
Comparative Analysis of the Cox Semi-parametric and Weibull Parametric Models on Colorectal Cancer Data

Usman Umar^{1,*}, Marafa Haliru Muhammad^{1,2}

¹Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria

²Planning Division, Usmanu Danfodiyo University Teaching Hospital, Sokoto, Nigeria

Email address:

Uusman07@gmail.com (U. Umar), halirumarafa5@gmail.com (M. H. Muhammad)

*Corresponding author

To cite this article:

Usman Umar, Marafa Haliru Muhammad. Comparative Analysis of the Cox Semi-parametric and Weibull Parametric Models on Colorectal Cancer Data. *International Journal of Data Science and Analysis*. Vol. 6, No. 1, 2020, pp. 41-47. doi: 10.11648/j.ijdsa.20200601.15

Received: January 26, 2020; **Accepted:** February 13, 2020; **Published:** March 17, 2020

Abstract: The survival and hazard functions are key concepts in survival analysis for describing the distribution of event times. The survival function gives, for every time, the probability of surviving (or not experiencing the event). The hazard function gives the potential that the event will occur, per time unit, given that an individual has survived up to the specified time. While these are often of direct interest, many other quantities of interest (e.g., median survival) may subsequently be estimated from knowing either the hazard or survival function. This research was a five-year retrospective study on data from a record of colorectal cancer patients that received treatments from 2013 to 2017 in Radiotherapy Department of Usmanu Danfodiyo University Teaching Hospital, Sokoto, being it one of the cancer registries in Nigeria. 9 covariates were selected to fit colorectal cancer data using Cox and Weibull Regression Models. From the result it is concluded that the predictor variables could significantly predict the survival of colorectal cancer patients using Cox. Also the result of the Weibull Proportional Hazard Model shows that the model is adequate enough to predict the survival of the colorectal patients. The A. I. C result shows that, according to our colorectal cancer data, the semi-parametric Cox regression model performed better than the parametric Weibull proportional hazards model. However, in the present study, the Cox model provided an efficient and a better fit to the study data than Weibull model.

Keywords: Survival, Cancer, Cox, Hazards, Weibull, Model

1. Introduction

Colorectal cancer (CRC) is a tumour of the colon and rectum. Most cases of CRC are sporadic; meaning there are no known hereditary (genetic) components, and it develops slowly over several years through adenomatous polyps (Brenner *et al.*, [1]). Changes in bowel habits, blood in the stool, and anaemia are cardinal symptoms and signs of CRC. In later stages, fatigue, anorexia, weight loss, pain, jaundice, and other signs and symptoms of locally advanced and metastatic disease occur. CRC is traditionally diagnosed by sigmoidoscopy and colonoscopy using biopsy. There are several ways to treat colorectal cancer depending on the cancer stage and where the tumour is localized. The main treatment is surgery; however, chemotherapy and radiation therapy can also use (Potter & Hunter, [2]).

Approximately 1.4 million new cases of colorectal cancer and almost 700 000 deaths occurred worldwide in 2012 (Arnold *et al.*, [3]). Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. The event can be death, occurrence of a disease, marriage, divorce, etc. The time to event or survival time can be measured in days, weeks, years, etc. For example, if the event of interest is death, then the survival time can be the time in years until a person dies (Hosmer, Lemeshow, and May, [4]).

According to Hosmer *et al.*, [4] Observations are called censored when the information about their survival time is incomplete; the most commonly encountered form is right censoring. Censoring is an important issue in survival analysis, representing a particular type of missing data. Censoring that is random and non-informative is usually

required in order to avoid bias in a survival analysis.

The survival and hazard functions are key concepts in survival analysis for describing the distribution of event times. The survival function gives, for every time, the probability of surviving (or not experiencing the event) up to that time. The hazard function gives the potential that the event will occur, per time unit, given that an individual has survived up to the specified time. While these are often of direct interest, many other quantities of interest (e.g., median survival) may subsequently be estimated from knowing either the hazard or survival function (Hosmer *et al.*, [4]).

Many countries today have population-based cancer registries. Their task is to collect and store information on all cases of cancer in the country and produce statistics of the incidence of cancer, and the survival of cancer patients. They play an important role in analysing the impact of cancer in the community. In Nigeria, for example, there are ten (10) population-based cancer registries owned by the Federal Government located at various tertiary hospitals across the country, according to Nigerian National System of Cancer Registries (NSCR, [5]). In most part of Africa, cancer burden is under reported due to lack of or inaccurate population statistics, which makes age specific incidence rate impossible or inaccurate (Abdulkareem, [6]).

This study was to estimate the population based colorectal cancer survival analysis using Cox and Weibull regression models, in order to ascertain the one that better fits colorectal cancer data in population-based research. The specific objectives were to:

Describe the survival function using Kaplan-Meier (K-M) approach, and then compare the survival curves using Log-rank tests. Fit the two models used in the survival analysis using data on colorectal cancer. Test for Cox Proportional Hazards assumptions using both the statistical test and graphical method. Estimate the survival function, hazard function, using Cox and Weibull Proportional Hazards Models and the effect of the covariates on patients from the colorectal cancer data collected. Ascertain performance, efficiency and flexibility of the models using AIC test.

The origin of survival was traced back to World War II (Dickman and Hakulinen, [7]). Survival analysis is a series of procedure that analyse timing of events (Kleinbaum and Klein, [8]). Dickman [9] listed some developments in statistical methods for population-based cancer survival analysis.

Adejumo and Ahmadu [10] disclosed that the shape parameter of the weibull model does not depend on the performance of cox proportional hazard model. Quantin *et al.*, [11] compared statistical models in terms of performance of different regression models for censored survival data in modeling the impact of prognostic factors on all-cause mortality in colon cancer. Similarly, Ahmad *et al.*, [12] sees the Cox

Model as multivariate Semi-parametric Regression Model in regarding Colon cancer.

Abdulkabir *et al.* [13] point out that the shape parameter of the Weibull Model does not depend or affect the performance

of the Cox Proportional Hazard Model. He further asserts that both models perform similarly when the distributional assumptions are not met except when sample size is small and the Weibull Model out-perform the Cox model when the distributional assumption are met and the shape parameter is known.

According to Wang *et al.* [14] Weibull Distribution is the best model for survival analysis of genetic associations in HNFIB of cancer patients when Cox proportional hazard and parametric models are compared.

The leading cause of death and disabilities worldwide is cancer which affects more than 14 million people annually (W. H. O., [15]). Knut *et al.* [16] consider colorectal cancer (CRC) as a complex disease that almost 40% of the surgically cured patients experience cancer recurrence within 5 years. Cancer control refers to all actions taken to reduce the frequency and impact of cancer (Armstrong, [17]).

Zaki [18] found a general formula for generating survival data on the computer through the fundamental relation between hazard rate and survival function.

Yuan [19] assumed an exponential form for the baseline hazard function and combined Cox Proportional Hazard Regression for the survival study of a group of lung cancer patients. The covariates in the hazard function are estimated by maximum likelihood estimation following the proportional hazards regression analysis. Although the proportional hazards model does not give an explicit baseline hazard function, the baseline hazard function can be estimated by fitting the data with a nonlinear least square technique.

Nigeria contributed 15% to the estimated 681,000 new cases of cancer that occurred in Africa in 2008 (Sylla, [20]). Similar to the situation in the rest of the developing world, a significant proportion of the increase in incidence of cancer in Nigeria is due to increasing life expectancy, reduced risk of death from infectious diseases, increasing prevalence of smoking, physical inactivity, obesity as well as changing dietary and lifestyle patterns (Sylla, [20]).

2. Material and Method

This research was a five-year retrospective study on data from a record of colorectal cancer patients that received treatments from 2013 to 2017 in Radiotherapy Department of Usmanu Danfodiyo University Teaching Hospital, Sokoto. A purposive sampling was considered in selecting UDUTH being it one of the cancer registries in Nigeria.

The research was designed to follow the subsequent procedure. The first stage was the discussion and formulation of Cox Proportional-Hazards Regression Model and Weibull Proportional Hazard Model. Finally, the data from one of the cancer registries (Usmanu Danfodiyo University Teaching Hospital, Sokoto) were collected for the following estimates: (Survival, Hazard and Median Survival Functions), efficiency and Performance of the fitted Model using AIC.

Software: The *R* programming language has sufficient packages required to carry out the research work. And SPSS

was used for data entries and arrangements.

2.1. Cox Proportional-Hazards Regression

The Proportional Hazards Model, proposed by Cox [22], has been used primarily in medical testing analysis to model the effect of secondary variables on survival. Its strength lies in its ability to model and test many inferences about survival without making any specific assumptions about the form of the life distribution model.

Most interesting survival-analysis research examines the relationship between survival typically in the form of the hazard function and one or more explanatory variables (or *covariates*).

The most common are linear-like models for the log hazard. For example, a parametric regression model based on the exponential distribution:

$$\log_e h_i(t) = h_o(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \quad (1)$$

Or equivalently,

$$h_i(t) = h_o(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}) \quad (2)$$

$$= h_o(t) \times e^{\beta_1 x_{i1}} + e^{\beta_2 x_{i2}} + \dots + e^{\beta_k x_{ik}} \quad (3)$$

where

$h_i(t)$ =Denotes the Hazards Function

$h_o(t)$ Is the Baseline Hazards

β_i Represents the Relative Risk

x_{ij} Represents the Covariates

$$i = 1, 2 \dots N$$

$$j = 1, 2 \dots k$$

Where i are indexes subjects, $x_{i1}, x_{i2} \dots x_{ik}$ are the values of the covariates for the i^{th} subject.

This is therefore a linear model for the log-hazard or a multiplicative model for the hazards itself. The model is *parametric* because, once the regression parameters $h_o(t), \beta_1, \dots, \beta_k$ are specified, the hazard function $h_i(t)$ is fully characterized by the model, the regression constant represents a kind of *baseline hazard* when all of the x 's are 0. Other parametric hazard regression models are based on other distributions commonly used in modelling survival data such as the Weibull distributions.

Fully parametric hazard regression models have largely been superseded by the Cox model [22], which leaves the baseline hazard function $h_o(t) = \log_e h_i(t)$ unspecified:

$$\log_e h_i(t) = h_o(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

or equivalently,

$$h_i(t) = h_o(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}) \quad (4)$$

The Cox Model is termed *semi-parametric* because, while the baseline hazard can take any form, the covariates enter the model through the *linear predictor*

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \quad (5)$$

Notice that there is no constant term (intercept) in the linear predictor: The constant is absorbed in the baseline hazard. The Cox Regression Model is a *Proportional-Hazards Model*:

Consider two observations, i and i' , that differ in their x -values with respective linear predictors

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \quad (6)$$

And

$$\eta_{i'} = \beta_1 x_{i'1} + \beta_2 x_{i'2} + \dots + \beta_k x_{i'k} \quad (7)$$

The hazard ratio for these two observations is

$$\frac{h_i(t)}{h_{i'}(t)} = \frac{h_o(t)e^{\eta_i}}{h_o(t)e^{\eta_{i'}}} = \frac{e^{\eta_i}}{e^{\eta_{i'}}} = e^{\eta_i - \eta_{i'}} \quad (8)$$

This ratio is constant over time. In this initial formulation, the research assumed that the values of the covariate x_{ij} are constant over time.

As we will see later, the Cox model can easily accommodate *time-dependent covariates* as well.

2.2. Parametric Model

The parametric proportional hazards model is the parametric versions of the Cox Proportional Hazards Model. It is given with the similar form to the Cox PH models. The hazard function at time t for the particular patient with a set of k covariates (x_1, x_2, \dots, x_k) is given as follows:

$$\begin{aligned} h(t|X) &= h_o(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k) \\ &= h_o(t) \exp(\beta'X). \end{aligned} \quad (9)$$

The key difference between the two kinds of models is that the baseline hazard function and is assumed to follow a specific distribution when a fully parametric PH model is fitted to the data, whereas the Cox model has no such constraint. The coefficients are estimated by partial likelihood in Cox model but maximum likelihood in parametric PH model. Other than this, the two types of models are equivalent. Hazard ratios have the same interpretation and proportionality of hazards is still assumed. A number of different parametric PH models may be derived by choosing different hazard functions. The commonly applied models are exponential, Weibull, or Gompertz Models.

2.3. Weibull Proportional Hazard Model

Suppose that survival times are assumed to have a Weibull distribution with scale parameter λ and shape parameter α so that the survival and hazard function of a $W(\lambda, \alpha)$ distribution are given by

$$S(t) = \exp(-\lambda t^\alpha), h(t) = \lambda \alpha (t)^{\alpha-1} \quad (10)$$

With $\lambda, \alpha > 0$. The hazard rate increases when $\alpha > 1$, and decreases when $\alpha < 1$ as time goes on. When $\alpha = 1$, the hazard rate remains constant, which is the special exponential case.

Under the Weibull PH model, the hazard function of a particular patient with covariates (x_1, x_2, \dots, x_k) is given by

$$h(t|X) = \lambda\alpha(t)^{\alpha-1} \exp(\beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k)$$

$$h(t|X) = \lambda\alpha(t)^{\alpha-1} \exp(\beta'X) \tag{11}$$

We can see that the survival time of this patient has the Weibull distribution with scale parameter $\lambda \exp(\beta'x)$ and shape parameter α . Therefore, the Weibull family with fixed α possesses PH property. This shows that the effects of the explanatory variables in the model alter the scale parameter of the distribution, while the shape parameter remains constant. The corresponding survival function is given by

$$S(t|X) = \exp\{-\exp(\beta'x) \lambda t^\alpha\}. \tag{12}$$

After a transformation of the survival function for a Weibull distribution, we can obtain

$$\log\{-\log S(t)\} = \log \lambda + \alpha \log t. \tag{13}$$

2.4. Model Evaluation Using Akaike Information Criterion (AIC)

To select a model that performs best among models we use Akaike Information Criterion (AIC) proposed by Akaike

[23]. AIC is a measure of the goodness of fit of an estimated statistical model (Akaike, [23]). The AIC is an operational way of trading off the complexity of an estimated model against how well the model fits the data. Given a set of data, several competing models may be ranked according to their corresponding AIC, and the one having the smallest AIC is the best. AIC is the first model criterion to gain widespread acceptance. It was an extension to the maximum likelihood principle. AIC is given by the formula:

$$AIC = -2\{\ln(\text{likelihood})\} + 2k \tag{14}$$

where

Likelihood=The probability of the data in a given model

K=The number of the parameters in the model

For the number of parameters, k=2 for the Cox and Weibull proportional hazard models (Klein and Moeschberger, [24]).

3. Result of Findings

3.1. Results from Cox Proportional Hazard Model

Log Likelihood=-68.097.

Table 1. Results from Cox Proportional Hazards Model.

Covariates	Coef	Exp (coef)	Se (coef)	Z	P	
Age	-19.290	0.000	0.009	-2038.040	0.000	***
Age At Diagnosis	19.290	238600000	0.009	2038.049	0.000	***
Associated Comps.	0.122	1.130	0.299	0.410	0.682	
Family History	-20.000	0.000	10540	-0.002	0.998	
Sex	1.139	3.123	0.449	2.535	0.011	*
Stage	-0.871	0.418	0.328	-2.661	0.008	**
Tribe	0.069	1.071	0.276	0.250	0.802	
Type Of Colorectal	-0.405	0.667	0.237	-1.712	0.087	
Type Of Treatment	0.353	1.423	0.414	0.852	0.394	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' .

Table 2. Cox Proportional Hazards Model with C. I for colorectal Cancer Patients.

Covariates	Exp (coef)	Exp (coef)	lower. 95	upper. 95
Age	0.000	238500000	0.000	0.000
Age At Diagnosis	238600000	0.000	234200000	243000000
Associated Comps.	1.130	0.885	0.629	2.030
Family History	0.000	483300000	0.000	Inf
Sex	3.123	0.320	1.295	7.532
Stage	0.418	2.390	0.22	0.795
Tribe	1.071	0.933	0.624	1.839
Type Of Colorectal	0.667	1.499	0.42	1.06
Type Of Treatment	1.423	0.703	0.632	3.206
Concordance=0.831 (se=0.036)				
Likelihood ratio test=51.350 on 9 df, p=0.0000				
Wald test=17.390 on 9 df, p=0.0400				
Score (logrank) test=30.100 on 9df p=0.000				

Table 2 shows that concordance is put at 0.831 with 0.036 SE, the likelihood ration test stands at 51.359 with 9 d.f and the P-Value as 0.0000 which shows that the likelihood ration test is highly significant. The Logrank test stands at 30.100 with 9 d.f, so it is highly significant.

The Wald test equals to 17.390 with 9 d.f also it is significant. Looking at the P-values of the test results above

i.e. the Concordance, the Likelihood ratio, the Log-rank and Wald test we can say that the predictor variables can significantly predict the survival of colorectal cancer patients. Also Age of the patient and Age at Diagnosis are the most effective covariates which relates to the survival of patients as we have seen them displayed in the output of R-programme.

3.2. Results from Cox Proportional Hazards Assumption Using Statistical Test

Table 3. Results from Cox Proportional Assumptions Using Statistical Test.

Covariates	rho	Chisq	P
Age	-0.089	0.169	0.681
Age At Diagnosis	-0.089	0.169	0.681
Associated Comps.	-0.001	0.000	0.995
Family History	-0.506	0.000	0.999
Sex	0.145	0.587	0.444
Stage	0.192	0.966	0.326
Tribe	-0.045	0.098	0.754
Type Of Colorectal	-0.308	4.170	0.051
Type Of Treatment	0.0196	0.010	0.922
GLOBAL	NA	6.170	0.723

From the result of table 3 the test is not statistically significant for each of the covariates, and the global test is also not statistically significant. Therefore, we can assume the proportional hazards (which mean that proportion hazards assumptions are met).

3.3. Results from Cox Proportional Hazard Assumptions Graphical Method

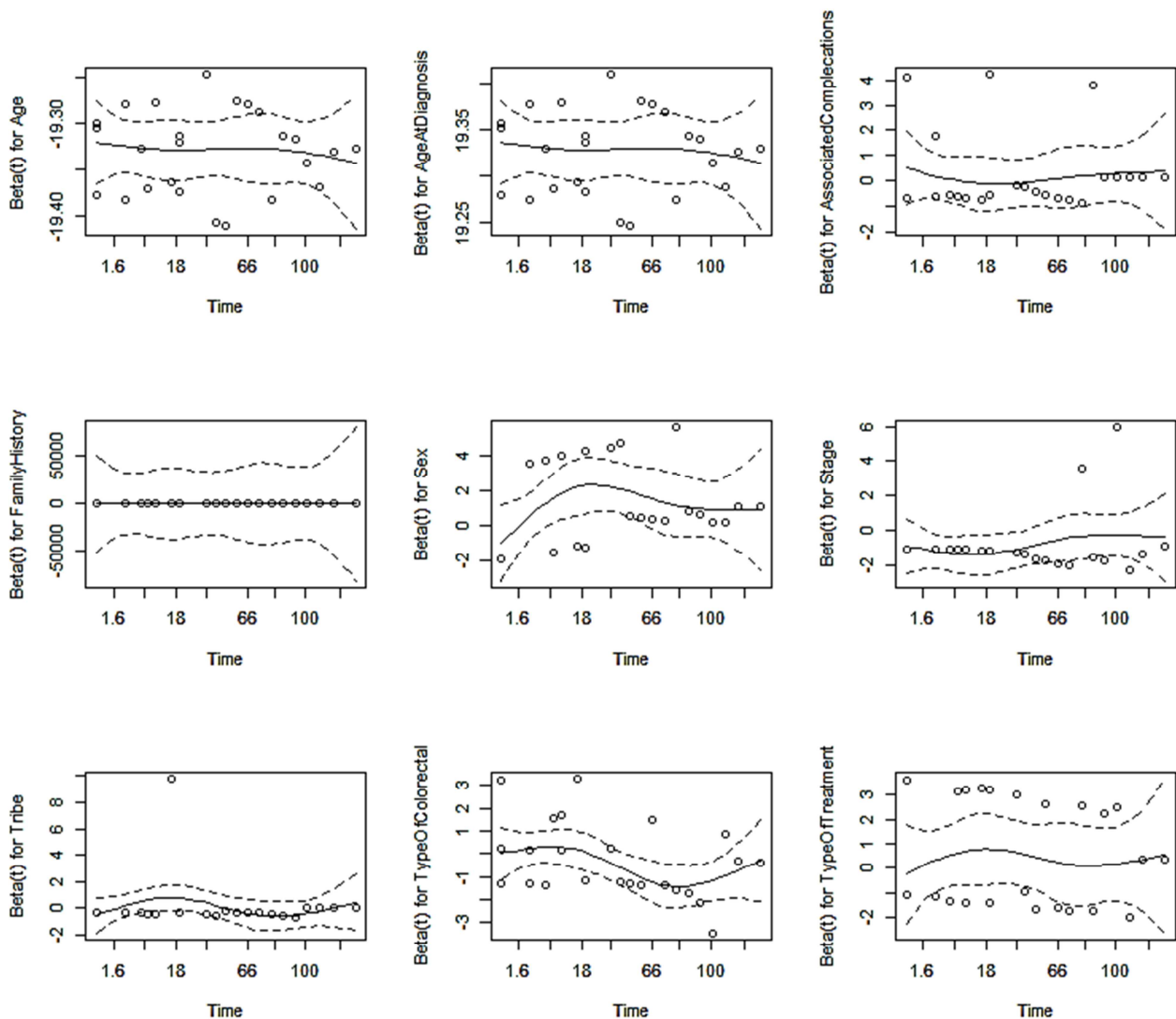


Figure 1. Cox Proportional Hazard Assumptions Graphical Method.

In the figure 1, the solid line is a smoothing spline fit to the plot, with the dashed lines representing a standard-error band

around the fit. Systematic departures from a horizontal line are indicative of non-proportional hazards, since proportional

hazards assumes that estimates $\beta_1, \beta_2, \dots, \beta_9$ do not vary much over time. From the graphical inspection, there is no pattern with time. The assumption of proportional hazards appears to be supported for the covariates.

3.4. Results from Analysis of Weibull Proportional Hazard Model

Log. Likelihood=-139.890.

Table 4. Results from Analysis of Weibull Proportional Hazard Model.

Covariate	Mean	Coef	Exp (Coef)	Se (Coef)	Wald p
Age	44.732	-2.010	0.134	0.697	0.004
Age At Diagnosis	43.402	2.010	7.460	0.697	0.004
Associated Comps.	0.026	0.136	1.146	0.339	0.687
Family History	1.198	-14.581	0.000	659.417	0.982
Sex	1.368	0.985	2.678	0.466	0.034
Stage	0.969	-0.891	0.410	0.356	0.012
Tribe	1.046	0.046	1.047	0.283	0.870
Type Of Colorectal	1.037	-0.420	0.657	0.234	0.073
Type Of Treatment	1.413	0.271	1.311	0.465	0.560
Log (scale)		-12.382	0.000	814.011	0.988
Log (shape)		-0.211	0.810	0.160	0.188
Events	27				
Total time at risk	8172				
Log. Likelihood	-139.890				
LR test statistic	52				
Degrees of freedom	9				
Overall p-value	0.000				

From the result of table 4, it is observed that the chi-square value (log-Rank test) is 52 with 9 d.f and p-value is 0.000. Age of the patients and Age at Diagnosis are the most effective covariates with p-values=0.001 which are highly significant and related to the survival of patients. The maximum log likelihood for Weibull is -139.890, while the overall p-value is 0.000 which show that the model is adequate enough to predict the survival of the Colorectal Cancer patients.

3.5. Results from the Comparison Between Cox and Weibull Proportional Hazard Model

Table 5. Summary of the Results of the Analysis and Akaike Information Criterion.

Covariates	Cox Model		Weibull Model	
	Coef	P	Coef	P
Age	-19.290	0.000	-2.010	0.004
Age At Diagnosis	19.290	0.000	2.010	0.004
Associated Complecations	0.122	0.682	0.136	0.687
Family History	-20.000	0.998	-14.58	0.982
Sex	1.139	0.011	0.985	0.034
Stage	-0.871	0.008	-0.891	0.012
Tribe	0.069	0.802	0.046	0.870
Type Of Colorectal	-0.405	0.087	-0.420	0.073
Type Of Treatment	0.353	0.394	0.271	0.560
Akaike Information Criterion (AIC)	140.194		283.780	

From Table 5, Cox Regression Model has the least AIC value (140.194); this means it has the best performance in terms of handling colorectal cancer data; Weibull Regression Model has the highest AIC value (283.780); so, it performs less.

4. Conclusion

The results of this study shows that, according to our colorectal cancer data, the semi-parametric Cox regression model could better determine the factors associated with the colorectal cancer disease than the parametric Weibull proportional hazards model. However, in the present study, the Cox model provided an efficient and a better fit to the study data than Weibull model. Therefore, it would be better for researchers of the health care field to consider this model in their researches concerning the colorectal cancer disease if the assumption of proportional hazards is fulfilled.

References

- [1] Brenner, H., Kloor, M., & Pox, C. P. (2014). Colorectal cancer. *The Lancet*, 1490-1502.
- [2] Potter, J. D., & Hunter, D. (2008). Colorectal Cancer. In H.-O. Adami, D. Hunter, & D. Trichopoulos, *Textbook of Cancer Epidemiology* (pp. 275-297). New York: Oxford University Press, Inc.
- [3] Arnold, M., Sierra, M. S., Laversanne, M., & Soerjomataram, I. (2017). Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, pp. 683-691.
- [4] Hosmer D. W., Lemeshow S., and May S. (2008). *Applied Survival Analysis: Regression Modeling of Time- to- Event Data*. 234-654.

- [5] Nigerian National System of Cancer Registries (NSCR) 2018.
- [6] Abdulkareem, F., (2009). Epidemiology and Incidence of Common Cancer in Nigeria. Presentation of Cancer Registration and Epidemiology Workshop. April, 2009.
- [7] Dickman, P. W. and Hakulinen, T. (2008). Population-Based Cancer Survival Analysis. John Wiley and Sons, UK 23-65.
- [8] Kleinbaum D G and Klein M. (2005) *Survival Analysis* New York: Springer. 16-53.
- [9] Dickman, P. W. (2010). An Introduction and Some Recent Development in Statistical Methods for Population-Based Cancer Survival Analysis. *Statistical Methods for Population-Based Cancer Survival Analysis*. Milan, 33-55.
- [10] Adejumo, A. O. and Ahmadu, A. O. (2016) A Study of the Slope of Cox Proportional Hazard and Weibull Models. *Science World Journal* Vol 11 (No 3) 2016.
- [11] Quantin, C., Michal, A., Thierry, M., Gillian, B., Todd, M., Mohammed, A., et al., (1999) Variation over Time of the Effects of Prognostic Factors in a Population based Study of Colon Cancer: Comparison of Statistical Models. *American Journal of Epidemiology*, Vol. 150, 11-20.
- [12] Ahmed E. F., Paul W. V., Don H., (2007) Modeling survival in colon cancer: a methodological review. *Molecular Cancer* 2007, 6: 15 doi: 10.1186/1476-4598-6-15.
- [13] Abdulkabir M., Ahmadu A. O., Udokang A. E., Raji S. T., (2015). *Gradient Curve of Cox Proportional Hazard and Weibull Models*. Automatic Control of Physiological State and Function, 2:2 <http://dx.doi.org/10.4172/2090-5092.1000108>
- [14] Wang Kesheng, Xuefeng Liu, Yue Pan, Daniel Owusu1, Chun Xu. (2017) Comparison of Cox Regression and Parametric Models. *Journal of Data Science* 16 (2017), 423-442.
- [15] World Health Organization. 2013a Cancer Control: A Global Snapshot in 2015 <http://www.who.int/cancer/cancer-snapshot-2015/en/> Accessed on 9 November 2016.
- [16] Knut A. M., Dejan Ignjatovic, Marianne A. M. (2017) Tailored Treatment of Colorectal Cancer: Surgical, Molecular, and Genetic Considerations. *Clinical Medicine Insights: Oncology*. DOI: 10.1177/1179554917690766.
- [17] Armstrong, B. K. (1992). The Role of Cancer Registry in Cancer Control. *Cancer Causes and Control*. 3: 569-579.
- [18] Zaki A. (2015) *Log-linearity for Cox's regression model*. University of Oslo.
- [19] Yuan, Xingchen, "Survival Model and Estimation for Lung Cancer Patients." (2005). *Electronic Theses and Dissertations*. Paper 1002. <http://dc.etsu.edu/etd/1002>.
- [20] Sylla BS, Wild CP (2011). A million Africans a Year Dying from Cancer by 2030: What can cancer research and control offer to the continent? *Int J Cancer*.
- [21] Collett, D.(2003). *Modelling Survival Data in Medical Research*. Chapman and Hall, London, 37-87.
- [22] Cox D: Regression Models and Life Tables (with Discussion). *J Roy Stat Soc B* 1972, 4: 187-220.
- [23] Akaike H (1977). A new look at the statistical model identification. *IEEE Trans Automatic Control*, 19, 716-23.
- [24] Klein J, Moeschberger M. (1997) *Survival Analysis: Techniques for Censored and Truncated Data*, New York: Springer-Verlag.