression showed similar results to previous studies. In multivariate analysis, hormone therapy, tumor size, tumor grade, lymph node ratio, and education were identified as prognostic factors. The 10th percentile of DFS increases for a case with hormone therapy and high-level school and decreases for a case with tumor size > 5, tumor grade 3, and high lymph node ratio. In the 20th percentile, tumor size > 5, tumor grade 3, and high lymph node ratio decrease DFS time, and high-level school increases DFS time. The results showed that effect of some

factors were significant at the special percentile but were not significant at the other percentile. For example, the 10th percentile of DFS for patients with hormone therapy was 23.85 months greater than patients who did not receive this treatment. On the contrary, the effect of hormone therapy was not significant for the 20th percentile quantile. It shows that factors may have heterogeneous effects on DFS. In comparison with the CQR model, some of the variables have a significant effect with Cox model. The standard Cox proportional hazard models do not have enough power to analyze data with survival trends like those shown in Figure 1 for several reasons. First, the assumption of proportional hazards can fail when survival curves have plateaus at their tails. Second, survival plots with long plateaus may indicate heterogeneity within a patient population. The Cox models are restricted because the proportional hazards assumption does not allow the sign of a covariate to vary between patients with shorter survival time and patients with longer survival time. Therefore, it is necessary to use methods that model the heterogeneous effects of predictors.

Zhao et al[2] studied independent prognostic factors for breast cancer patients and discuss the role of postmastectomy radiotherapy in these patients.[2] They showed tumor size and

ER/PR status were independent predictors of risk of recurrence. Karihtala et al[18] investigated the role of early breast cancer prognostic factors in metastatic disease. The results confirm ER status as a primary prognostic factor in metastatic breast cancer. Furthermore, it also suggests that the presence of initial lymph node metastases could serve as a prognostic factor in recurrent breast cancer.[18] Akbari et al[5] studied prognostic factors effective in recurrence and death in Iranian breast cancer patients with a retrospective study conducted by reviewing data acquired from 1604 female breast cancer patients who were admitted to Cancer Research Center at Shahid Beheshti University of Medical Sciences. Based on their results biologic marker, ER and PR status had the most influence in early recurrence, unlike late recurrence, where the stage of disease had a more important role. However, lymph vascular invasion has been an effective factor either in early or late recurrence.[5] Potential prognostic factors that affect the DFS were investigated by Koca et al.[19] In univariate analysis, the number of pathologic axillary lymph nodes, triple-negative status, and tumor size were found to be the prognostic factors affecting the DFS and, in the multivariate analysis, triple-negative status was the only independent prognostic factor that affected the DFS adversely.[19] Patients diagnosed with the disease between 2003 and 2005 and identified through the institution's cancer hospital records were analyzed by Diniz et al[20] to analyze the DFS at 5 years and prognostic factors in women with non-metastatic invasive breast cancer. This study showed the main variables associated with DFS were lymph node involvement, use of hormone therapy, and education level.[20]

Some limitations should be taken into consideration in interpreting the results of this study. The high proportion of patients censored (86.1%) implies limit percentiles to estimate in the Laplace regression model. In addition, breast cancer in the study population was diagnosed when patients showed up at hospitals for primary treatment; as a result, the lead time bias could be a potential problem. Despite these views, health policy tumor-related factors can use the results of this study to develop plans to reduce breast cancer recurrence. Among the prognostic factors, the lymph node ratio showed a close relationship with recurrent breast cancer. The findings indicated that developing public screening and educational programs through the health care system with more emphasis on low-educated women is needed among Iranian women.

Author Contributions
AY designed the model and the computational framework and analyzed the data. AY wrote the manuscript with support from SH. SH contributed to sample preparation. All authors discussed the results and contributed to the final manuscript. All authors have read and approved the manuscript.

ORCID iDs
Akram Yazdani [iD] https://orcid.org/0000-0002-8353-1653
Shahpar Haghighat [iD] https://orcid.org/0000-0003-2015-3197

Supplemental Material
Supplemental material for this article is available online.

REFERENCES
1. Santucci C, Carioli G, Bertuccio P, et al. Progress in cancer mortality, incidence, and survival: a global overview. *Eur J Cancer Prev*. 2020;29:367-381.
2. Zhao JM, An Q, Sun CN, et al. Prognostic factors for breast cancer patients with T1–2 tumors and 1–3 positive lymph nodes and the role of postmastectomy radiotherapy in these patients. *Breast Cancer*. 2021;28:298-306.
3. Francies FZ, Hull R, Khanyile R, Dlamini Z. Breast cancer in low-middle income countries: abnormality in splicing and lack of targeted treatment options. *Am J Cancer Res*. 2020;10:1568-1591.
4. Yazdani A, Yaseri M, Haghighat S, Kaviani A, Zeraati H. Investigation of prognostic factors of survival in breast cancer using a frailty model: a multicenter study. *Breast Cancer* (Auckland). 2019;13:1178223419879112.
5. Akbari ME, Akbari A, Nafissi N, et al. Prognostic factors of recurrence (early and late) and death in breast cancer patients in Iranian women. *Int J Cancer Manag*. 2016;9:e5747.
6. Northouse LL, Mood D, Kershaw T, et al. Quality of life of women with recurrent breast cancer and their family members. *J Clin Oncol*. 2002;20:4050-4064.
7. Lee J, Kim SH, Kang BJ. Prognostic factors of disease recurrence in breast cancer using quantitative and qualitative magnetic resonance imaging (MRI) parameters. *Sci Rep*. 2020;10:1-13.
8. Saad ED, Squifflet P, Burzykowski T, et al. Disease-free survival as a surrogate for overall survival in patients with HER2-positive, early breast cancer in trials of adjuvant trastuzumab for up to 1 year: a systematic review and meta-analysis. *Lancet Oncol*. 2019;20:361-370.
9. Yazdani A, Akbari H. Association of CA 15-3 and CEA with liver metastases in patients with breast cancer. *Curr Cancer Ther Rev*. 2020;16:332-336.
10. Yazdani A, Yaseri M, Haghighat S, Kaviani A, Zeraati H. The comparison of censored quantile regression methods in prognosis factors of breast cancer survival. *Sci Rep*. 2021;11:1-12.
11. Xue X, Xie X, Strickler HD. A censored quantile regression approach for the analysis of time to event data. *Stat Methods Med Res*. 2018;27:955-965.
12. Portnoy S. Censored regression quantiles. *J Am Stat Assoc*. 2003;98:1001-1012.
13. Yazdani A, Zeraati H, Yaseri M, Haghighat S, Kaviani A. Laplace regression with clustered censored data. *Comput Stat*. 2022;37:1041-1068.
14. Koenker R, Bassett G Jr. Regression quantiles. *Econometrica*. 1978;46:33-50.
15. Hong HG, Christiani DC, Li Y. Quantile regression for survival data in modern cancer research: expanding statistical tools for precision medicine. *Precis Clin Med*. 2019;2:90-99.
16. Faradmal J, Roshanaei G, Mafi M, Sadighi-Pashaki A, Karami M. Application of censored quantile regression to determine overall survival related factors in breast cancer. *J Res Health Sci*. 2016;16:36-40.
17. Bottai M, Zhang J. Laplace regression with censored data. *Biometrical J*. 2010;52:487-503.
18. Karihtala P, Jääskeläinen A, Roininen N, Jukkola A. Prognostic factors in metastatic breast cancer: a prospective single-centre cohort study in a Finnish University Hospital. *BMJ Open*. 2020;10:e038798.
19. Koca B, Kuru B, Karabicak I, Yuruker SS, Ozen N. Prognostic factors affecting disease-free survival in patients at age 35 or younger with invasive breast cancer. *Ann Ital Chir*. 2014;85:249-253.
20. Diniz RW, Guerra MR, Cintra JRD, Fayer VA, Teixeira MTB. Disease-free survival in patients with non-metastatic breast cancer. *Rev Assoc Méd Bras*. 2016;62:407-413.