Original Article

Determination of the Risk Factors for Breast Cancer Survival using the Bayesian Method, Yazd, Iran

Abstract

Background: There are numerous sophisticated studies which have investigated risk factors of breast cancer (BC). The purpose of this paper is to use benefits of Bayesian modeling to involve such prior information in determining factors affecting the survival of women with BC in Yazd city. **Materials and Methods:** The checklist included the characteristics of the patients and the factors studied. Then, from the records of patients referred to Radiotherapy Center of Shahid Ramezanzadeh, who had BC, from April 2005 to March 2012, the survival of 538 persons was recorded in the census. Data were analyzed by R software version 3.4.2, and 0.05 was considered the significance level. **Results:** The mean age of BC diagnosis was 48.03 ± 11016 years. The Bayesian Cox regression showed that surgery (hazard ratio [HR] =1.631 95% PI; 1.102–2.422), ki67 (HR = 3.260. 95% PI; 1.6308–6.372), stage (HR = 5.620, 95% PI; 4.079–7.731), lymph node (HR = 1.765, 95% PI; 1.127–2.790), and ER (HR = 2. 600 95% PI; 2.023–3.354) were significantly related to survival time. **Conclusion:** The parametric and cox models were compared with standard error, and Cox model was selected as an optimal model. Accordingly, stage, ki67, lymph node, ER, and surgery variables had a positive effect on death hazard.

Keywords: *Bayesian method, breast cancer, regression analysis, risk factors, survival analysis*

Vida Pahlevani, Morteza Mohammadzadeh[1], Nima Pahlevani[2], Vajiheh Nayeb Zadeh

Department of Biostatistics and Epidemiology, Shahid Sadoughi University of Medical Sciences and Health Services, Yazd, [1]Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, [2]Department of Medical Sciences, Kashan University, Kashan, Iran

Introduction

Cancer is a chronic disease that dedicates a high rate of death to itself in many societies in recent decades. Breast cancer (BC) is the most frequent malignancy in women around the world. Its incidence and mortality rates have been rising in Asian countries.[1] The available cancer treatments are generally costly, along with multiple complications and the response rate to treatment is not complete in many cases.[2] Furthermore, BC is the most common cause of death in female 40–44 years old in many advanced and developing countries. It is the second leading cause of cancer death after lung cancer.[3] It is estimated that cancer death will reach 13.1 million in 2030. However, 40% of deaths from cancer can be prevented by identifying and controlling the risk factors.[4] BC is abnormal and malignant proliferation of breast tissue cells, which is generally divided into two groups of carcinoma *in situ* (noninvasive) and invasive cancer.[5] According to the latest statistics, the incidence of BC in Iranian female is

27.5/1000 people. The 5-year survival rate in these patients ranges from 48% to 84%, and the overall survival rate is 72%.[6] This cancer is a type of hormone-related disease and malignant proliferation of epithelial cells that cover the breast lobule and lactiferous duct. Sixteen percent of all cancers in Iran are related to BC, which is ranked first among females.[7] In this paper, two tumor markers of ER, PR Ki67, and Her2 were applied. In survival analysis, time variable is usually called as the "survival time" because this variable determines the length of time that a person has "survived" during the follow-up period.[8] In the majority of medical studies, methods such as Cox regression are used when the purpose is to evaluate the survival distribution. Cox regression is a much more popular choice than parametric regression, because the nonparametric estimate of the hazard function offers you much greater flexibility than most parametric approaches, but the proportional hazard (PH) assumption for all predictor variables in model is an important and fundamental hypothesis for this model that proportion hazard has been constant

Address for correspondence:
Dr. Vajiheh Nayeb Zadeh, Shahid Sadoughi University of Medical Sciences and Health Services, Yazd, Iran.
E-mail: vajihenayebzade16@gmail.com

Access this article online
Website: www.advbiores.net
DOI: 10.4103/abr.abr_152_19
Quick Response Code:

over time. Only if the PH assumption is being satisfied, we are allowed that fit Cox regression on data.[9] Hitherto many studies have been conducted on the Cox regression model, but according to a systematic study, only 5% of these studies have considered the PH assumption.[10] In parametric models, survival times have a known probability distribution such as Weibull, exponential, lognormal, and log logistic.[11] Medical communities use Bayesian methods for survival analysis due to limitations such as the small sample size or lots of censorship. The main reason for this introduction is to combine the previous information and the data in the Bayesian method, which makes the estimations accurate and riches the findings. Optimal decisions can be made by observing new data and arguing about the probability distribution. In this paper, the parametric and Cox models are fitted on the data using the Bayesian method considering different prior distributions and meta-analyses. Initially, the various survival parametric models including exponential, Weibull, Gompertz, log logistic, lognormal, generalized Fisher, and generalized gamma and Cox were fitted by the Bayesian method on data. The final model was selected based on the Bayesian standard deviation criterion, and accordingly, factors affecting on survival of BC were determined.

The goal in this study has two folds: our primary purpose is to investigate effect of tumor markers on patients' survival such that involve prior information about them in our analysis. Second, find optimal model between several parametric and Cox model in Bayesian perspective. Bayesian models can combine the previous information with the data. Thus, when rich sources of prior information about risk factors are available, this approach can overcome classical statistical models.

Materials and Methods

Initially, a checklist containing the patients' specifications and all the studied factors(age, size of the tumor, lymph node involvement, primary metastasis, stage of disease, type of pathology (vascular or neurological lymph invasion), grade of disease, tumor markers (Her2, ER, PR, and Ki67), type of surgery (mastectomy or breast-conserving therapy [BCT]), collateral treatments (post-surgery radiotherapy, post-surgery chemotherapy, hormone therapy), distance of metastasis, and survival of patients) were prepared. Then, the patients' file in the archives of Shahid Ramezanzadeh Radiotherapy Center in Yazd, who had BC was examined from the beginning of 2005 to the end of 2012, and they were called to determine the survival of 538 patients. In addition, the time of patients' death was given from the provincial health center. In this analytical and survival analysis study, exploratory factor analysis was used to identify correlated predictor variables and to reduce the dimension of data, due to the lots of predictor variables in model. The variables were allocated to five independent factors using exploratory factor analysis.

Finally, seven variables including Her2, Ki67, estrogen receptor (ER), age (less than or over 40), type of surgery (mastectomy or BCT), stage of disease (primary or advanced), and lymph node involvement (positive or negative) were considered as a risk factor for the survival of BC according to the results of factor analysis and expert opinion.[12] In the next step, the parametric models fitted with different survival models including exponential, Weibull, Gompertz, log logistic, lognormal, generalized Fisher, and generalized gamma on the variables.[13] The optimal parametric model was selected using Akaike information criterion (AIC), and then, the optimal model fitted by Bayesian method.

Akaike criterion was used to determine the optimal parametric model. Akaike criterion was presented by Akaike in 1974 to assess the goodness of fitting models and is defined as follows:

AIC = −2 (log [likelihood]) +2 (a + c)

Where, a is the number of model parameters and c is a constant coefficient that is different from that of the applied model, and the smaller Akaike criterion means better fitting. Furthermore, the Bayesian Cox model was fitted, and finally, the optimal model was determined based on Bayesian standard error (SE).

In this study, packages SurvMisc,[14] icenReg,[15] and Coda[16] were used in the R software for Bayesian survival analysis. In all tests, α =0.05 was considered a significant level.

Results

In this study, 538 patients had BC which 109 patients were died (20.3%). Using the Kaplan–Meier method, the survival rates of 1, 3, and 8 years old were estimated 0.976, 0.898, and 0.737, respectively. Mean age of patients was 48.03 ± 11.66 years old and the mean survival time of patients was of 97.64 ± 4.23 months. Table 1 reports descriptive statistics. In the parametric modeling, all listed parametric models fitted on these seven variables, and "Log Normal model" was selected as the optimal parametric model (AIC = 464.77).Then, the log normal model was fitted with the Bayesian method. Gibbs sampling method was used for Markov chain Monte Carlo (MCMC) technique. The fitting results of Bayesian Lognormal model were given in Table 2. The mean and variance of significant variables were determined by studying different studies for Cox Bayesian method. The normal distribution was selected as the informative prior distribution. Bayesian Cox model was considered in the multiple form. For parameters h, hk, and β used bottom prior distribution:

$$h_k \overset{iid}{\sim} Ga(r_0 h, r_0), k = 1,..., M$$

$$h \sim N (h_0, \upsilon_0^2)$$

Table 1: Frequency of patients with breast cancer in terms of the variables affecting the disease (*n*=538)

Variables		Frequency (Percent)	Survival time mean (Standard Deviation)
Age	<40 [ref]	132(%24.5)	84.81 (6.17)
	>40	406(%75.5)	86.98 (3.98)
Ki67	low	29(%9.2)	87.26 (8.22)
	high	287(%90.8)	87.74 (4.45)
Surgery	mas	226(42%)	93.52 (3.80)
	BCT	312(58%)	100.71 (3.43)
Stage	Elementary	100(%18.6)	4.1210 (6.56)
	Advanced	438(%81.4)	(3.77) 37.123
ER	neg	158(%37.4)	90.97 (7.29)
	pos	264(%62.6)	104.24 (4.31)
Her2	neg	259(%69.3)	7.806 (2.48)
	pos	115(%30.7)	9.878 (3.94)
Lymph node	no	176 (33%)	102 0.48 (6.26)
	yes	357 (67%)	98 0.16 (3.58)

Table 2: Estimate Of Bayesian Log Normal

	RR (Risk Raito)	Mean	Naïve SE	PI€
mu	-	6.29	<0.001	(5.58 , 7.32)
log_s	-	0.48	<0.001	(0.17, 0.84)
age.	0.83	-0.17	<0.001	(-1.07, 0.64)
Ki67	0.88	-0.12	<0.002	(-1.26 , 1.11)
ER	1.49	0.40	<0.001	(-0.27 ,1.14)
MAS	0.67	-0.39	<0.001	(-1.20, 0.34)
Her2*	0.47	-0.74	<0.001	(-1.59, -0.06)
lymph.node	0.91	-0.08	<0.001	(-0.87 , 0.68)
stage	0.41	-0.88	<0.002	(-2.13 , 0.13)

€Probability Interval (Credible Interval)

$$\beta \mid \mu_0, \sum 0 \underset{\sim}{iid}\ N_p\ (\mu_0, \sum 0)$$

Posterior distribution for β is

$$f(\beta \mid else) \alpha\ exp\{\sum_{i=1}^{n}(-A_0\ (t_i)\ e^{x_i'\beta} + x_i'\beta) - \frac{1}{2}$$

$$(\beta - \mu_0)' \sum_0^{-1}(\beta - \mu_0) - \frac{1}{2}z'\ (C^{-1} - I_n)z\}$$

Moreover, Schonfeld test[18] was used to testing the proportion hazard assumption (PH). The results of the test were satisfactory (Global *P* = 0.399).

In the Bayesian method, MCMC is used so that it estimates the posterior distribution of the parameters under the specific loss function by performing the large sampling. In this research, Gibbs sampling was performed with indicators (nburn = 6000; nsave = 60,000; nskip = 20 + 4; niter = nburn + nsave) and the square error function was considered. Therefore, the posterior distribution mean was calculated as the final estimate of HR for each variable. These samples should be random without any particular process (they should not have autocorrelation). Geweke and Raftery Lewis[17] indicators were used to randomize

samples (close to 1) and lack of autocorrelation (<2.6). In addition, the statistical inferences are done using probability interval (credible interval) instead of *P* value. Figure 1 shows the Bayesian survival for significant variables in this study, and Table 3 was given the result of the Bayesian Cox method.

Discussion

The use of the Bayesian factors is a good idea for a measure of comparing parametric models and Cox model, but in practice, they are not measured in statistical packages in the Bayesian survival analysis, and their manual calculation is not possible due to a high number of sample size and variables. Hence, the SE has been selected in this paper as a comparison in this paper.[18] Based on this, the Bayesian Cox model was selected as the optimal model because it had a minimum of standard error between models, so based on this, the factors affecting BC were chosen. In this study, the age variable was not significant, which is similar to Tazhibi *et al*.'s study. who studies the survival data of 996 patients with BC in Isfahan to evaluate the risk factors in metastatic patients.[19] This result can be due to the use of the same cut point for the age variable. The Ki67 variable is significant, and patients with a Ki67-positive rate have 3.260 times greater death risk than the negative one. This result has been confirmed in other studies.[20] The role of Ki67 in BC has always been disagreement.[21] In Nishimura *et al*. study, a high level of Ki67 was associated with a low survival,[22] but in Bryan study, there was no relationship between Ki67 and androgen receptor.[23]In addition, For the type of surgery variable, patients with mastectomy survived 1.631 times less than BCT. Moslemi *et al*. proposed BCT method because of breast maintain and mental relaxation for patients, but unlike our results, it was not significant.[24] Saadatmand *et al*. investigated the effect of the stage disease on the survival of 173,797 patients. It was reported that mastectomy had more death hazard, and the Her2 variable was not significantly similar to this study.[25] However, by using of the Bayesian method in our study even with a small sample size, we obtained similar results with Saadatmand's study, The stage of disease was significant in most studies, such as Rakha *et al*.[26] In this study, patients with advanced stage of disease have 5.620 times higher death hazard than those in the primary stages of the disease.[27] In addition, the ER was significant and patients with positive ER were 2.6 times more death hazard than negative ER patients, which were similar to the most studies.[28] Finally, the lymph node variable involvement was statistically significant, and patients with lymph node involvement had 1.765 times higher death hazard than patients without lymph node involvement. The results were similar to the results of Khodarahmi *et al*. (2015)[29] and Fallahzadeh *et al*.[30] in the same year. Medically, it has been proven that lymph node involvement

Table 3: Estimates for Bayesian Cox Regression

| Variables		Multiple Bayesian Cox regression model				
		Hazard ratio	95% Credible interval	Raftery lewiss	Geweke	SE
age	<40 (ref)	0.83	0.39-1.92	1.18	-0.68	<0.001
	>40					
Ki67	low (ref)	3.26	1.63-6.37	1.13	-0.48	<0.001
	High					
Surgery	BCT (ref)	1.63	1.10-2.42	1.11	-1.21	<0.001
	Mas					
Stage	Elementary (ref)	5.62	4.07-7.73	1.08	1.14	<0.001
	Advanced					
ER	neg (ref)	2.59	2.02-3.35	1.08	0.22	<0.001
	Pos					
Her2	neg (ref)	0.94	0.63-1.41	1.1	1.58	<0.001
	Pos					
lymph node	No (ref)	1.76	1.12-2.79	1.07	2.00	<0.001
	Yes					
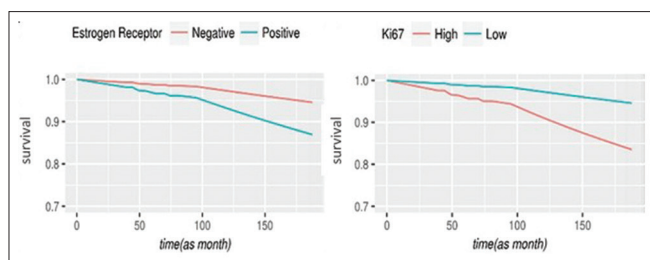
Figure 1: Survival plot for significant tumor markers. Width (Px): 624, Height (Px): 244

is one of the most influential factors in BC.[31] Shahireh Haghighat *et al*. evaluated the survival rate in BC patients. In a longitudinal study, they evaluated 623 patients with BC who referred to the Center of Breast Disease in Jihad University in 1997–2006. They concluded that HR ratio according to Cox analysis showed that lymph node involvement (HR = 2.52) and negative ER (HR = 2.60) significantly were important factors correlated with the survival rate of the patients.[32] Joseph Ibrahim, a researcher of Bayesian survival, reported in 2011[33] if there was no information about the parameters from previous studies, and the prior distribution of parameters was considered to be noninformative; then, the classical Cox and Bayesian Cox lead to the same findings. In this study, the informative prior distributions have been used, so results can be cited with a high power. One of the main problems of the research is the lack of enough sample size. However, high number of sampling has some limitations. The Bayesian method allows accurate results even in low samples based on information from meta-analysis and scientific resources. Combination of Bayesian and parametric methods in the survival analysis for diseases such as cancer gives more accurate results in terms of estimation error compared to the classic methods. In some of diseases such as BC that mortality rate in this cancer is lower than other cancers, Cure models are a better alternative to survival analysis.[34]

The combination of the Bayesian method and Cure models can be considered a substitute proposal in feature studies.

Conclusion

The Bayesian Cox model was selected as the optimal model because it had a minimum of standard error between models, so based on this, the factors affecting BC were chosen. In this study, stage, ki67, lymph node, ER, and surgery variables had a positive effect on death hazard.

Acknowledgment

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

References

1. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, *et al*. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature 2015;526:131-5.
2. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: Systematic review and meta-analysis to update the 2009 US preventive services task force recommendation effectiveness of breast cancer screening. Ann Intern Med 2016;164:244-55.
3. Kazemi A, Omid E, Amin MM, Nesaee P. A Survey on breast cancer status in Kurdistan province on Medical Geography viewpoint During 2006 - 2010.hsr. 2015;11:459–72.
4. Potosky AL, O'Neill SC, Isaacs C, Tsai HT, Chao C, Liu C, *et al*. Population-based study of the effect of gene expression profiling on adjuvant chemotherapy use in breast cancer patients under the age of 65 years. Cancer 2015;121:4062-70.
5. McGale P, Taylor C, Correa C, Cutter D, Duane F, Ewertz M, *et al*. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer

mortality: Meta-analysis of individual patient data for 8135 women in 22 randomised trials. Lancet (London, England). 2014;383:2127–35.

6. Jafari-Koshki T, Schmid VJ, Mahaki B. Trends of breast cancer incidence in Iran during 2004-2008: A Bayesian space-time model. Asian Pac J Cancer Prev 2014;15:1557-61.

7. Sharifian A, Pourhoseingholi MA, Emadedin M, Rostami Nejad M, Ashtari S, Hajizadeh N, *et al*. Burden of breast cancer in Iranian women is increasing. Asian Pac J Cancer Prev 2015;16:5049-52.

8. Cleves M, Gould W, Gould WW, Gutierrez R, Marchenko Y. An Introduction to Survival Analysis Using Stata, Second Edition. Stata Press; 2008. 398p.

9. Cox DR, Oakes D. Analysis of Survival Data. Boca Raton: Chapman and Hall/CRC; 2017. 212 p. Bennett S. Analysis of survival data by the proportional odds model. Stat Med 1983;2:273-7.

10. Klein JP, Houwelingen HC van, Ibrahim JG, Scheike TH. Handbook of Survival Analysis. CRC Press; 2016. 635p.

11. Zhou H, Hanson T. spBayesSurv: Bayesian Modeling and Analysis of Spatially Correlated Survival Data. R Package Version; 2014. p. 1.

12. Kabacoff RI. R in action: Data analysis and graphics with R. Simon and Schuster; 2015.

13. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence diagnosis and output analysis for MCMC. R News. 2006 Mar; 6:7–11.

14. Therneau TM. Extending the Cox Model. In: Lin DY, Fleming TR, editors. Proceedings of the First Seattle Symposium in Biostatistics. New York, NY: Springer US; 1997. p. 51–84. (Lecture Notes in Statistics).

15. Elandt-Johnson RC, Johnson NL, Statistiker M. Survival models and data analysis. Wiley Online Library; 1980.

16. Tazhibi M, Fayaz M, Mokarian F. Detection of prognostic factors in metastatic breast cancer. J Res Med Sci 2013;18:283-90.

17. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: Prognostic and predictive potential. Lancet Oncol 2010;11:174-83.

18. Vera-Badillo FE, Chang MC, Kuruzar G, Ocana A, Templeton AJ, Seruga B, *et al*. Association between androgen receptor expression, Ki-67 and the 21-gene recurrence score in non-metastatic, lymph node-negative, estrogen receptor-positive and HER2-negative breast cancer. Journal of clinical pathology. 2015;68:839–43.

19. Nishimura R, Osako T, Okumura Y, Hayashi M, Toyozumi Y, Arima N. Ki-67 as a prognostic marker according to breast cancer subtype and a predictor of recurrence time in primary breast cancer. Exp Ther Med 2010;1:747-54.

20. Bryan RM, Mercer RJ, Bennett RC, Rennie GC, Lie TH, Morgan FJ. Androgen receptors in breast cancer. Cancer 1984;54:2436-40.

21. Moslemi D, Gholizadeh PA, Hajian K, Sum SH, Pourghasem M, Jahantigh R. Comparison of Modified Radical Mastectomy with Breast Conservative Therapy and Radiotherapy in Patients with Breast Cancer; 2012.

22. Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MM. Influence of tumour stage at breast cancer detection on survival in modern times: Population based study in 173,797 patients. BMJ 2015;351:h4901.

23. Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. Cancer 2007;109:25-32.

24. Ren Y, Black DM, Mittendorf EA, Liu P, Li X, Du XL, *et al*. Crossover effects of estrogen receptor status on breast cancer-specific hazard rates by age and race. PLoS One 2014;9:e110281.

25. Vostakolaei FA, Broeders MJ, Rostami N, van Dijck JA, Feuth T, Kiemeney LA, *et al*. Age at Diagnosis and Breast Cancer Survival in Iran. Simon MS, editor. International Journal of Breast Cancer. 2012 Nov 22;2012:517976.

26. Khodarahmi S, Rezaianzadeh A. The role of prognostic factors on the survival of breast cancer patients: Bayesian approach. Iran J Epidemiol 2015;11:23-33.

27. Fallahzadeh H, Momayyezi M, Akhundzardeini R, Zarezardeini S. Five year survival of women with breast cancer in Yazd. Asian Pac J Cancer Prev 2014;15:6597-601.

28. de Boniface J, Frisell J, Andersson Y, Bergkvist L, Ahlgren J, Rydén L, '. Survival and axillary recurrence following sentinel node-positive breast cancer without completion axillary lymph node dissection: The randomized controlled SENOMAC trial. BMC Cancer 2017;17:379.

29. Haghighat S. Survival rate and its correlated factors in breast cancer patients referred to Breast Cancer Research Center. Iran J Breast Dis 2013;6:28-36.

30. Ibrahim JG, Chen MH, Sinha D. Bayesian semiparametric models for survival data with a cure fraction. Biometrics 2001;57:383-8.

31. Othus M, Mitchell A, Barlogie B, Morgan G, Crowley J. Cure-Rate Survival Models and Their Application to Cancer Clinical Trials. In: Matsui S, Crowley J, editors. Frontiers of Biostatistical Methods and Applications in Clinical Oncology [Internet]. Singapore: Springer; 2017. p. 165-78. Available from: https://doi.org/10.1007/978-981-10-0126-0_11. [Last accessed on 2021 Jun 26].

32. Haghighat S. Survival rate and its correlated factors in breast cancer patients referred to Breast Cancer Research Center. Iran J Breast Dis. 2013;6:28¨C36.

33. Ibrahim JG, Chen M-H, Sinha D. Bayesian survival analysis. Wiley Online Library; 2005.

34. Othus M, Mitchell A, Barlogie B, Morgan G, Crowley J. Cure-Rate Survival Models and Their Application to Cancer Clinical Trials. Front Biostat Methods Appl Clin Oncol. Springer; 2017. p. 165¨C178.