FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 7 No. 5, October, 2023, pp 181 - 184
DOI: https://doi.org/10.33003/fjs-2023-0705-2011

STUDY ON THE PERFORMANCE OF SOME PARAMETRIC PROPORTIONAL HAZARD MODELS AND SEMI-PARAMETRIC MODEL IN THE ANALYSIS OF BREAST CANCER DATA

[1]Zoramawa, A. B., [1]Usman, U., [2]Magaji, B. A., *[3]Rabiu, M.

[1]Department of Statistics Usmanu Danfodiyo University Sokoto State, Nigeria
[2]Department of Community Health Usmanu Danfodiyo University Sokoto State, Nigeria
[3]Department of Mathematics and Statistics Al-Qalam University Katsina State Nigeria

*Corresponding authors' email: muntarirabiu@auk.edu.ng

ABSTRACT
A study is conducted on medical records of 416 breast cancer patients. Analysis was performed using the R software version R3.6.3, and the level of significance was set at 0.05. The work employed three models which were based on Exponential, Weibull and Cox Regression models. The Weibull proportional model (AIC=1959.038) was the most appropriate model among the considered models, based on the Akaike information criterion (AIC). Results of the best fitted model showed that the survival time of breast cancer patients is significantly affected by age, age at diagnosis, and treatment taken at 95%.

Keywords: Semi-Parametric Model, Parametric Models, AIC, Breast Cancer

INTRODUCTION

Basically, there are two regression models used for survival data Cox (1972), semi-parametric model and parametric model. Efron (1977) as a result of less assumptions of Cox regression model, researchers most often prefer Cox regression model instead of parametric models but some reasons were made under certain situations, parametric models estimate parameters more efficient than semi-parametric model.

The analyses are often difficult when subjects under study refused to stayed in the trial, or when some of the subjects may not experience the event before the end of the study, though they would have experience the event, or lose touch with them in the course of study. The main difference between the normal conventional regression model and survival analysis is the presence of censoring. To reach an appropriate fit for parametric models, it is better that right-censorship does not exceed 40 to 50 percent Nardi *et al.*(2003 and Royston **(**2001).

MATERIALS AND METHODS

A recorded Breast Cancer patient's data set with thirteen covariates was employed for this study. The following are the covariates considered: Age, Menopause, Parity, Tumor Grade, Treatment Taken, Age at First Birth, Alcohol Consumption, Comorbidity, Smoking Status, Family History, Menarche, Age at Last Birth, Age at Diagnosis. The event of interest is survival time, and event and censoring are coded as 1 and 0 respectively.

The Cox proportional hard model

Faruk A (2018) the semi-parametric is used to measure the effects of covariates on the survival time, and it can be represented by the relationship of the hazard function, the baseline hazard function, and one or more covariates in the form

$h(t) = h_o(t)\, exp(\beta^t X)$

Where t is the survival time, h(t) is the hazard function, $h(t)$ is the hazard function, $h_o(t)$ is the baseline hazard function which is unspecified, β is a column vector of the regression coefficients, and X is a column vector of the covariates.

The semi-parametric model assumes that the hazard ratio for any two subjects in the population is constant over time. This feature is also known as the proportional hazard assumption and it can be defined as the ratio of the hazard functions for two individuals with different values of covariates X_1 and X_2. The hazard ratio is given by:

$$H(t) = \frac{h_o(t)\, exp(\beta_2 X_2)}{h_o(t)\, exp(\beta_1 X_1)} = \frac{exp(\beta_2 X_2)}{exp(\beta_1 X_1)} = exp(\beta^t(X_2 - X_1))$$

FarukA (2018) the hazard ratio $H(t) = exp(\beta^t(X_2 - X_1))$ is independent of time. It showed that the hazard ratio for any two individuals is constant over time. This property is also known as the proportional hazard assumption.

Weibull distribution

A random variable T has two-parameters with hazard, density and survivorship functions

$h(t, \lambda, \gamma) = \lambda\gamma t^{\gamma-1}$
$f(t, \lambda, \gamma) = \lambda\gamma t^{\gamma-1}\, exp(-\lambda t^{\gamma})$
$S(t, \lambda, \gamma) = exp(-\lambda t^{\gamma})$

Where $\lambda > 0$ and $\gamma > 0$ are the scale and shape parameters respectively.

The hazard function under the weibull proportional hazard model:

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

The survival function under weibull proportional hazard is:

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

It has increasing hazard if the shape parameter $\gamma > 1$ and decreasing hazard if $\gamma < 1$.

Exponential Distribution

A random variable T has the exponential distribution with the following hazard, density, and survivorship functions.

$h(t, \lambda) = \lambda$
$f(t, \lambda) = \lambda\, exp(-\lambda t)$
$S(t, \lambda) = exp(-\lambda t)$

The hazard function of exponential proportional hazard is:

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

Where $\lambda > 0$

Model Evaluation Using Akaike Information Criterion (AIC)

The Akaike information criteria (AIC) is a mathematical tool used to evaluate how well a model fits the data and to determine which model best fitted the data set by comparing the considered models. The Akaike information criterion (AIC) is computed from the covariates used to build the model and the maximum likelihood estimate of the model.

AIC=-2(log-likelihood)+2K

Where, K is the number of model parameters, and Log-likelihood is a measure of model fit.

RESULTS AND DISCUSSION

Results of Cox Proportional Hazard Model

Table 1: Analysis of Cox Proportional Hazard

Covariates	HR	se(β)	P
Age	0.8576	0.02851	7.14e-08
Parity	1.0210	0.03686	0.5709
Menopause	1.0210	0.03686	0.5748
Age at First Birth	1.0080	0.01774	0.6567
Age at Last Birth	1.0060	0.01254	0.6475
Age at Diagnosis	1.1680	0.02695	8.06e-09
Family History	1.0320	0.1858	0.8645
Alcohol Consumption	0.8981	0.1788	0.5480
Smoking Status	1.0990	0.3053	0.7574
Menarche	1.0580	0.04282	0.1896
Tumor Grade	1.0430	0.08611	0.6250
Treatment Taken	0.9004	0.05979	0.0792
Comorbidity	1.0240	0.1358	0.8602

Results from the Cox Proportional Hazard model presented in Table 1 indicates that Age (HR=0.8576, p-value=7.14e-08), and Age at diagnosis (HR=1.1680, p-value=8.06e-09) at 95% are significantly associated with the survival time of breast cancer patients. The age at diagnosis showed high risk of mortality.

Results of Weibull Proportional Hazard model

Table 2 Analysis of Weibull Model

Covariates	HR	se(β)	P
Age	1.2114	0.03229	2.9e-09
Parity	0.9659	0.04310	0.4211
Menopause	0.9976	0.00501	0.6300
Age at First Birth	0.9850	0.02065	0.4629
Age at Last Birth	0.9960	0.01464	0.7864
Age at Diagnosis	0.8242	0.03048	2.2e-10
Family History	0.9339	0.21793	0.7537
Alcohol Consumption	0.8436	0.21109	0.4204
Smoking Status	0.8585	0.35950	0.6714
Menarche	0.9407	0.04979	0.2194
Tumor Grade	0.9344	0.10092	0.5011
Treatment Taken	1.1509	0.06993	0.0445
Comorbidity	0.9893	0.15870	0.9461

In Table 2, the weibull regression model shows Age (HR=1.2114,p-value=2.9e-09), and Age at diagnosis (HR=0.8242, p-value=2.2e-10), and also Treatment taken (HR=1.1509, p-value=0.0445) at 95% are statistically significant, and age, age at diagnosis, and treatment taken indicated higher risk of mortality.

Results of Exponential Proportional Hazard model

Table 3: Analysis of Exponential

Covariates	HR	se(β)	P
Age	1.1977	0.02694	2.1e-11
Parity	0.9631	0.03709	0.31
Menopause	0.9973	0.00429	0.53
Age at First Birth	0.9817	0.01790	0.30
Age at Last Birth	0.9964	0.01263	0.78
Age at Diagnosis	0.8350	0.02525	9.1e-13
Family History	0.9102	0.18830	0.62
Alcohol Consumption	1.1935	0.18365	0.34
Smoking Status	0.8484	0.31063	0.60

Menarche	0.9432	-0.05852	0.17
Tumor Grade	0.9312	0.08770	0.42
Treatment Taken	1.1519	0.06070	0.02
Comorbidity	0.9784	0.13660	0.87

The result of exponential model in the Table 4 shows, age (HR=1.1977, p-value=2.1e-11), age at diagnosis (HR=0.8350, p-value=9.1e-13) and treatment taken (HR=1.1519, p-value=0.02) at 95% are significantly associated with the survival of patients. The hazard ratio of age and treatment taken indicated high risk of mortality.

Table 4: Comparison of Semi-parametric Model and the Parametric Proportional Hazard Models Using Akaike Information Criteria

MODELS	AIC VALUES
COX REGESSION MODEL	2522.257
WEIBULL MODEL	1959.038
EXPONENTIAL MODEL	1968.628

Results

The Cox regression model and proportional hazard models were fitted, from the table 2 and table 3 we see that three covariates namely age, age at diagnosis, and treatment taken are the most predicted covariates of the survival time of the breast cancer patients, while for Cox regression model, the result in table 1 showed that only two covariates were significantly associated with the survival time of the breast cancer patients. The considered models were compared by Akaike information criteria (AIC), the results in the table 4 showed that the semi-parametric and the considered parametric models have disparities in testing the significance of the covariates. The Akaike information criteria value of Cox regression model (AIC=2522.257), Weibull model (AIC=1959.038), and Exponential model (AIC=1968.628), in this regard, the whole assessment of the parametric models is far better than the Cox proportional hazard model, based on the Akaike information criteria the weibull model is the best fitted model among the models considered.

Discussion

Sharma *et al.* (2019), pointed out in their study to compare the efficiency of some accelerated failure time models (log-normal, exponential, log-logistic, and weibull) AIC was calculated and Weibull model found to be the best for the breast cancer data set, which is consistent with our findings where the weibull was the best fitted model among the models considered.

Vallinayagam *et al.* (2014) compared the performance of some parametric models including log-logistic, gompertz, exponential lognormal and weibull for Breast cancer data set. It indicated that the lognormal model was best fitted model more than other models based on deviance which is inconsistent with our study.

Hayat *et al.* (2010) compared five parametric models (Log-normal, Weibull, Log-logistic, Gamma, and Gompertz), the result revealed that age covariate was not significantly associated with the survival time of the breast cancer patients. The evaluation of the Akaike information criteria (AIC) showed that the Gompertz model was the best fitted using the Breast cancer data set from Ege University Cancer Research Centre. Their study was not in conformity with some other related works which indicated Age as a risk factor.

Magaji *et.al* (2017) compared the survival rates of colorectal cancer patients of some Asian countries, patients from the Chinese ethnic group had lower survival rates compared to their counterparts. The more advanced staging and late presentation were the most significant variables of colorectal cancer survival as obtained from Cox proportional hazard regression analysis.

Nuttawich *et al.* (2019) evaluate the performance of semi-parametric and two parametric models (Weibull and Log-logistic), the Cox regression model was the best model with the smallest value of Akaike information criteria using breast cancer data set, the composition of their models was limited to two parametric models.

Akintunde *et al.*(2019) studied the performance of some parametric models (Log-logistic, Exponential, Weibull, Log-normal), and Cox regression model, the Breast Cancer data set was used. The result showed that the log-normal model was the best fitted model when compared with the other models considered including the Cox regression model for a real life data set, but by using the simulated data set the results revealed that the exponential model was the best fitted model for a sample size 10 at low, moderate and high percentage censoring, while Cox regression model would be the best model for sample size 50, 100, and 500, at low, moderate and high censoring based on the Akaike information criteria (AIC).

CONCLUSION

Despites many survival models recommended by different researchers, with different methodologies in their works, this study has recommended the Weibull model as the best fitted model among the considered models for the breast cancer data set used, based on the Akaike information criteria (AIC). Many studies can be conducted as further research to evaluate the effect of covariates and the performance of the models by varying the sample sizes and the percentage of the censorship of the data set.

REFERENCES

Akintunde M.O., V. A. Micheal, J. O. Oyekunle& A. A. Agbona (2019), Comparison Of Cox Proportional Hazard And Parametric Models Of A Breast Cancer Data,*Pioneer Journal of Theoretical and Applied Statistics, 18(2) 1-14.http://www.pspchv.com/content-PJTAS.html*

Akaike H (1974), A New Look at the Statistical Model Identification.*IEEE. Transaction andAutomatic Control AC-19, 716-23.https://ieeexplore.ieee.org/document/1100705*

Alfensi Faruk (2018), The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data, Journal of Physics: Conference Series, 974 012008

Bello ArkillaMagaji, Foong Ming Moy, April Camilla Roslani& Chee Wei Law (2017), Survival rates and predictors of survivalamong colorectal cancer patients in

aMalaysian tertiary hospital. *BMC Cancer17:339, DOI 10.1186/s12885-017-3336-z*

Cox D.R (1972), Regression Models and Life-Tables *Journal of the Royal Statistical Society.SeriesB.34(2),187-220https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1972*

Efron B (1977**),**The Efficiency of Cox**'s** Likelihood Function for Censored Data. *Journal of the American Statistical Association, 72(359), 557-565. https://www.tandfonline.com/doi/abs/*10.1080/01621459.1977.10480613

Nardi A.& Schemper M. (2003), Comparing Cox and parametric models in clinical studies, *Statistics in Medicine, 22(1), 3597–3610 DOI: 10.1002/sim.1592*

Nuttawich Thongphet, PhisanuChiawkhun, Walaithip Bunyatisai, & ImjaiChitapanarux (2019), Comparison of Cox Regression and Parametric Models for Survival of Breast Cancer Patients with 1-3 Positive Lymph Nodes, *Burapha Science Journal 24(1) 190-200, Science.buu.ac.th/ojs246/index.php/sci/article/view/2468/0*

Royston **(**2001)**,** The Lognormal Distribution as a Model for Survival Time in Cancer, With an Emphasis on Prognostic Factors, *StatisticaNeerlandica, Netherlands Society for Statistics and Operations Research, 55(1), 89-104.https://ideas.repec.org/a/bla/stanee/v55* y2001i1p89-104.html

V. Vallinayagam, S. Prathap, &P. Venkatesan (2014), Parametric Regression Models in the Analysis of Breast Cancer Survival Data, *International Journal of Science and Technology 3(3) 163-167.https://www.researchgate.net/publication/296265075...*

Zelen M., Mary C.& Dannemiller **(**1961**),** The Robustness of Life Testing Procedures Derived from the Exponential Distribution**.** Technometrics***:3(1), 29-49https://www.tandfonline.com/doi/abs/10.1080/00401706.1961.10489925***