International Journal of Innovative Mathematics, Statistics & Energy Policies
12(1):1-8, Jan.-Mar., 2024

 www.seahipaj.org **ISSN: 2467-852X**

A Comparative Study Of Some Survival Models On The Analysis Of Prostrate Cancer Data In Dalhatu Araf Specialist Hospital Lafia, Nasarawa State, Nigeria

Attah Akuembo Samuel, Ahmad Abdulkaldir & Rashid Bello

Department of Mathematical Sciences,
Faculty of Sciences, Abubakar Tafawa Balewa University, Bauchi, Nigeria
meetsuccessfulsamuel@gmail.com; ahmed4statistics@gmail.com; arasheed@gmail.com

ABSTRACT
This research study is to carry out a comparative study of some survival models on the analysis of prostrate cancer data in Dalhatu Araf Specialist Hospital Lafia. The research work aim at identifying significant factors that are associated with survival of prostate cancer patients and to find out if any of our models (parametric and semiparametric model) outperform the other. A secondary data from the medical data sets on the demographic and clinical history of 27 prostate cancer patients was collected from Dalhatu Araf Specialist Hospital from 2020 to 2022. This study used the maximum likelihood estimation (MLE) method in estimating the parameters of the survivor functions and all the data analysis were performed using STATA 13 software. Cox regression and some selected parametric models (Exponential, Weibull and Gompertz) were utilized to detect factors influencing survival time of patients. The models comparison were made with Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The finding on the Exponential model shows that age, stage and grade are the variables influencing the survival time of patients with prostate cancer. Results according to the Weibull model shows that age, education status, alcohol status and smoking habit status are the variables influencing the survival time of patients with prostate cancer. Finding from the Gompertz model shows that age, education status, family history, stage and grade are the variables influencing the survival time of patients with prostate cancer. Results on the Cox model shows that age, education status, alcohol status, smoking, stage and grade are the variables influencing the survival time of patients with prostate cancer. According our results, the Weibull model (with the smallest AIC = 21.257; BIC=13.482) provided a better fit and the most efficient model than the Exponential model, Gompertz model and Cox model. Lastly, the results suggested that the parametric regression models demonstrate more reliable, interpretable and a better-fit survival models than the semiparametric model (Cox model) for the study of patients with prostate cancer than the (Cox model).
Keywords: Cox model, Gompertz model, Exponential mode, Weibull model, Prostate cancer and Survival analysis.

INTRODUCTION
The purpose of data analysis is to find out useful information in order to make reasonable decisions and suggest conclusions in those various studies and researches. There are many methods for data analysis, like the study of an event that happen during the time such as time series or survival analysis etc. One of

the methods of data analysis is called survival analysis and will be used throughout this research work, to which much attention was given in the research field for several decades (Sam and Krongs, 2008; Kargarian-Marvasti et al., 2017; Ahmedi et al., 2020). Survival analysis aims to estimate the three survival (survivorship, density, and hazard) functions, denoted by S(t), f(t) and h(t), respectively. The survival function S(t) gives the probability of surpassing a given time t without an event occurring, and it's the inverse of the cumulative distribution function, F(t) (Ryan Wu, 2021; Hong, 2017). The hazard function h(t) signifies the immediate probability per unit of time for the event to take place, given that the individual has survived up to time t". Survival analysis was designed for study of death, but it has now been extended to time to event, where event here includes: contacting a disease, equipment failures, earthquakes, automobile accidents, stock market crashes, job terminations, births, marriages, divorces (Xian 2012; Kartsonaki, 2016).

These techniques include parametric method such as; Weibull, exponential, log-logistic, log-normal, and Gompertz given by (George et al., 2014; Ahmedi, 2020), non-parametric method such as; Kaplan Meier, Nelson-Alan, life table, and semiparametric method such as; Cox Proportional hazard model (Abbas et al., 2015; Ahmedi, 2020). The parametric models are in high demand because of their predictive power while the semi-parametric and non-parametric models are used for survival time analysis due to their flexibility (Xian 2012; Kartsonaki, 2016; Kim, HeeJin 2022).

Statement of the Problem

The application of statistics in medicine and health science relies heavily on the modeling of survival data. There is uncertainty about the prevalence estimation of prostate cancer data because of the rarity of diagnosis and a lack of high quality studies. Several researchers have used the Cox model (semi-parametric model), non-parametric models and the parametric models to model survival time in diseases such as acute leukemia, liver cirrhosis, breast cancer, lung cancer, kidney transplant, and so on (Saeed and Seyyed, 2017; Danial et al., 2018). Thus, previous researches (Ahad Alizadeh et al., 2013; Danial et al., 2018) have not fully investigate the effects of the explanatory variable (covariates) on the survival models and the survival function. Hence, there is need to investigate the best survival model between parametric models and semi-parametric model (Cox model) using criterion and models that has not been fully explored. In addition, there is need to increase the number of attributed variables that are significant predictors of survival time such as education status, alcoholic status and smoking abuse status that may improve the performance of the survival models in order to understand the characteristics of health behaviors associated with survivorship for prostate cancer patients.

In view of this, the research conducted a comparative analysis of several models to examine prostrate cancer data in Dalhatu Araf Specialist Hospital Lafia. The aim was to determine the most suitable model using maximum likelihood estimation, comparing parametric semi parametric methods. The investigation involved assessing the prerequisites for estimating model parameters and exploring the impact of demographic and clinical information on patient survival time.

Research Aim and Objectives

The aim of this research study is to carry out a comparative study of some survival models on the analysis of prostrate cancer data in Dalhatu Araf Specialist Hospital Lafia, Nasarawa State. The desired aim will be achieved through the following objectives:

1. To fit some selected parametric methods and a semi-parametric method (Cox model), in order to render it suitable with the prostate cancer patient's data.
2. To estimate the model parameters for the survivor and hazard functions using the Maximum Likelihood Estimation (MLE).
3. To investigate the effects of the explanatory variable (demographic, clinical and pathology factors) in the survival time of prostate cancer patients.
4. Comparing different survival models Semiparametric model (Cox model) and some selected parametric models (Exponential, Weibull, and Gompertz model) to determine the best fitting model of prostate cancer data in DASH, Lafia using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values.

RESEARCH METHODOLOGY

We carried out a comparative study of some survival models on the analysis of prostrate cancer data in specialist hospital, Lafia. The method of maximum likelihood was used to estimate the model parameters in order to choose the best fit model among some selected parametric models and semiparametric model, we employed the use of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Kassambara, 2018) for this.

ANALYSIS RESULTS AND DISCUSSION

The researcher used STATA 13 software to carry out an analysis on a secondary data consisting of 27 prostate cancer medical records or history of patients from year 2020 to year 2022.

Proportionality Assumption

Before using the Cox regression model, PH assumption for each of the variables were investigated on how they satisfy the proportional hazard assumptions. This provides the basis for the selection of suitable variables to be included in the model. PH assumption need to be tested using either Graphical test, Time-dependent covariates test or Goodness of fit (Schoenfeld residual).

Overall Performance of the Cox Model

The Cox-Snell residual plot is a useful diagnostic tool for assessing the fit of a survival model and can provide insights on how well the model aligns with the observed survival data. The plot has displays the cumulative hazard residuals against the expected values of the survival model. The residuals should follow a straight line at a 45-degree angle. Deviations from this line suggest departures from the assumed model. The Cox-Snell residuals were plotted for eight categorical variables with time recorded in days. The categorical variables considered for the study were Educational Status, Marital Status, Alcohol Habit Status, Smoking Status, Family history, Clinical stage, Grade and Status of prostate cancer.

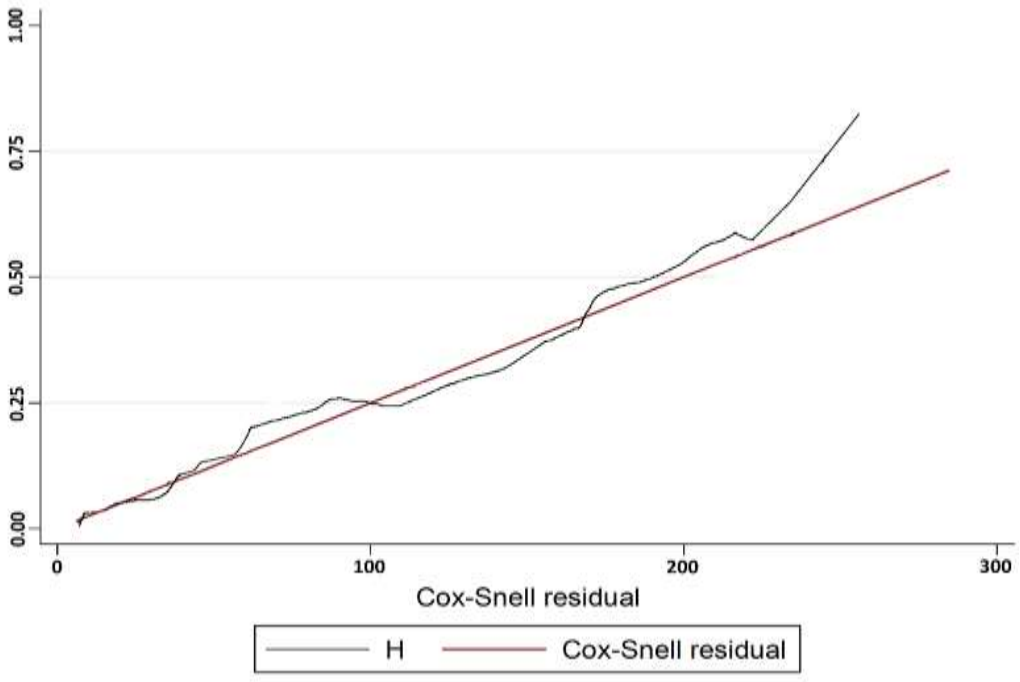

Figure 1: A Cox-Snell residual plot

The plot from figure 1 shows the Cox-Snell residuals (together with their cumulative hazard function) obtained from fitting semiparametric (Cox) model to the survival data via maximum likelihood

estimation. Results shows the lines corresponding to the Cox-Snell residuals of the Cox model are closest to the line through the origin, indicating that these model fit the data well. The Cox model (which assume proportional hazards) do seem to fit the data well, indicating that the proportional hazards assumption is not violated.

PH assumption

A global test was performed for the whole model, which looks at the combination of all these covariates in the survival model. Variables, which are not significant at 0.05 are said to have satisfied the PH assumptions.

Table 2: Proportional Hazard Assumption Test for the Covariates

Variables	Chisq	Df	P
Age	0.562	1	0.247
Education	0.328	1	0.594
Marital Status	1.212	1	0.152
Alcohol Status	1.086	1	0.354
Smoking Status	0.567	1	0.487
Family History	2.410	1	0.410
Stage	3.432	1	0.235
Grade	1.231	1	0.141
Global Test	17.291	9	0.301

Table 2 presents the results for the test of proportionality assumption for the covariates. The proportional hazards models assume that the hazard ratio is proportional over time. It can be observed that the result of Schoenfeld test showed that none of the variables violated the PH assumption ($p>0.05$). Therefore, we could use the Cox proportional hazard model for this study.

The Estimation of Factors associated with the Survival Models

The covariates parameters were estimated using the maximum likelihood estimator to understand the characteristics of health behaviors associated with survivorship for prostate cancer patients. To understand this better, a Hazard Ratio was estimated and P-value are used. If the hazard ratio is greater than one, it is an indication that the risk of getting the event increases as the survival rate of the event decreases. The results of the findings were presented in the Table 3:

Table 3: The Effects of the Explanatory Variables on the Survival Time

Model	Variables	Age	Edu	Mar	Alcoh	Smok	FamH	Stage	Grade
Exponential	HR (SE)	1.260	0.396	1.401	1.536	1.642	2.294	1.256	1.044
		(0.294)	(0.356)	(0.621)	(0.144)	(0.247)	(0.558)	(1.876)	(0.502)
	P-Value	0.026	0.054	0.743	0.338	0.176	0.410	0.024	0.000
Weibull	HR (SE)	1.831	0.54	1.051	1.043	1.059	1.076	1.217	0.83
		(0.167)	(0.174)	(0.154)	(0.016)	(0.034)	(0.071)	(0.820)	(0.275)
	P-Value	0.041	0.030	0.352	0.039	0.042	0.242	0.141	0.501
Gompertz	HR (SE)	1.670	0.720	1031	2.041	1.912	1.074	1.042	1.181
		(0.264)	(0.015)	(0.052)	(0.012)	(0.647)	(0.067)	(0.798)	(0.296)
	P-Value	0.001	0.009	0.268	0.514	0.471	0.025	0.036	0.020
Cox	HR (SE)	1.001	0.805	1.206	1.271	1.732	1.470	1.330	1.502
		(0.049)	(0.680)	(0.690)	(0.848)	(1.829)	(0.668)	(0.968)	(0.395)
	P-Value	0.027	0.036	0.409	0.017	0.139	0.426	0.032	0.041

a. Hazard Ratio (HR); b. Standard Error (SE); c. The P-value;

After entering the variables in survival function, the results from table 3 on the Exponential model shows that age (P=0.026), stage (P=0.024) and grade (P=0.000) are the variables influencing the survival time of patients with prostate cancer. Results according to the Weibull model shows that age (P=0.041), education status (P=0.030), alcohol status (P=0.039) and smoking habit status (P=0.042) are the variables influencing the survival time of patients with prostate cancer. Findings on results from the Gompertz model shows that age (P=0.001), education status (P=0.009), family history (P=0.025), stage (P=0.036) and grade (P=0.020) are the variables influencing the survival time of patients with prostate cancer. Looking carefully at the results again, the Cox model shows that age (P=0.027), education status (P=0.036), alcohol status (P=0.017), smoking, stage (P=0.032) and grade (P=0.041) are the variables influencing the survival time of patients with prostate cancer.

According to results on Exponential model in table 3 shows the estimate of HR for age (HR =1.260), clinical stage (1.260) and grade (1.044) are all greater than one suggesting an increase in hazard as the survival time of patients with prostate cancer decreases since their P-values are significant. Finding based on the Weibull model in table 3 shows the estimate of HR for age (HR=1.831), alcohol status (HR=1.043), smoking habit status (HR=1.059), are all greater than one suggesting an increase in hazard as the survival time decreases since their P-values are significant. Thus, the hazard rate on education status (HR=0.501) is less than one suggesting a decrease in hazard rate as the survival time of prostate cancer patients increases for the Weibull model since the P-values is significant. Again, results based on the Gompertz model in table 3 shows the estimate of HR for; age (HR=1.670), family cancer history (HR=1.074), clinical stage (HR=1.042) and grade (HR=1.181) are all greater than one suggesting an increase in hazard as the survival time of prostate cancer patients decreases given their P-values are significant. Thus, hazard rate on education status (HR=0.720) is less than one suggesting a decrease in hazard rate as the survival time of prostate cancer patients for the Gompertz model an increase since the P-value is significant. Lastly, results based on the Cox model in table 3 shows the estimate of HR for; age (HR=1.001), alcohol status (HR=1.271), smoking habit status (HR=1.732), clinical stage (HR=1.330) and grade (HR=1.502) are all greater than one suggesting an increase in hazard. This will bring about decreases in the survival time of prostate cancer patients since their P-values are significant. Thus, hazard rate on education status (HR=0.805) is less than one suggesting a decrease in hazard rate as the survival time of prostate cancer patients increase for the Gompertz model since the P-value is significant.

The Models Comparison

To identify the best model between some selected parametric and semiparametric model (Cox model), the likelihood ratios, AIC and BIC values are used to select the best-fitting model. The best model will have the lowest value for AIC and BIC in table 4:

Table 4: Model Summary Statistics Table

Type of Model	Model	Obs	Df	AIC	BIC	-2log(likelihood)
Semiparametric	Cox	27	9	145.436	157.099	-63.718
Parametric(PH)	Exponential	27	10	30.000	42.958	-5.000
	Weibull	27	6	21.257	13.482	-16.628
	Gompertz	27	11	23.580	29.320	22.787

Results from the table 4, each models were compared with each other, shedding light on the determination of the best-fitted model under various situations. It can be seen that the Weibull model (AIC = 21.257 and BIC =13.482) has the smallest value of AIC and BIC respectively than Exponential model (AIC=30.000; BIC=42.958.), Gompertz model (AIC= 23.580; BIC= 29.320) and Cox model (AIC= 145.436; BIC=157.099). This implies that Weibull model provided a better fit to our data than the other models in this study. Furthermore, results shows that parametric regression models demonstrated more reliable, interpretable and a better fit to survival model for prostate cancer patients' data compared to the Cox model as shown in the results obtained because they all have smaller AIC and BIC than semiparametric model.

Decision on Research Hypothesis 1

Results shows that combination of our joint covariates on Cox model (p=0.040), Exponential model (p=0.043), Weibull model (p=0.000) and Gompertz model (p=0.000) have significant effects on survival time of prostate cancer patients. This is also an indicative of the joint statistical significance of the overall covariate on survival time of prostate cancer patients was significant and therefore a reliable indicator of the study findings at the combination of all these covariates in the model. From our results, we accept our null hypothesis ( ) which stated that the explanatory variable (covariates) have effects on the survival function to recovery of patients with prostate cancer. Thus, we reject the alternate hypothesis ( ), we stated that the explanatory variable (covariates) do not have influence or effects on the survival models to recovery of patients with prostate cancer.

Decision on Research Hypothesis 2

Our finding from Table 4 revealed that the parametric regression models demonstrate more reliable, interpretable survival models and a better fit for survival data in prostate cancer patients compared to the Cox model as shown in the results obtained. Therefore, we reject the null hypothesis) that says the semi parametric model (Cox model) provides more reliable survival models, interpretable survival models and a better fit for survival data in prostate cancer patients and accept the alternate hypothesis (which says the parametric models provides more reliable survival models, interpretable survival models and a better fit for prostate cancer data.

SUMMARY

This research study is to carry out an analysis of Cox Proportional Hazard Model and some parametric models in the study of prostate cancer patient in DASH Lafia, Nasarawa State. The research work aim at identifying significant factors that are associated with survival of prostate cancer patients and to find out if any of our models (parametric and semiparametric model) outperform the other. A secondary data sourced from medical data on the Demographic and clinical history of 27 prostate cancer patients was collected from Dalhatu Araf Specialist Hospital from 2020 to 2022. The reviewed of existing research works and the method adopted for model selection were also reviewed. We adopted the Maximum Likelihood Estimation (MLE) in estimating the parameters of the survivor functions by obtaining of the models given. The used of AIC and BIC are employed in choosing our best fit model.

Furthermore, the plot from figure 1 shows the lines corresponding to the Cox-Snell residuals of the Cox model are closest to the line through the origin, indicating that these model fit the data well. The Cox model (which assume proportional hazards) do seem to fit the data well, indicating that the proportional hazards assumption is not violated.

The results from table 3 on the Exponential model shows that age, stage and grade are the variables influencing the survival time of patients with prostate cancer. Results according to the Weibull model shows that age, education status, alcohol status and smoking habit status are the variables influencing the survival time of patients with prostate cancer. Findings on results from the Gompertz model shows that age, education status, family history, stage and grade are the variables influencing the survival time of patients with prostate cancer. Looking carefully at the results again, the Cox model shows that age, education status, alcohol status, smoking, stage and grade are the variables influencing the survival time of patients with prostate cancer.

Finding based on the Exponential model in table 3 shows the estimate of HR for age, clinical stage and grade are all greater than one suggesting an increase in hazard as the survival time of patients with prostate cancer decreases since their P-values are significant. Finding based on the Weibull model in table 3 shows the estimate of HR for age, alcohol status, smoking habit status are all greater than one suggesting an increase in hazard as the survival time decreases since their P-values are significant. Thus, the hazard rate on education status (HR=0.501) is less than one suggesting a decrease in hazard rate as the survival time of prostate cancer patients increases for the Weibull model since the P-values is significant. Again, results based on the Gompertz model in table 3 shows the estimate of HR for; age, family cancer history, clinical stage and grade are all greater than one suggesting an increase in hazard as the survival

time of prostate cancer patients decreases given their P-values are significant. Thus, hazard rate on education status (HR=0.720) is less than one suggesting a decrease in hazard rate as the survival time of prostate cancer patients for the Gompertz model an increase since the P-value is significant. Lastly, results based on the Cox model in table 3 shows the estimate of HR for; age, alcohol status, smoking habit status, clinical stage and grade are all greater than one suggesting an increase in hazard. This will bring about decreases in the survival time of prostate cancer patients since their P-values are significant. Thus, hazard rate on education status (HR=0.805) is less than one suggesting a decrease in hazard rate as the survival time of prostate cancer patients increase for the Gompertz model since the P-value is significant.
Our findings from table 4 shows that Weibull model (AIC = 21.257; BIC=13.482) will provide a better-fit model to our data than the Exponential model, Gompertz model and Cox model. Furthermore, results shows that the parametric regression models demonstrate more reliable, interpretable and a better-fit survival models than the Cox model for the study of patients with prostate cancer as obtained in table 4.

CONCLUSION

Emanating from the findings, we found out that the combination of our joint covariates on Cox model (p=0.040), Exponential model (p=0.043), Weibull model (p=0.000) and Gompertz model (p=0.000) have significant effects on survival time of prostate cancer patients. We may conclude that our model are effective, flexible and fits well to the survival model of patients with prostate cancer. Results shows that the hazard rate on education status (for the cases of Weibull and Gompertz model) is less than one suggesting a decrease in hazard rate as the survival time of prostate cancer patients increases since the P-value is significant.
We concluded that the Weibull model with the smallest (AIC = 21.257; BIC=13.482) provided a better fit and the most efficient model to be used than the Exponential model, Gompertz model and Cox model. In summary, our findings suggest and affirm that parametric regression models offer greater reliability, interpretability, and a more suitable fit for survival analysis in the study of prostate cancer patients compared to the semiparametric model (Cox model).

REFERENCES

A. Kassambara (2018). Regression model validation. Regression model accuracy metrics: R-square, AIC, BIC, CP and more.

Abdul-Fatawu Majeed, (2020). Accelerated failure time models: An application in insurance attrition. The Journal of Risk Management and Insurance Vol. 24 No. 2. Universit_e de Pau et des Pays de l'Adour, Pau, France.

Ahad Alizadeh,Reza Ali Mohammadpour, Mohammad Reza Barzega, (2013). Comparing cox model and parametric models in estimating the survival rate of patients with prostate cancer on radiation therapy. J Mazandaran Univ Med Sci; 23 (100):21-29 http://jmums.mazums.ac.ir/article-1-2170-en.html

Cox, D. R (1972). Regression Models and Life Tables (with Discussion). Journal of Royal Statistics Society: Series B (Methodology) 34(2):187–202.

Danial Habibi, Mohammad Rafiei1, Ali Chehrei, Zahra Shayan, Soheil Tafaqodi1, (2018). Comparison of survival models for analyzing prognostic factors in gastric cancer patients. Asian Pac J Cancer Prev, 19(3), 749-753.DOI:10.22034/APJCP.2018.19.3.749

Datwyler C. and Stucki T., (2011). Parametric survival models. Retrieved from https://stat.ethz.ch/education/semesters/ss2011/seminar/contents/handout.9.pdf

Felix Noyanim Nwobi and Chukwudi Anderson Ugomma. (2014). A Comparison of methods for the estimation of Weibull distribution parameters. Metodoloskizvezki, 11(1):65.

Frank E Harrell Jr., (2015). Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis. Springer.

George, B., Seals, S, & Aban I., (2014). Survival analysis and regression models. Journal of Nuclear Cardiology, 21(4).

Hong S, (2017). Evaluation of goodness-of-fit tests for the cox proportional hazards model with time-varying covariates. (Master's thesis). Retrieved from https:// scholarcommons.sc.edu/etd/4302.

Kartsonaki, C. (2016). Survival analysis. Diagnostic histopathology 22 (7), 263–270.

Kim, HeeJin, Sunghun Kim, and Eunjee Lee. (2022). Cox proportional hazards regression for interval censored data with an application to college entrance and parental job loss. Economies 10: 218. https://doi.org/10.3390/economies 10090218.

Nasejje Justine, (2013). Application of survival analysis methods to study under-five child mortality in Uganda. A thesis of department of statistics and computer science, University of KwaZulu-Natal.

Patricia A. Laryea, (2015). Survival analysis of the average time to handling a claim in the insurance industry: A case study of an automobile insurance company in Ghana. A Thesis of Mathematics Department, Kwame Nkrumah University of Science and Technology.

Ryan Wu, (2021). A contribution to variable selection for the cox proportional hazards model with high-dimensional predictors. A thesis of master of science in the department of statistics and data science.

S. A. ElHafeez et al., (2021). Methods to analyze time-to-event data: The cox regression analysis.

Saeed Hosseini Teshnizi and Seyyed Mohammad Taghi Ayatollahi., (2017). Comparison of cox regression and parametric models: Application for assessment of survival of pediatric cases of acute leukemia in Southern Iran. Asian Paciific Journal of Cancer Prevention, 18 (4), 981-985.DOI:10.22034/APJCP. 2017.18. 4.981

Sam and Krongs (2008), Survival analysis approach to reliability, survivability and prognostics and health management (PHM). IEEE Aerospace Conference.

Shaikha Ahmedi, Faiz A M Elfaki, Iing Lukman & N A Kabbashi (2020). Cox's model for prison partly interval censored data. Conf. Series: Journal of Physics: Conf. Series 1489 (2020) 012032 IOP Publishing doi:10.1088/1742-6596/1489/1/012032.

Xian Liu (2012).Survival Analysis models and applications. First edition, John Wiley & Sons Ltd, the Atrium, Southern Gate, Chichester, United Kingdom.