JRHS
Journal of Research in Health Sciences
journal homepage: www.umsha.ac.ir/jrhs

Original Article

Survival Analysis of Breast Cancer Patients using Cox and Frailty Models

Javad Faradmal (PhD)[a★], Atefeh Talebi (MSc)[a★], Abbas Rezaianzadeh (PhD)[b] and Hossein Mahjub (PhD)[c*]

[a] Department of Biostatistics & Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

[b] Department of Epidemiology, School of Health and Nutrition, Research Center for Health Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

[c] Department of Biostatistics & Epidemiology and Research Center for Health Sciences, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

★ The first and second authors have the same contribution.

ARTICLE INFORMATION

Article history:
Received: *09 August 2012*
Revised: *28 October 2012*
Accepted: *02 December 2012*
Available online: *05 December 2012*

Keywords:
Breast neoplasm
Cox model
Frailty

**** Correspondence***
Hossein Mahjub (PhD)
Tel: *+98 811 8380292*
Fax: *+98 811 8380509*
E-mail: *mahjub@umsha.ac.ir*

ABSTRACT

Background: Cox proportional hazard (CPH) model is the most widely used model for survival analysis. When there are unobserved/unmeasured individuals factor, then the results of the Cox proportional hazard model may not be reliable. The purpose of this study was to compare the results of CPH and frailty models in breast cancer (BC) patients.

Methods: A historical cohort study was carried out using medical records gathered from the Fars Province Cancer Registry. The dataset consisted of 769 women having BC referred to Shiraz Namazi Hospital, south of Iran. These patients had been followed for 6 years. After selecting the most important prognostic risk factors on survival, CPH and gamma-frailty Cox models were used to estimate the effects of the risk factors.

Results: The results of CPH model showed that, tumor characteristics and number of involved lymph nodes increase the mortality hazard of BC ($P < 0.05$). In addition, the frailty model showed that there is at least a latent factor in the model ($P = 0.005$).

Conclusion: Both of the frailty and CPH model emphasis that the early detection of BC improves survival in BC patients.

Citation: Faradmal J, Talebi A, Rezaianzadeh A, Mahjub H. Survival Analysis of Breast Cancer Patients using Cox and Frailty Models. J Res Health Sci. 2012;12(2):127-130.

Introduction

Breast cancer (BC) is the most common hormone-dependent cancer in women. This cancer is the most frequent cancer among Iranian women comprising 23.6% of all recorded cases of cancer[1-2]. In case of diagnosis of BC in early stages; the chance of recovery increases[1]. Survival analysis techniques were used for analyzing time to event analysis[1, 3-5]. Cox proportional hazard (CPH) model is the most widely used models for survival analysis with the strong assumption, proportional hazard (PH) assumption, which is reasonable in short follow-up studies[6].

When there is at least one unaccounted predictors in the model, a random component designed to account for variability due to unobserved individual-level factors called frailty[7]. The model is such that, events (e.g. death) happen sooner for those who, are more frailty[8-10]. Hereditary, genetic characteristics, growth and living environment are factors that caused difference between the patients. Regardless of the factors if widespread models such as CPH regression are used then the estimates are biased and the variances of the parameters are underestimated [11].

Several studies have been done for analysis of survival and/or hazard in BC patients in which CPH model is the most widely used model [12]. Li et al. used gamma frailty model for analysis of survival in BC patients[13]. Wienke et al. used a correlated log normal frailty in order to analysis BC data from the Swedish Twin Registry [14]. Perperoglou and Keramopoullos fitted CPH model with time-dependent covariates in BC patients [6]. Rezaianzadeh and colleagues used CPH model in BC patients in the south of Iran. In their study, they examined factors affecting survival of patients using CPH model [15]. Bellera et al. examined the effects of time-dependent variables on survival in BC patients [16].

Due to probably unmeasured factors affecting on survival time of BC patients in the south of Iran, the goal of present study was to compare the factors affect on BC survival using CPH and frailty model.

Methods

In this historical cohort study, we obtained the information of 769 women diagnosed BC who met the following criteria.

(i) no metastasis at the time of diagnosis, (ii) no previous BC, (iii) patients who underwent surgery including axillary dissection as the first treatment followed by chemotherapy and then radiotherapy. These data were gathered from patients' medical records referred to Shiraz University Cancer Registry Centre during 2000-2005, South of Iran. There were 11 recorded factor containing baseline characteristics and pathological factors. The time (days) elapsed since the cancer diagnosis until death was measured as outcome variable. Some variables containing tumor size, tumor grade, number of involved lymph nodes and age at diagnosis were enforced to include in the models.

Statistical analysis

The CPH model is a semi-parametric model with PH assumption [7]. The PH assumption was assessed using Schoenfeld residuals. Nevertheless it is possible, considering the effects of some predictors are depending on time. Such predictors are called time-dependent variables. If time-dependent variables are considered, a modified version of CPH model, extended CPH, that no longer satisfies the PH assumption may still be used [7].

If we entered all variables in the model, since missed values was relatively large in the gathered data, there were only 50 patients experiencing death in the data. So, in the present study, selected variables based on importance in other studies were tumor size, grade, number of involved lymph nodes and age at diagnosis[1, 15]

In the first step, we have fitted an extended CPH model to the data. When the number of involved lymph nodes was enter in the model as a categorical variable (0, 1-5, 6-10 and more than 10 lymph nodes) a weak sign of time-dependent effect was appeared (p=0.048). Whereas, when the mentioned variable was considered as a quantitative ones in the extended CPH model, its effect was not longer significant (P=0.06). So, in the CPH model the variable was considered as a qualitative. Since, the status of involved lymph nodes is an important prognosis factor in hazard of death or metastasis regardless of its effect, this variable also was entered in the model.

Besides, since some important factors such as hormonal status neither was recorded in the data, it was possible to have a great heterogeneity between patients. As in CPH model the unobserved variables is not considered into account, so a frailty model can be utilized. The frailty factor is an unobserved multiplicative effect on the hazard function that is assumed to have a g(α) distribution with unit mean and unknown variance of θ. In this study we considered gamma distribution for frailty component in Cox model [13]. Finally, we compared the results of CPH and frailty models using concordance index[17]. Concordance is defined as proportion of pairs observation having shorter survival times also have the larger risk score[17].

Level of Significance for statistical tests was 0.05. The R software version 2.15.2 was used for statistical analysis [18].

Results

Of patients included in the analysis, 617 (80.2%) patients were alive at the end of the study and 152 (19.8%) patients died. In this study, the mean (sd) and median of follow up time were 33.1 (16.1) and 31.5 months, respectively. The overall mean (sd) of survival time was 62.6 (1.3) months. The mean (sd) of patients' age at diagnosis was 46.6 (10.9) ranging from 19 to 80 years. Mostly, the patients were 35-59 years old, married with less than 3 children (Table 1).

Table 1: Patients' Characteristics at diagnosis by the patient status at the end of follow-up

Demographical characteristics	Alive		Dead		Total	
	Number	Percent	Number	Percent	Number	Percent
Education						
Illiterate	116	84.7	21	15.3	137	100.0
Primary	125	86.2	20	13.8	145	100.0
High	81	86.2	13	13.8	94	100.0
University	49	80.3	12	19.7	61	100.0
Total	371	84.9	66	15.1	437	100.0
Smoking						
Non-Smoker	493	81.4	113	18.6	606	100.0
Smoker	122	75.8	39	24.2	161	100.0
Total	615	80.2	152	19.8	767	100.0
Menopause Status						
Menopause	265	81.0	62	19.0	327	100.0
Premenopause	352	79.6	90	20.4	442	100.0
Total	617	80.2	152	19.8	769	100.0
Age Group						
<=35	90	78.3	25	21.7	115	100.0
>35 & <50	283	79.5	73	20.5	356	100.0
>=50	244	81.9	54	18.1	298	100.0
Total	617	80.2	152	19.8	769	100.0
Marital Status						
Single	52	76.5	16	23.5	68	100.0
Married	507	81.0	119	19.0	626	100.0
Divorced & Widowed	57	77.0	17	23.0	74	100.0
Total	616	80.2	152	19.8	769	100.0
Menarche Age						
<=13 Yrs	317	83.4	63	16.6	380	100.0
>13yrs	203	79.3	53	20.7	256	100.0
Total	520	81.8	126	18.2	636	100.0
Children						
<=3 Child	314	82.2	68	17.8	382	100.0
4-8 Child	279	79.7	71	20.3	350	100.0
>8 Child	21	65.6	11	34.4	32	100.0
Total	614	80.4	150	19.6	764	100.0

Table 2 shows some pathological factors by the patient status at the end of follow-up. The results of this Table indicated that majority of women have tumor ranging 2-5 cm and no or less than 5 involved lymph nodes.

Table 2: Pathological characteristics of the breast cancer women

Tumor characteristics	Alive		Dead		Total	
	Number	Percent	Number	Percent	Number	Percent
Side						
Left	286	80.1	71	19.9	357	100.0
Right	331	80.3	81	19.7	412	100.0
Total	617	80.2	152	19.8	769	100.0
Tumor grade						
Poorly-differentiated	70	60.9	45	39.1	115	100.0
Moderately-differentiated	373	80.9	88	19.1	461	100.0
Well-differentiated	174	90.2	19	9.8	193	100.0
Total	617	80.2	152	19.8	769	100.0
Involved lymph node						
Negative Lymph node	257	95.2	13	4.8	270	100.0
1-5 Lymph node	230	83.9	44	16.1	274	100.0
6-10 Lymph node	68	69.4	30	30.6	98	100.0
More than 10 Lymph node	62	48.8	65	51.2	127	100.0
Total	617	80.2	152	19.8	769	100.0
Tumor size						
≤2 cm	204	89.5	24	10.5	228	100.0
2-5 cm	302	80.5	73	19.5	375	100.0
≥5 cm	111	66.9	55	33.1	166	100.0
Total	617	80.2	152	19.8	769	100.0

The test based on Schoenfeld residuals showed that all of variables met the PH assumption (Global *P*=0.060). We provide the results of CPH and frailty model in Table 3 and 4, respectively. As it can be seen, tumor size, tumor grade, status of involvement of lymph nodes and number of involved lymph nodes are significantly related to the hazard rate in both models. The result of gamma-frailty Cox model indicates that the hazard of death due to BC in women with a tumor bigger than 5 cm is about 2.3 times compared to the women with a tumor less than 2 cm in diameter (*P*=0.015). In addition it can be seen that increase in the number of involved lymph nodes in lymph node positive patients increases in the death hazard (*P*=0.002).

Table 3: Adjusted hazard rate (HR) estimation for prognostic risk factors on survival using Cox proportional hazard model (Concordance= 0.792, SE = 0.026)

Factors	Coefficient	SE	HR	95%CI	*P* value
Tumor size					
2-5cm	0.326	0.239	1.386	0.868, 2.212	0.172
≥5 cm	0.620	0.252	1.860	1.135, 3.047	0.014
Involved lymph nodes					
Number	0.051	0.008	1.053	1.037, 1.069	<0.001
Positive lymph nodes					
Number	1.348	0.304	3.849	2.123, 6.980	<0.001
Tumor grade					
Moderately differentiated	0.348	0.259	1.417	0.853, 2.353	0.179
Poorly differentiated	0.991	0.283	2.695	1.549, 4.691	<0.001
Age group					
35-50 yrs	-0.307	0.236	0.735	0.463, 1.167	0.192
≥50 yrs	-0.221	0.245	0.802	0.496, 1.296	0.367

Table 4: Adjusted hazard rate estimation for prognostic risk factors on survival using gamma-frailty Cox model (Concordance = 0.936, SE = 0.026)

Factors	Coefficient	SE	HR	95% CI	*P* value
Tumor size					
2-5	0.377	0.319	1.458	0.781, 2.72	0.240
≥5	0.842	0.347	2.321	1.177, 4.58	0.015
Involved lymph nodes					
Number	0.136	0.018	1.145	1.105, 1.19	<0.001
Involved lymph nodes					
Number	1.189	0.377	3.283	1.568, 6.88	0.002
Tumor grade					
Moderately differentiated	0.553	0.329	1.738	0.913, 3.31	0.093
Poorly differentiated	1.310	0.385	3.706	1.742, 7.89	0.001
Age					
35-50	-0.126	0.340	0.882	0.453, 1.72	0.710
≥50	-0.086	0.352	0.917	0.460, 1.83	0.811
Variable					
Frailty	1.920	-	-	-	0.005

A negative coefficient for age categories indicated that younger women (age at diagnosis <=35) had worse prognoses, compared to 35-50 years (HR=1.36; *P*=0.192) as well as older than 50 years one (HR=1.25; *P*=0.367). However, these findings were not significant statistically.

Table 4 shows that the variance of frailty is significantly greater than zero (θ=1.92; *P*=0.005). It shows that there are latent factors that affect on hazard of death. The concordance index in the frailty model is 0.936 which is bigger than 0.792 in the CPH model (*P* < 0.001).

Discussion

In the present study, we well fitted the CPH and gamma-frailty Cox models to estimate the adjusted hazard of BC patients who underwent surgery. Our results showed that tumor characteristics have direct effect on hazard of death. These findings are consistent with Faradmal et.al[1] findings. Schmitt et al. studied time-dependent effects of biological factors in BC patients [19]. This study performed on 314 BC patients during 58 months of follow-up. The significant variables in the CPH model include vessel invasion, tumor grade, tumor necrosis, tumor size, lymph node involvement status and the status of hormones. Time-dependent variables, number of involved lymph nodes and estrogen receptors with time function $g(t) = t$, were considered in the extended Cox model [19]. In the present study, we enforced some variables as important factors in hazard of death in BC patients. Categorized number of lymph nodes was a time-dependent factor[19]. However when we enter it into the CPH model in its original scale, it was not longer a time-dependent variable.

Perperoglou et al. fitted several models in long-term survival in BC patients [6]. The main variables were age at diagnosis, tumor size, number of involved lymph nodes, tumor grade, the status of chemotherapy treatment, the status of receiving radiotherapy and surgery. Then CPH, extended Cox, frailty and cure model were fitted. Time-dependent variable was, status of receiving radiotherapy and time function was considered as $g(t)=log(t)$. Using Brier score and coefficient of determination, a comparison between these models was made. They concluded that the extended Cox and then the frailty model showed proper fit. In the current study, gamma-frailty Cox model was fitted to the hazard of BC patients. The difference between our study and Perperoglou et al. is due to the employment of long-term follow up in their study.

In the study by Rezaianzadeh and et al., only CPH model was fitted. Three main factors included in the model were number of involved lymph nodes, tumor grade and tumor size[15]. Whereas, in our study were fitted CPH and frailty models to the data.

Bellera et al. fitted CPH and extended Cox models. In their article, age, tumor size, tumor grade, the number of involved lymph nodes, peritumoral vascular invasion, status of hormone receptors, Her2, and Mib1 considered as important variables. In that study, grade of tumor was considered as time-dependent variable with time function $g(t)=t$ [16]. In the present study, the time-dependent variable is number of involved lymph node when it categorized in some groups.

Although, this study was carefully prepared, but there are some limitations. First, hormonal factors as important factors on hazard have not recorded in the questionnaire. Second, the follow-up time was short. Finally, there are too many missing observations in some of the recorded factors.

Conclusion

Due to the results of the frailty model, it can be conclude that using more sophistic statistical model that consider the role

of latent variables such as environmental and genetic factors, may increase the efficacy of the further analysis. Also it can be seen that detection of BC in early stages can reduce the hazard of death which highlights the screening role.

Acknowledgments

This article is a part of MSc thesis supported by Hamadan University of Medical Sciences. We would like to thank the Deputy of Education as well as Deputy of Research and Technology of Hamadan University of Medical Sciences for funding this study.

Conflict of interest statement

None declared.

Funding

This study was funded by the Vice-chancellor of Research and Technology of Hamadan University of Medical Sciences

References

1. Faradmal J, Kazemnejad A, Khodabakhshi R, Gohari MR, Hajizadeh E. Comparison of three adjuvant chemotherapy regimes using an extended log-logistic model in women with operable breast cancer. *Asian Pac J Cancer Prev.* 2010;11(2):353-358.
2. Haghpanah S, Amini M, Kherad M, Sadeghimehr R. Knowledge and Practice of Patients with Breast Cancer about Complication of Chemotherapy. *J Res Health Sci.* 2006;6(1):28-32.
3. Aghamolaei T, Zare S, Tavafian S, Abedini S, A P. IUD Survival and Its Determinants; a Historical Cohort Study. *J Res Health Sci.* 2007;7(2):31-35.
4. Soltanian AR, Mahjub H. A Non-parametric Method for Hazard Rate Estimation in Acute Myocardial Infarction Patients: Kernel Smoothing Approach. *J Res Health Sci.* 2012;12(1):19-24.
5. Mirzaee Ramhormozi S, Moghimbeigi A, Mahjub H, Soltanian AR. Birth Distance Influential Factors: A Multilevel Recurrent Events Approach. *J Res Health Sci.* 2010;10(2):98-103.
6. Perperoglou A, Keramopoullos A, van Houwelingen HC. Approaches in modelling long-term survival: an application to breast cancer. *Stat Med.* 2007;26(13):2666-2685.
7. Kleinbaum DG, Klein M. *Survival analysis: a self-learning text*. 2nd ed. New York: Springer; 2005.
8. Collett D. *Modelling survival data in medical research*. 2nd ed. London: Chapman & Hall/CRC; 2003.
9. Ma Z, Krings AW. *Multivariate Survival Analysis (I): Shared Frailty Approaches to Reliability and Dependence Modeling*. Moscow: Idaho; 2008.
10. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. *J Gerontol A Biol Sci Med Sci.* 2007;62(7):722-727.
11. Jonker MA, Bhulai S, Boomsma DI, Ligthart RS, Posthuma D, Van der Vaart AW. Gamma frailty model for linkage analysis with application to interval-censored migraine data. *Biostatistics.* 2009;10(1):187-200.
12. Tsodikov A. Semi-parametric models of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Stat Med.* 2002;21(6):895-920.
13. Li H, Zhong X. Multivariate survival models induced by genetic frailties, with application to linkage analysis. *Biostatistics.* 2002;3(1):57-75.
14. Wienke A, Locatelli I, Yashin AI. The Modelling of a Cure Fraction in Bivariate Time-to-Event Data. *Austrian Journal of Statistics.* 2006;35(1):67–76.
15. Rezaianzadeh A, Peacock J, Reidpath D, Talei A, Hosseini SV, Mehrabani D. Survival analysis of 1148 women diagnosed with breast cancer in Southern Iran. *BMC Cancer.* 2009;9:168.
16. Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med Res Methodol. 2010;10:20.
17. Terry Therneau. A Package for Survival Analysis in S. R package version 2.36-14. 2012 .
18. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 2012; available from: http://www.R-project.org.
19. Schmitt' M, Thomssen' C, UIm K, Seiderer A, Harbeck N, Hofler H, et al. Time-varying prognostic impact of tumour biological factors urokinase (uPA), PAI-1 and steroid hormone receptor status in primary breast cancer. *Br J Cancer.* 1997;76(3):306-311.