IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 5 Ver. V (Sep. - Oct.2016), PP 16-21
www.iosrjournals.org

Comparing Cox Proportional Hazard Model and Parametric Counterpart in the Analysis of Esophagus Cancer Patient Data

Rinku Saikia[1] and Manash Pratim Barman[2]
[1] *Research Scholar, Department of Statistics, Dibrugarh University, Dibrugarh, Assam, INDIA.*
[2] *Assistant Professor , Department of Statistics, Dibrugarh University, Dibrugarh , Assam, INDIA.*

Abstract: *Survival analysis is the analysis of statistical data in which the outcome variable of interest is time until an event occurs. In this paper, attempt has been made to find the best fitted model for studying the survival time of esophagus cancer patients of Assam. The present retrospective study is conducted on medical records of 178 patients. Akaike Information Criterion(AIC) and Baysian Information Criteria (BIC) and* R^2 *are used to identify the best fitted model. From the study it is found that Cox PH model is better than the other parametric counterparts for the esophagus cancer patients data.*

Keywords: *Cox PH model, AIC, BIC, esophagus cancer patients,* R^2 .

I. Introduction

Survival analysis is the analysis of statistical data in which the outcome variable of interest is time until an event occurs. The survival time, a non- negative random variable which is a length of time that is measured from the start of a study time to the time that the event of interest occurs. The event may be death, disease, etc. There are various characteristics in survival data. The characteristics are presents of censored observation, skewed distribution, lack of normality in distribution, etc. Because of these characteristics the traditional statistical methods or techniques cannot be applied to this type of data. Various methods have been developed to analyze this data. They are non- parametric (Kaplan-Meier method and log-rank test), semi parametric method (Cox Proportional Hazard (PH) model) and parametric methods (Parametric PH model and the Accelerated failure time model (AFT)). In medical science research though the semi-parametric Cox PH model is the most widely used to analyze the survival data, Parametric models are considered more preferable in the situations when the correct form of the parametric model is exactly known.(Ravangard et. al;)

Cox PH model is a semi parametric model in which the baseline hazard function is unspecified or has no particular form. In case of Cox PH model there is no particular parametric form for hazard and time. If the baseline hazard function has a specific parametric form such as Exponential, Weibull , Gompertz then the model is considered as a parametric proportional hazard model.

The main objectives of this paper are (i) to compare the semi parametric Cox PH model with its parametric counterparts and find the best fitted model (ii) to study the effect of different factors on the survival of esophagus cancer patients by using the best fitted model.

II. Materials And Methods

The study was taken up in a historical cohort and information from the medical charts of patients with esophagus cancer in Assam Medical College Hospital (AMCH) Dibrugarh , Assam. The period of the study was from 1st January 2007 to 31st December 2009. All the patients diagnosed with esophagus cancer during 1st January 2007 to 31st December 2008 were included in the study. Cases diagnosed during 2009 were excluded due to limited follow up (i.e., through 2009). During the inclusion period of the study a total of 178 patients were diagnosed with Esophagus cancer in AMCH. A pre-designed, pretested questionnaire was used for the collection of data. Information about age, sex, extension of the disease at the time of diagnosis, cancer directed collected from the hospitals records. The patients were considered as censored if they were alive beyond 31st December 2009, died due to other causes or loss to follow up. After collecting the hospitals records, a household survey were conducted to collect the information about the survival status of the patients, date of expired (if he/she expired), continuation of treatment and socioeconomic status of the patients. Also, a re-verification of the information collected from the hospital was made during the household visit. The extension of the disease includes the stages: localized (confined to the esophagus, with no evidence of spread to surrounding organs/tissues or no regional lymph nodes); regional (invasion beyond the organ to surrounding organs/tissues or no regional lymph nodes); distant/metastatic (spread to remote organs/tissues directly or by discontinuous metastasis) and unknown. Survival (in months) was estimated from the month of diagnosis until death, loss to follow up ,or the end of 2009. Patients are categorized into three groups based on the cancer directed treatment.

Those who are treated with surgery and others; other than surgery and who were not treated were termed as no treatment.

(i) Cox PH model: The Cox PH is proposed by Cox in 1972. Here the effect of covariate acts multiplicatively (proportionally) with respect to hazard .

The Cox PH model is given by

$$h(t,x) = h_0(t)\exp(\sum_{i=1}^{p} \beta_i x_i) \dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

where $h_0(t)$ is the baseline hazard function, exp is the exponential expression to the linear sum, (this sum is over p explanatory variable), x_i is the explanatory or predictor variable and β_i is the regression coefficient.

Cox model is widely employed model in survival analysis.

III. Parametric Proportional Hazard (PH) Model

The parametric PH model is the parametric version of Cox PH model. It is similar with the form of Cox PH model. The Cox model given in equation (1), the hazard function is unspecified If the baseline hazard function is assumed to follow a specific distribution such as Weibull, Gompertz etc. then the PH is called parametric PH model. In Cox PH model the coefficients are estimated by partial likelihood but in parametric PH model the Maximum likelihood method is used to estimate the parameters. By choosing different hazard function different parametric PH model may be derived such as Exponential, Weibull , Gompertz etc.. In this paper the researchers have considered three form of PH models which are explained in following…

(ii) Exponential PH model: In the Exponential PH model it is assumed that the hazard function is constant over time. The Survival and hazard function of Exponential model are

$$s(t) = \exp(-\lambda t)$$

And

$$h(t) = \lambda$$

Under exponential PH model, the hazard function is

$$h(t,x) = \lambda \exp(\sum_{i=1}^{p} \beta_i x_i)$$

Here hazard function follows Exponential distribution. That's why it is called Exponential PH model.

(iii) Weibull PH model: The Weibull PH model is the generalization of exponential distribution with parameter λ and shape parameter γ. So

$$s(t) = \exp(-\lambda t^{\gamma})$$

$$h(t) = \lambda\gamma(t)^{\gamma-1} \qquad \gamma, \lambda > 0$$

When $\gamma > 1$, the hazard rate increases.
When $\gamma < 1$ the hazard rate decreases and
When $\gamma = 1$, the hazard rate remains constant.

Under the Weibull PH model , the hazard function is

$$h(t,x) = \lambda\gamma(t)^{\gamma-1} \exp(\sum_{i=1}^{p} \beta_i x_i)$$

Here the Weibull PH model with scale parameter $\lambda\exp(\sum_{i=1}^{p} \beta_i x_i)$ and shape parameter γ.

The survival function of Weibull PH model is

$$s(t,x) = \exp\{-\exp(\sum_{i=1}^{p} \beta_i x_i)\lambda t^{\gamma}\}$$

(iv) Gompertz Proportional Hazard Model : The hazard function of the Gompertz distribution is given by

$$h(t) = \lambda e^{\theta t}, 0 \leq t \leq \infty \; and \; \lambda > 0$$

the survival function of the Gompertz distribution is given by

$$S(t) = \exp\left\{\frac{\lambda}{\theta}\left(1 - e^{\theta t}\right)\right\}$$

The density function is

$$f(t) = \lambda e^{\theta t} exp\left\{\frac{\lambda}{\theta}\left(1 - e^{\theta t}\right)\right\}$$

The Gompertz PH model model is given by

$$h(t, x) = \lambda e^{\theta t} \exp\left(\sum_{i=1}^{p} \beta_i \, x_i\right)$$

Various goodness of fit Test:

(i) **AIC:** To compare various semi-parametric and parametric models Akaike Information Criterion (AIC) is used. The AIC is proposed by Akaike (Akaike, 1974). It is a measure of goodness of fit of an estimated statistical model. For the model in this study, AIC is computed as follows

$$AIC = -2(log - likelihhod) + 2(P + K)$$

Where P is the number of parameters and K is the number of coefficients (excluding constant) in the model. For P=1, for the exponential, P=2, for Weibull, Log-logistic, Lognormal etc.The model which as smallest AIC value is considered as best fitted model.

(v) BIC: The Baysian Information Criteria (BIC) is given by Schwarz (Schwarz, 1978). It is computed as follows

$$BIC = -2(log - likelihood) + (P + K) * \log(n)$$

Where P is the number of parameters in the distribution, K is the number of coefficients and log(n) is the number of observations. The distribution which has the lowest BIC value is considered as best fitted model.

(vi) R^2: It is also a goodness of fit test. This statistic is calculated as follows:

$$R_p^2 = 1 - \{\exp[\frac{2}{n}(L_0 - L_p)]\}$$

Where L_p is the log-likelihood for the fitted model with p covariates and L_0 is the log likelihood for the model with no covariates.

On the basis of Akaki's Information Criteria (AIC), Baysian Information Criteria(BIC) and R^2 , the best fitted model is identified. After identifying the best fitted model the effect of different variables such as age, sex, location, socio-economic status, status of the patients ,stage of the patient and cancer directed treatments are used to fit the models on the survival for this esophagus cancer patient. Cox- Snell residuals (Cox and Snell, 1968) plot is used to fit the goodness of fit graphically. All the data are analyzed with the help of computer software package R version 3.3.1.

IV. Results

A total of 178 individuals diagnosed with esophagus cancer during the study period are included in the study. The average age of the patients are 59.13 years (s.d. 1.16 years). There are male preponderance in the sample with 67.4% and 32.6% are female. Among the patients about 27% are diagnosed at Distant stage while the stage at the time of diagnosis could not specify for 18% of the patients. The detail demographic, treatment and disease profile of the esophagus cancer patients are presented in the table I .

The Cox PH model and its parametric counterpart following Exponential, Weibull , Gompertz distribution are fitted with the same data set to assess the best fitted model which explains the survival data of esophagus cancer. The various explanatory variables consider are sex, location, socio-economic status, stage of the patients and cancer directed treatment.

Different statistical measures such as AIC, BIC and R^2 as discussed in the methodology section are estimated for the models on the consideration. Cox-Snell residual plots are also drawn for the fitted models. By observing this plots, one can have idea about the best fitted model. The values of AIC, BIC and R^2 are presented in table II . From table II , can be observed that AIC and BIC values of parametric PH assuming Exponential, Weibull and Gompertz distribution are more or less similar but for the Cox PH model these values are much less then the parametric counterparts. For the Cox PH model, the AIC and BIC values are 1207.51 and 1210.52 respectively. In case of R^2, all the fitted models registered more or less similar values. By observing the Cox-Snell residuals, also the Cox PH is found to be best fitted which is presented in Fig1. Thus we can conclude that the Cox PH model is the best fitted to studying the survival time of esophagus cancer data.

As Cox PH model is found to be the best fitted one, the interpretation about the effect of different independent variables on the survival times is made by using Cox PH model. The results are presented in table III. From the table it is seen that, the middle and higher socio-economic group have lower risk of dying than the lower socio- economic group. The patients undergo the cancer directed treatment other than surgery and the patients who have not taken any treatment experiencing a significantly higher risk of 1.60 times (95% C.I. 1.01 to 2.52) and 3.49 times(95% C.I. 1.82 to 6.70) respectively of dying than that of patients who undergo surgery and others treatment. The stage at the time of diagnosed is a prominent factor for better survival of esophagus cancer patients of Assam. The risk of dying among patients diagnose with Regional, Distant stage are 1.97 (95% C.I. 1.13 to 3.44) and 4.52 (95% C.I. 2.52 to 8.08) times more than that of patients diagnosed in Localized stage. The patients whose stage remain unknown at the time of diagnosis are experiencing a significantly higher role of 2.77 (95% C.I. 1.48 to 5.19). From the result presented in table III, it can be observed that residential status, sex, and age of the patients have no significant influence on the survival of esophagus cancer patients. Cancer directed treatment has a significant role.

V. Discussion And Conclusion

In this paper, attempt has been made to find the best fitted model for studying the survival time of esophagus cancer patients. To meet the objectives semi- parametric Cox PH model and parametric PH models following distributions Exponential, Weibull, Gompertz, are fitted to survival data of esophagus cancer patients collected from AMC, Dibrugarh , Assam. Different statistical measures such as AIC, BIC, R^2, Cox- Snell residuals plots are used to find the best fitted model. From the results, the Cox PH model is found to be the best fitted model. Researches conduct in the past shows that the best fitted model for studying survival data may vary with the objective under study.

Marazzi et al., (1998) in their study showed that none of the considered parametric models (including Lognormal, gamma, weibull) appreared to fit satisfactory to describe the length of stay data. Austin et al., (2002) concluded that the generalized liner model were better than the linear models fo predicting length of stay after CABG surgery.

Nardi and Scheme (2003) compared Cox PH and parametric models in three clinical trial studies mainly performed at Vienna University Medical School. They used Normal –deviate residuals (Nardi,1999) to verify the parametric model assumptions. Their study showed that Weibull model was superior to other parametric model.

Pourhoseingholi et al., (2007) compared Cox regression and Parametric models in the analysis of the patients with gastric carcinoma and found that lognormal model fitted better than other models . Revangard et al., (2011) compared Cox PH model with Parametric models(including exponential, weibull, gompertz, log-normal ,log-logistic and gamma) in the study of length of stay in Tehram and from AIC and Cox Snell residuals it showed that the gamma model was the best fitted model.

Vallinayagam et al.,(2014) compared some parametric models including exponential weibull, gompertz, log-normal and log- logistic for Breast cancer data. It was found that the lognormal model was fitted better than the other model.

The researcher fails to find any previous attempts to find the best fitted model for studying survival data of esophagus cancer patients by using this parametric PH model and Cox PH model.

In this study, the result shows that Cox PH model is better than the other models in case of explaining the survival esophagus cancer data. The factor, cancer directed treatments, has a significant role in case of survival of esophagus cancer patients. The patients who undergo the cancer directed treatment other than surgery has lower risk of dying than the patients who has underwent the treatment of surgery and its combinations. Socio- economic status has also play the significant role on the survival of esophagus cancer patients. With reference to lower socio economic status patients, the middle and higher socio economic patients has higher chances of survival. The stage at diagnosis of patients is also a responsible factor in case of survival of esophagus cancer patients. The probability of survival of a patient diagnose in early stage is significantly higher than patients diagnose in advance stages. From the observations it is found that, the age of the patients at the time of diagnosis has no significant impact on the esophagus cancer patients. Also, the patients belonging to both rural and urban area are experiencing more or less similar risk of time. The sex of the patients is also not found to be a significant factor which can influence the survival .

Appendix

Table I: Demographic, Treatment and Disease profile of the esophagus cancer patients

Characteristics	**Frequency(%)**
Location	
Rural	84 (47.2)
Urban	94 (52.8)
Sex	
Male	120 (67.4)
Female	58 (32.6)
Age	
Less than 50	33 (18.5)
50 to70	118 (66.3)
Above 70	27 (15.2)
Cancer Directed Treatment	
Surgery & others	49 (27.5)
Other than Surgery	106 (59.5)
No treatment	23 (12.9)
Socio- economic status	
Lower	24 (13.5)
Middle	133 (74.7)
Higher	21 (11.8)
Stage	
Localized	34 (19.1)
Regional	67 (37.6)
Distant	49 (27.5)
Unknown	28 (15.7)

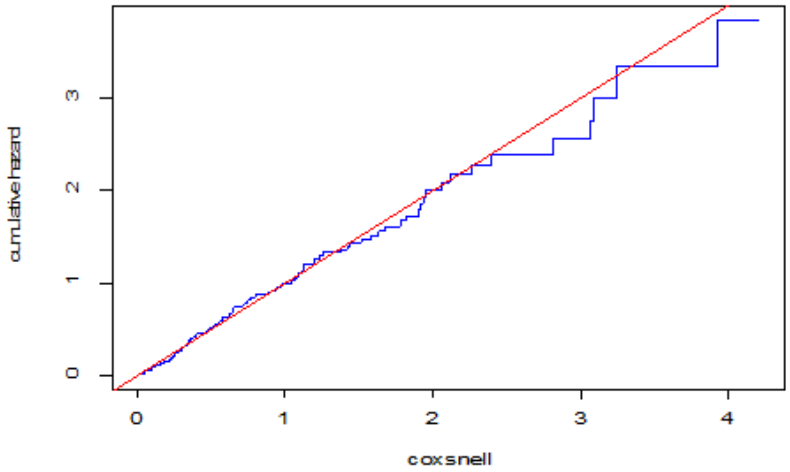
Table II : Goodness of fit of the Model on the basis of AIC and R^2

Models	**AIC**	**BIC**	R^2
Exponential	1928.06	1931.06	32.83%
Weibull	1920.67	1923.82	36.21%
Gompertz	1924.56	1930.06	34.86%
Cox PH	1207.51	1210.52	34.80%

Table III : Results of the Cox PH model

Characteristics	**Hazard Ratio(HR)**	**95% confidence Interval**
Location		
Rural	Reference	Reference
Urban	1.30	.91-1.86
Sex		
Male	Reference	Reference
Female	.95	.66-1.37
Age		
Less than 50	Reference	Reference
50 to70	.91	.56-1.45
Above 70	1.49	.82-2.68
Cancer Directed Treatment		
Surgery & others	Reference	Reference
Other than Surgery	1.60	1.01-2.52
No treatment	3.49	1.82-6.70
Socio- economic status		
Lower	Reference	Reference
Middle	.46	.28-.76
Higher	.32	.15-.69
Stage		
Localized	Reference	Reference
Regional	1.97	1.13-3.44
Distant	4.52	2.52-8.08
Unknown	2.77	1.48-5.19

Fig1: Cox Snell Residuals for Cox PH model



References

[1]. Akaike H.: A New Look at the Statistical Model Identification. *IEEE. Transaction and Automatic Control AC*-19. 1974, 716-23.

[2]. Austin PC, Rothwell DM, TU JV.:A Comparison of Statistical Modeling Stratigies for Analyzing Length of Stay after CABG surgery, *Health Services and Outcomes Research Methodology*, . 2002; 3(2):107-133.

[3]. Cox D.R., and Snell E.J.: A General Definition of Residuals (with discussion), *Journal of the Royal Statistical Society*, A. 1968.

[4]. Marazzi A, Paccaud F, Ruffieux C, Beguin C.:Fitting the Distributions of Length of Stay by Parametric Models, *Med Care*, 1988; 36(6)-915-927

[5]. Nardi, Alessandra and Schemper, Michael : Comparing Cox and Parametric models in Clinical Studies, *Statistics in Medicine, Statist Med.* 2003; 22:3597-3610.

[6]. Pourhoseingholi,M.A.,E. Hajizadeh,B.Moghimi Dehkordi, A. Safaee,A. Abadi and M.R. Zali: Comparing Cox regression and parametric models for survival of patients with gastric carcinoma, *Asian Pacific Journal of Cancer Prevention ,2007*; 8 (3):412-416

[7]. Ravangard, R., Arab M., Rashidian, A.,Akbarisari, A., Zare A., Zeraat, H. : Compared Cox Model and Parametric Models in the Study of Length of Stay in a Tertiary Teaching Hospital in Tehran, Iran. *Acta MedicaIranica*, 2011; 49(10): 650-658.

[8]. Schwarz, Gideon E. : Estimating the Dimension of a Model, *The Annals of Statistics*, vol 6, No.2, (1978),pp-461-469.

[9]. Vallinayagam, V., Prathasarathy, S., and Venkatesan, P.: Parametric Regression Models in the Analysis of Breast Cancer Survival Data , *International Journal of Science and* Technology, *2014* ;3(3) 2049-7318.