

Survival Analysis of Breast Cancer Patients Using Additive Hazards Regression Models

Goutam Barman^{1*}, Babulal Seal²

^{1*}Department of Statistics, Krishnagar Government College, Nadia, INDIA

²Department of Statistics and Informatics, Aliah University, Kolkata, INDIA

*Corresponding Author: goutambrmn@gmail.com

Available online at: www.isroset.org

Received 14th Apr 2017, Revised 26th Apr 2017, Accepted 25th May 2017, Online 30th Jun 2017

Abstract— Breast cancer is the most common malignant disease for females. Cox proportional hazards model are mostly used model to analysis the effects of prognostic factors to the breast cancer patients. In this study Aalen's additive hazards model and Lin-Ying's additive hazards model are used for survival analysis of breast cancer patients and compare with the results obtained by Cox proportional hazards model. The proportional hazards assumption was tested by using Schoenfeld residuals and p -value less than 0.05 was consider statistically significant. Also overall survival rate were estimated by the Kaplan-Meier product limit method. 686 patients with breast cancer and seven standard prognostic factors, namely age at diagnosis, menopausal state, tumor size, tumor grading, no. of involved nodes, progesterone and estrogen receptor were entered into analysis. Two models, Cox and Lin-Ying's models are given the similar results. Four covariates namely tumor size, tumor grade III, nodes and progesterone receptor showed significant impact on the breast cancer patient's data in both hazard models. Neither Cox model nor Lin-Ying model found age at diagnosis, menopausal state, tumor grade I, II and estrogen receptor as a significant prognostic factor. On the other hand, Aalen's model shows that tumour grade III is not statistically significant but the other results are similar with Cox and Lin-Ying models. Generally, the Cox and additive hazards models give different pieces of information about the risk factors. So, to get more accurate results about the risk factors, it is desire to use these two models in parallel. However, if the proportional hazards assumption is not satisfied, additive hazards model is an appropriate alternative for the Cox model otherwise both models are appropriate.

Keywords— Breast cancer, prognostic factor, Cox model, additive hazards models, survival analysis

I. INTRODUCTION

Breast cancer is the leading cause of death from cancer among women worldwide, accounting for 25% of all cases [1]. In 2012, 1.68 million breast cancer patients were diagnosed and from this disease, nearly 522,000 patients died [1]. In Germany, each year, approximately 57,000 new patients with breast cancer are registered on account, 27.8% of all cancer patients in German women [2]. So the prognostic factors identification associated with survival from this disease is very important.

In survival analysis of medical research, the major interests are to establish the relationship between the prognostic factors and a patient's time to death by choosing appropriate regression models. The Cox-PH model is the most widely popular model in survival analysis [3]. This model is intuitive, simple to fit and understanding the results is very easy. The hazards ratio for each explanatory variable is assumed to be constant over time. The validity of the analyses using by this model relies heavily on the PH assumptions and the model cannot include time-varying

covariates effects. Most researchers used Cox-PH model for the prognostic studies of breast cancer patients.

An unknown baseline hazard involved in Cox-PH model and this baseline hazard may assumed in a particular parametric form, namely, log-normal, log-logistic, Weibull, generalized gamma, etc. If the choice of the parametric baseline hazard is incorrect, the estimation will be inconsistent and biased [4, 5]. Also the Cox model may give biased conclusion when the PH assumption does not fulfill. Then additive hazards model is an alternative approach for survival analysis. Aalen's additive hazards model and Lin-Ying's additive hazards model are taken in this study. Several authors used additive hazards model for the analysis of the risk factor.

The primary aim of this study is to apply the additive hazards model to the study of the prognostic factors of the breast cancer patients and to compare the results obtained by the Cox model. A brief discussion about the dataset and the survival models, viz. Cox proportional hazards model, Aalen's additive hazards model and Lin-Ying's additive hazards model, selected for this study, is provided below.

II. MATERIALS AND METHODS

Study population:

The data used in this study for 686 women diagnosed with breast cancer from the period July 1984 to December 1989 were reported by German Breast Cancer Study group. Among 686 patients, 299 had an event for recurrence-free survival and 171 died. Seven standard prognostic factors namely age, menopausal state, tumor size, tumor grading, no. of involved nodes, progesterone and estrogen receptor are used in this analysis.

Statistical analysis:

For multivariate analysis of the prognostic factors, Aalen’s additive hazards model, Lin-Ying’s hazards model and Cox proportional hazards model were used and statistical analysis were performed in statistical software R-3.4.0 and p-value less than 0.05 was consider statistically significant. The proportional hazards assumption was tested by using Schoenfeld residuals. In this section, all models for analysis are reviewed.

a. Cox Proportional Hazards Model [3]:

Recently, in biomedical studies, the most popular regression model for survival analysis is the Cox proportional hazards model (simply, Cox model). In this model, the covariates have a multiplicative effect on some unknown unspecified baseline hazard function and the regression coefficients are constant. The Cox model in terms of the hazard function, associated with the covariate $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$, is defined as:

$$h(t | X_i) = h_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})$$

where $h_0(t)$ is the baseline hazard function and β_i ’s are the regression coefficients.

b. Aalen’s Additive Hazards Model [6, 7]:

This is an alternative model with the comparison of the Cox proportional hazards model but not popular. In this model, the covariates have an additive effect with some unknown baseline hazard and the regression coefficients are functions of time. So the effect may change with time. The Aalen’s additive hazard model (simply, Aalen’s model) in terms of the hazard function, associated with the covariate $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ is defined as:

$$h(t | X_i) = h_0(t) + \beta_1(t)X_{i1}(t) + \beta_2(t)X_{i2}(t) + \dots + \beta_p(t)X_{ip}(t)$$

where $h_0(t)$ is the baseline hazard function and β_i ’s are the regression coefficients.

c. Lin and Ying’s Additive Hazards Model [8]:

This model is closer with the Cox model. In this model, the covariates have an additive effect with some unknown unspecified baseline hazard function and the regression coefficients are constant. The Lin and Ying’s additive hazard

model (simply, Lin-Ying’s model) in terms of the hazard function, associated with the covariate $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$

is defined as:

$$h(t | X_i) = h_0(t) + \beta_1(t)X_{i1}(t) + \beta_2(t)X_{i2}(t) + \dots + \beta_p(t)X_{ip}(t)$$

where $h_0(t)$ is the baseline hazard function and β_i ’s are the regression coefficients.

The following section discuss and summarizes the results obtained by selected models in this study.

III. RESULTS AND DISCUSSION

From July 1984 to December 1989, 41 centres recruited 720 patients, of whom about two-thirds were randomized. Among them 686 (95.3%) patients with seven standard prognostic factors were completely available which are given in Table 1. These patients are taken as the basic population in this paper. At the time of diagnosis, the mean age was 53.05 ± 10.12 years. The mean and median of overall survival time were 44.03 and 44.6 months respectively. Of the patients, 506 (73.8%) had tumor size greater than 20 mm. Overall survival rate were estimated by the Kaplan-Meier product limit method.

Table 1. Patients characteristics with respect to prognostic factors

Variable	Category	No.	Percentage
Age at diagnosis	≤ 45	153	22.3
	> 45	533	77.7
Menopausal state	pre	290	42.3
	post	396	57.7
Tumor size (mm)	≤ 20	180	26.2
	> 20	506	73.8
Tumor grading	I	81	11.8
	II	444	64.7
	III	161	23.5
No. of nodes involved	< 10	583	85.0
	≥ 10	103	15.0
Progesterone receptor	< 20	269	39.2
	≥ 20	417	60.8
Estrogens receptor	< 20	262	38.2
	≥ 20	424	61.8

The results for proportional hazards assumptions testing are shown in **Table 2** and a plot of Schoenfeld residuals for the variable “prog_recp” is shown in **Figure 1**.

Table 2. Test for proportional hazards assumptions

	rho	χ^2	p-value
age	-0.097264	1.59	0.2078
menopause	0.098536	1.63	0.2015
size	-0.021331	0.0778	0.7804
as.factor (grade) 2	0.000335	0.0000195	0.9965
as.factor (grade) 3	-0.035762	0.222	0.6372
nodes	0.045602	0.357	0.5500
prog_recp	0.156444	4.43	0.0354*
estrg_recp	0.008156	0.0110	0.9165
GLOBAL	NA	11.9	0.1544

*statistically significant

From Table 2, the correlation between the Schoenfeld residuals for the variable ‘prog_recp’ and ranked survival time is 0.156444 with a p-value of 0.0354. This significant p-values proof that the proportional hazards assumption is not satisfied for the variable ‘prog_recp’. The p-values for the other variables are not significant suggest that there is not enough evidence to reject the proportional hazards assumptions for these variables. The global test for the entire model is not significant with $p = 0.1544$. This global test offers evidence that the proportional hazards assumption is satisfied for that model.

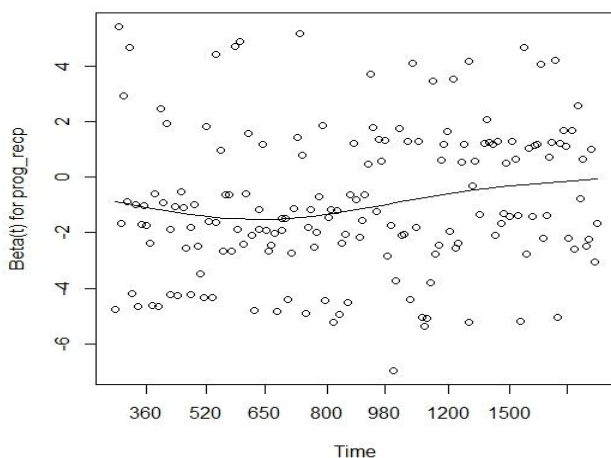


Figure 1. Schoenfeld residuals for ‘prog_recp’

The fitted curve does not look like horizontal and consequently the PH assumption for the variable ‘prog_recp’ does not satisfied. The Cox-PH model, Aalen’s additive hazards model and Lin-Ying’s additive hazards model was used separately to investigate the influence of several factors on the survival times.

Table 3. Multivariate analysis of prognostic factors for patients with BC using the Cox Proportional Hazards Model

Variable	Hazard Ratio	95% CI	p-value	
Age at diagnosis	0.8157	(0.5031, 1.3225)	0.4087	
Menopausal state	1.2256	(0.8055, 1.8647)	0.3422	
Tumor size	1.5514	(1.0421, 2.3097)	0.0305	
Tumor grading	I vs. II	2.0710	(0.8937, 4.7994)	0.0895
	I vs. III	2.4925	(1.0271, 6.0486)	0.0435
No. of node	2.7674	(1.9740, 3.8795)	<0.0001	
Progesterone receptor	0.3980	(0.2730, 0.5803)	<0.0001	
Estrogens receptor	0.8347	(0.5792, 1.2031)	0.3327	

Table 4. Multivariate analysis of prognostic factors for patients with BC using the Aalen’s Additive Hazards Model

Variable	Regression coefficient	Standard error	p-value	
Age at diagnosis	-0.000502	0.000478	0.294	
Menopausal state	0.000352	0.000386	0.362	
Tumor size	0.000700	0.000331	0.0346	
Tumor grading	I vs. II	0.000578	0.000316	0.0675
	I vs. III	0.001010	0.000562	0.0737
No. of node	0.003260	0.000761	< 0.0001	
Progesterone receptor	-0.002230	0.0000503	< 0.0001	
Estrogens receptor	-0.000388	0.000439	0.377	

Table 5. Multivariate analysis of prognostic factors for patients with BC using the Lin-Ying’s Additive Hazards Model

Variable	Regression coefficient	Standard error	p-value
Age at diagnosis	-0.0000383	0.0000412	0.341

Menopausal state		0.0000333	0.0000334	0.312
Tumor size		0.0000665	0.0000284	0.0195
Tumor grading	I vs. II	0.0000489	0.0000269	0.0735
	I vs. III	0.0000986	0.0000488	0.0418
No. of node		0.000286	0.0000670	< 0.0001
Progesterone receptor		-0.000190	0.0000434	< 0.0001
Estrogens receptor		-0.0000363	0.0000380	0.332

The results of the Cox and Lin-Ying’s models are given in Table 3 and Table 5 respectively. These two models give the similar results. Four covariates namely tumor size, tumor grade III, nodes and progesterone receptor showed significant ($p < 0.05$) impact on the breast cancer patient’s data in both hazard models. Neither Cox model nor Lin-Ying model found age at diagnosis, menopausal state, tumor grade I, II and estrogen receptor as a significant prognostic factor. On the other hand, the result of the Aalen’s model are given in Table 4. This model shows that tumor grade III is not statistically significant where according to other two models this covariate is statistically significant. The other results are similar with Cox and Lin-Ying models. The comparative results of these three models are shown in the following Table 6.

Table 6. Comparison of p-values for the prognostic factors by three models

Prognostic factors		Cox p-value	Aalen’s p-value	Lin-Ying’s p-value
Age at diagnosis		0.4087	0.294	0.341
Menopausal state		0.3422	0.362	0.312
Tumor size		0.0305*	0.0346*	0.0195*
Tumor grading	I vs. II	0.0895*	0.0675*	0.0735
	I vs. III	0.0435*	0.0737	0.0418*
No. of node		< 0.0001	< 0.0001	< 0.0001
Progesterone receptor		< 0.0001	< 0.0001	< 0.0001
Estrogens receptor		0.3327	0.377	0.332

Generally, the Cox and additive hazards models give different pieces of information about the risk factors. So, to get more accurate results about the risk factors, it is desire to use these two models in parallel. However, if the proportional hazards assumption is not satisfied, additive hazards model is an appropriate alternative for the Cox

model otherwise both models are appropriate. Our analysis suggest that tumor size, tumor grade, no. of involved nodes and progesterone receptor are important prognostic factors to increase survival of patients with breast cancer.

IV. CONCLUSION

Due to the results of additive hazards model, it can be conclude that when proportional hazards assumptions are satisfied, additive hazards model and Cox proportional hazards model both gives similar results. Also it can be shown that when PH assumption violated then additive hazards model is appropriate choice for survival analysis. Although in this situation we can also choose parametric models like lognormal, Weibull, Exponential etc. If the choice of the parametric models is not appropriate then the results will be biased.

In this study, tumor size, tumor grade III, nodes and progesterone receptor are the significant prognostic factors for breast cancer patients. However scope exists to analysis the risk factors of breast cancer from current dataset to increase the survival time of the patients.

ACKNOWLEDGEMENT

The authors would like to thank the German Breast Cancer Study group for providing the dataset.

REFERENCES

- [1] WCR, “World Cancer Report 2014”, World Health Organization, Switzerland, pp.1-120, 2014
- [2] Katalinic A, Pritzkeleit R, Waldmann A, “Recent trends in breast cancer incidence and mortality in Germany”, Breast Care (Basel) Vol.4, Issue.2, pp.75-80, 2009.
- [3] Cox DR, “Regression models and life-tables”, Journal of Royal Statistical Society, Vol.34, Issue.2, pp.187-220, 1972.
- [4] Mayer BD, “Unemployment insurance and unemployment spells”, Econometrica, Vol.58, Issue.4, pp.775-782, 1990.
- [5] Bhat CR, “A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity”, Transportation Research Part B: Methodological, Vol.30, Issue.3, pp.189-207, 1996.
- [6] Aalen OO, “A model for nonparametric regression analysis of counting process”, Lecture notes in statistics, Vol.2, Issue.2, pp.1-25, 1980.
- [7] Aalen OO, “A linear regression model for the analysis of life times”, Stat Med, Vol.8, Issue.8, pp. 907-925, 1989.
- [8] Lin DY, Ying Z, “Semi-parametric analysis of the additive risk model”, Biometrika, Vol.81, Issue.1, 61-71, 1994.