ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1835

Comparing Cox Proportional Hazard Model and Some Parametric Proportional Hazard Models for Analyzing the Survival Time of Patients with Cardiovascular Disease in Kebbi State, Nigeria

Adamu Buba[1]; Altine Muhammad Muhammad[2]; Gambo Musa[3]; Yakubu Aliyu[4]

[1]Department of Mathematics/Statistics Federal University Birnin Kebbi, Kebbi State Nigeria
[2,3]Department of Mathematics/Statistics Federal University Birnin Kebbi, Kebbi State Nigeria
[4]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria

Publication Date: 2026/05/27

Abstract: In this paper, we compare the results of the survival time of Cardiavascular patients using cox proportional hazard model, Weibull, log-normal, and Gompertz proportional hazard models. The analysis was based on data obtained from seventy eight patients suffering from cardiovascular disease in Federal Teaching Hospital, Birnin Kebbi, Kebbi State Nigeria. Information from these patients were obtained between 2020 to 2024. The log rank test was applied to compare between the survival curves of the patients based on their gender and other type of diseases (diabetes and hypertension). The results of the log rank test showed that, the survival time of the patients did not differ significantly based on gender, and patients with diabetes and those without diabetes. On the other hand, there is statistical significant difference between patients with hypertension and patients without hypertension. The data was then analyzed using the aforementioned survival models. To determine the best models, Akaike Information Criteria and Bayesian Information Criteria were used. The results of the study revealed that the cox proportional hazard model is more efficient in fitting the survival information. Finally, different cox proportional hazard models with interaction were fitted and likelihood ratio test was used to determine the most efficient model.

Keywords: *Survival Models, Cox Proportional Hazard Model, Cardiovascular Disease, Log Rank Test, Parametric Proportional Hazard Models.*

How to Cite: Adamu Buba; Altine Muhammad Muhammad; Gambo Musa; Yakubu Aliyu (2026) Comparing Cox Proportional Hazard Model and Some Parametric Proportional Hazard Models for Analyzing the Survival Time of Patients with Cardiovascular Disease in Kebbi State, Nigeria. *International Journal of Innovative Science and Research Technology*, 11(5), 1720-1725. https://doi.org/10.38124/ijisrt/26may1835

I. INTRODUCTION

Cardiovascular disease (CVD) is an aggregation of disorders of the heart and the blood vessels. Coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, congenital heart disease, deep vein thrombosis and pulmonary embolism are collectively named as cardiovascular diseases (CVDs) (Gaziano, 2001; Mendis *et al.,* 2011). It is the leading cause of death globally for both men and women. According to (Mendis *et al.,* 2011; Moran et al., 2014) more people died annually from CVDs than from any other disease. Cardiovascular diseases (CVDs) are the leading cause of mortality globally. In 2021 alone, CVD accounted for about 20.5 million deaths which is approximately one-third of all global deaths Di Cesare *et. al*., (2024). In addition, approximately 80% of the 20.5 million CVD-related deaths in 2021 occurred in low and middle income countries. Hence, the burden of CVD is not evenly distributed, it varies throughout the world in type and distributions especially between developed and developing nations. Socio economic factors including globalization, aging and accelerated urbanization continue to shape an evolving burden of CVD in Nigeria Ogah *et al*., (2023). CVD is a significant public health concern responsible for 11% of over 2 million Non-communicable Diseases (NCD) deaths in Nigeria annually. It is also responsible for a high burden of morbidity and disability. Most people with CVDs are not aware until catastrophes like stroke, heart attack or death occur.

Prevention of CVDs is the most prioritized issues to be considered but equivalently, scholars should pay attention on the way to prolong the life of affected CVD patients. Appropriate intervention should be made so as to reduce mortality and morbidity due to CVD. All potential stakeholders should spend considerable time to identify the most important risk factors that could be a cause for death of cardiovascular patients.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1835

Survival analysis is a collection of statistical techniques that is used in studying the occurrence and timing of events (Allison, 2010; Crowder, 2012; Ibrahim et al., 2001). It is applied in many fields such as Biology, Economics, Engineering, as well as Medicine (Allison, 2010; Ibrahim et al., 2001). Survival analysis involves modeling the time to the occurrence of an event data and in the current study, death is considered as the event. In survival analysis literature traditionally only a single event occurs, after which the organism or mechanism is dead or broken. Several methods have been developed for the analysis of survival data such as Kaplan-Meier, Log-rank test, Cox regression, Accelerated Failure Time (AFT), but due to complexity of data one may be popular than the others for predicting events (Cox, 1972; Hosmer *et al*., 2008; Kalbfleisch, and Prentice, 2002; Kaplan, and Meier,1958). Thus, to make realistic analysis, there is a need to find-out the most appropriate statistical model and thus, model comparison also is made. Moreover, since considering the entire data set is challenging in terms of time, human resource and finance, this research will consider samples so as to make inference about the population.

The rest of the paper is organized as follows: section 2 we discus on the cox proportional hazard model, Weibull, gamma and log-logistic proportional hazard models. Results and discussion is presented in section 3 and we finally conclude in section 4.

II. METHODOLOGY

➢ *Cox Proportional Hazard Model*

The most popular regression model for the investigation of survival is Cox proportional hazard model. It is a semi-parametric proportional hazard model and is defined as:

$$h(t) = h_0(t)\exp(\boldsymbol{X\beta}) \quad (1)$$

Where $h(t)$ is the hazard function, $h_0(t)$ is the baseline hazard function (the hazard value when the value of all covariates is zero), $\boldsymbol{X} = (x_1, x_2, \cdots, x_n)'$ is the vector of risk factors included in the model and $\boldsymbol{\beta}$ is the vector of parameters to be estimated. Cox's model has become the most used procedure for modeling the relationship of covariates to a survival outcome Pourhoseingholi et al (2007). The cox proportional hazard model is semi parametric model since $h_0(t)$ is unspecified. The model in equation (3) is parametric model when $h_0(t)$ is specified. Common choices include exponential, Weibull to mention but a few.

➢ *Parametric Proportional Hazard Models*

The model in equation is said to be parametric proportional hazard model when $h_0(t)$ is specified. For instance, it is exponential proportional hazard model when the base line hazard function is the hazard function of exponential distribution.

- *Exponential Distribution*

A random variable T with parameter λ has the exponential distribution, if its hazard, density and survival functions are respectively given by:

$$h(t) = \lambda \quad (2)$$

$$f(t) = \lambda\exp(-\lambda t) \quad (3)$$

And

$$S(t) = \exp(-\lambda t) \quad (4)$$

Where $\lambda > 0$ is the scale parameter. The exponential distribution is of practical importance because of its simplicity, memory less property and constant hazard rate function.

- *Weibull Distribution*

A random variable T with parameter λ and φ has the weibull distribution, if its hazard, density and survival functions are respectively given by:

$$h(t) = \lambda\varphi t^{\varphi-1} \quad (5)$$

$$f(t) = \lambda\varphi t^{\varphi-1}\exp(-\lambda t^{\varphi}) \quad (6)$$

And

$$S(t) = \exp(-\lambda t^{\varphi}) \quad (7)$$

Where $\lambda > 0$ is the scale parameter and $\varphi > 0$ is the shape parameter. The Weibull distribution reduced to the exponential distribution when the shape parameter (φ) takes the value one. It is widely used in modeling weather forecasts in meteorology, and defining the distribution of wind speed in radar modeling. It is preferred in survival data analysis and in in medicine. For instance, the distribution of the survival period of childhood leukemia patients was analysed using Weibull distribution (Viscomi et al., 2006).

- *Log-Normal Distribution*

A random variable T with parameter λ and φ has the log-normal distribution, if its hazard, density and survival functions are respectively given by:

$$h(t) = \frac{\exp\left(-\frac{(\log(t)-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma t}\Big/1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right) \quad (8)$$

$$f(t) = \frac{\exp\left(-\frac{(\log(t)-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma t} \quad (9)$$

And

$$S(t) = 1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right) \quad (10)$$

Where $\sigma > 0$. This distribution is used in survival analysis (Tai et al., 2007).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1835

- *Gompertz Distribution*

A random variable T with parameter λ and φ has the Gompertz distribution, if its hazard, density and survival functions are respectively given by:

$$h(t) = \lambda \exp(\varphi t) \tag{11}$$

$$f(t) = \lambda \exp(\varphi t) \exp\left(\frac{\lambda}{\varphi}(1 - \exp(\varphi t))\right) \tag{12}$$

And

$$S(t) = \exp\left(\frac{\lambda}{\varphi}(1 - \exp(\varphi t))\right) \tag{13}$$

Where $\lambda > 0$ and $\varphi \in (-\infty, \infty)$. This distribution is used frequently by medical researchers and biologists in modeling mortality ratio data. It is a growth model and has been used in relation with tumor development and in describing the distribution of adult lifespans by demographers (Preston, Samuel et al., 2001) and actuaries (Willemse and Koppelaar, 2000). The distribution is also considered for the analysis of survival data and in modelling the failure rates of computer codes (Ohishi, Okamura and Dohi, 2009).

> *Study Design and Setting*

The study is a retrospective study which is based on the CVD patients who were diagnosed of the disease in Federal Teaching Hospital Birnin Kebbi, Kebbi State, Nigeria between the years 2020 to 2024. The data obtained include: age of patients, gender and health conditions (Hypertension and Diabetes).

> *Participants*

The availability of data in the hospital was not sufficient. Information from the aforementioned variables was obtained from the record of eighty two (82) cardiovascular patients'. However, the survival time of four patients each was zero. Hence, the study was based on information from the records of seventy-eight (78) patients. In addition, the hypertensive and diabetes status of some patients in the data set were missing. Multiple imputation scheme was used to handle the missing observations. This method was used because missing data are estimated by replacing each missing value with several plausible values instead of a single estimate and it helps preserve variability and reduces bias compared with other methods such as mean substitution or complete-case analysis.

The required data was derived from the hospitals' record department in which information on the aforementioned variables were obtained from the patient's record file.

> *Evaluation Criteria*

- *Akaike Information Criteria and Bayesian Information Criteria*

The Akaike Information Criteria (*AIC*) and Bayesian Information Criteria (*BIC*) were used to compare between the parametric and semi-parametric proportional hazard models. The AIC and BIC are defined mathematically as:

$$AIC = -2ll + 2p \tag{14}$$

And

$$BIC = -2ll + p\log(n) \tag{15}$$

Respectively, where ll is the natural logarithm of the likelihood function, p is the number of parameters and n is the sample size. The model with the least value of these information criteria is regarded as the best model. In a situation where models are nested, the AIC and BIC becomes inappropriate. To overcome this problem, likelihood ratio test is employed.

- *Likelihood Ratio Test*

The likelihood ratio test is used in comparing two models that are nested. That is, it is used to test whether a fitted model to a given data set is statistically superior to the fits of the other model. The hypothesis of interest are $H_0 : \Phi \in \Phi_0$ and $H_1 : \Phi \in \Phi_1$ where Φ_0 and Φ_1 are the parameter space of the reduced model and that of the full model respectively. The test statistic is given by:

$$\tau = -2[ll_r - ll_f] \tag{16}$$

Where ll_r and ll_f are the log of the likelihood functions for the reduced model and the full model respectively. The statistic τ is asymptotically distributed as χ^2 with q (the difference between the number of parameters of the full and reduced models) degrees of freedom.

III. RESULTS AND DISCUSSION

We first compared between the survival time of the patients based gender, hypertension and diabetes status using log-rank test.

> *Log Rank Test*

We performed log rank test to ascertain the equality or otherwise of the survival time of the patients in different category (gender, hypertension and diabetes). The result of the test is given in Table 1.

Table 1 Log Rank Test of Equality Over Groups

Category	**chi SQ value**	**DF**	**P**
Gender	1.9	1	0.2000
Hypertension	15.4	1	0.0001
Diabetes	0.6	1	0.4000

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1835

From the results in Table 1, the survival experience of male patients did not differ significantly from that of female patients since $p = 0.2000$, so also patients with diabetes did not differ significantly from those without diabetes based on log-rank test: $\chi^2 = 0.6$, df = 1, p = 0.4000. On the other hand, patients with hypertension differ significantly from those without hypertension based on the log-rank test: $\chi^2 = 15.4$, df = 1, p = 0.0001. We then fit cox, exponential, weibull, log-normal and Gompertz proportional hazard models. The results of these fits are given in Table 2.

Table 2 Proportional Hazard Models Fitted to Cardiovascular Patients

Model	**Covariates**	**Coef**	**Exp (coef)**	**S.E (coef)**	**Log Lik**	**AIC**	**BIC**
Cox	Age	-0.0144	0.9857	0.0202	-43.9072	95.8143	99.14715
	Gender	-0.3857	0.6800	0.6850			
	Diabetes	0.3935	1.4821	0.6471			
	Hypertension	2.7154	15.1101	0.8729			
Exponential	Age	-0.0182	0.0199	0.9820	-96.0258	202.0516	213.8352
	Gender	-0.3752	0.6541	0.6871			
	Diabetes	0.7082	0.5512	2.0303			
	Hypertension	2.7427	0.6775	15.5291			
Weibull	Age	-0.0179	0.0200	0.9820	-95.9973	203.9946	218.1348
	Gender	-0.3750	0.6560	0.6870			
	Diabetes	0.7940	0.6650	2.2100			
	Hypertension	2.8400	0.8000	17.2000			
Log-normal	Age	0.0174	0.0181	1.0175	-95.4452	202.8905	217.0307
	Gender	0.6202	0.6490	1.8593			
	Diabetes	-0.2940	0.5855	0.7453			
	Hypertension	-2.8450	0.6507	0.0581			
Gompertz	Age	-0.0178	0.0199	0.9820	-95.8774	203.7549	217.8951
	Gender	-0.3300	0.6630	0.7190			
	Diabetes	0.87700	0.6470	2.4000			
	Hypertension	2.9600	0.8050	19.4000			

Table 2 gives the fits of the cox, exponential, Weibull, log-normal and Gompertz proportional hazard models. Based on the results, cox proportional hazard model is more efficient in fitting the survival data since it has the least AIC and BIC values. We then fit the cox proportional hazard model assuming interactions between the different covariates. Fifteen different models were fitted: the first five models were based on interactions with the hypertension covariate, the second five based on the diabetes covariate and the last five based on gender. In model 1 hypertension was assumed to interact with all other covariates, in model 2 no interaction with diabetes, in model 3 diabetes is not included, while model 4, the interaction between hypertension and gender is not include and in model 5, the covariate: gender, is not included. Similar procedure is followed for models 6 to 10 and models 11 to 15, for instance, models 12 to 15 are sub-models of model 11 and model 13 is a sub-model of model 12, also model 15 is a sub-model of model 14. The information criteria for these models are given in Table 3.

Table 3 Information Criteria for Cox Proportional Models

Model	**LogLik**	**AIC**	**BIC**
Model 1	-40.8174	95.63489	101.4674
Model 2	-40.9421	93.88417	98.88345
Model 3	-41.8061	93.61215	97.77822
Model 4	-42.9192	97.83842	102.8377
Model 5	-43.4267	96.85341	101.0195
Model 6	-42.6556	99.31117	105.1437
Model 7	-43.1712	98.34236	103.3416
Model 8	-50.6588	111.3175	115.4836
Model 9	-42.6961	97.39215	102.3914
Model 10	-43.2365	96.47296	100.639
Model 11	-40.9748	95.94966	101.7822
Model 12	-43.7047	99.40948	104.4088
Model 13	-50.6568	111.3136	115.4797
Model 14	-41.0026	94.0052	99.00448
Model 15	-42.1714	94.34277	98.50884

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1835

From Table 3, the best fitted models based on the log-likelihood are models 1, 2, 11, 14, 3 and 15, based on AIC are models 3, 2, 14, 15, 1 and 11 while based on BIC are models 3, 15, 2, 14, 10 and 5. Hence, we conclude based on the log-likelihood, AIC and BIC models 3, 2, 14 1 and 15 are more efficient in fitting the data. However, models 2 and 3 are sub-models of model 1 as explained earlier. Also, model 15 is a sub-model of model 14. Hence, a likelihood ratio test is used to test for the significance of the additional parameter(s).

Table 4 Likelihood Ratio Tests for the Fitted Models

Model comparison	τ	p
Model 1 Vs Model 2	0.2493	0.6176
Model 1 Vs Model 3	1.9773	0.3721
Model 2 Vs Model 3	1.7280	0.1887
Model 14 Vs Model 15	2.3376	0.1263

Hence, based on the p-values from Table 4, the sub-models are more efficient. Based on the p-values, models 2 and 3 are more efficient than model 1, so also, model 15 is more efficient than model 14. In addition, the *LRT* suggest that, model 3 is more efficient than model 2. Hence, model 3 is more efficient in fitting the survival time cardiovascular patients. The result of model 3 is given in Table 5.

Table 5 Result of the Most Efficient Model

Covariates	**coef**	**exp(coef)**	**se(coef)**	**z**	**p**
Age	-0.03677	0.9639	0.02573	-1.429	0.153
HypertensionYES	2.45469	11.64287	1.98277	1.238	0.2157
GenderM	2.26899	9.66967	1.57688	1.439	0.1502
Age:HypertensionYES	0.03421	1.0348	0.0387	0.884	0.3768
HypertensionYES:GenderM	-3.64396	0.02615	1.89833	-1.92	0.0549
Overall likelihood ratio test=22.19 on 5 df, p=0.0004821					

The fitted cox proportional hazard model in Table 5, include Age, hypertension, gender and the interactions between age and hypertension, and hypertension and gender.

IV. CONCLUSION

In this study, we fitted a semi-parametric proportional hazard model and some parametric proportional hazard models. Log rank test showed that, there is no significant difference between the survival time of the patients based on gender and diabetes status. It was also shown that, there is significant difference between the survival time of patients with hypertension and patients without hypertension. We then, compare the results of the survival times of cardiovascular patients using the AIC, BIC and likelihood ratio test. It was found that the most efficient model in fitting the survival time of cardiovascular patients.

REFERENCES

[1]. Allison, P. D. (2010). Survival analysis using SAS: a practical guide. Sas Institute.

[2]. Di Cesare, M., Perel, P., Taylor, S., Kabudula, C., Bixby, H., Gaziano, T. A. McGhie DV, Mwangi J., Pervan, B., Narula, J., Pineiro, D. & Pinto, F. J. (2024). The heart of the world. *Global heart*, *19*(1).

[3]. Gaziano, J. M. (2001). General considerations of cardiovascular disease. *Braunwald's Heart Disease: a Textbook of Cardiovascular Medicine*, *7*, 1-19.

[4]. Hosmer Jr, D.W., Lemeshow, S. and May, S., 2008. Applied survival analysis: regression modeling of time-to-event data (Vol. 618). Wiley-Interscience.

[5]. Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, 57(2):383–388.

[6]. Kalbfleisch, J.D. and Prentice, R.L., 2002. The Statistical Analysis of Failure Time Data 2002 New York. NY Wiley Crossref.

[7]. Kaplan, E.L.and Meier, (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457-481.

[8]. Mendis S, Puska P, Norrving B (2011). Global Atlas on Cardiovascular Disease Prevention and Control (PDF). World Health Organization in collaboration with the World Heart Federation and the World Stroke Organization, 3–18. ISBN 978-92-4-156437-3

[9]. Ogah, O. S., Orimolade, O. A., & Jinadu, T. O. (2023). Cardiovascular diseases in Nigeria: current status, threats, and opportunities. *Circulation*, *148*(19), 1441-1444.

[10]. Ohishi, K., Okamura, H., & Dohi, T. (2009). Gompertz software reliability model: Estimation algorithm and empirical validation. *Journal of Systems and software*, *82*(3), 535-543.

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

[12]. Preston, Samuel H.; Heuveline, Patrick and Guillot, Michel (2001), "Demography:measuring and modeling population processes",Oxford: Blackwell.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1835

[13]. Tai, P., Chapman, J. A. W., Yu, E., Jones, D., Yu, C., Yuan, F., & Sang-Joon, L. (2007). Disease-specific survival for limited-stage small-cell lung cancer affected by statistical method of assessment. *BMC cancer*, *7*(1), 31.
[14]. Viscomi S, Pastore G, Dama E, et al. (2006), "Life expectancy as an indicator of outcome in follow-up of population-based cancer registries: the example of childhood leukemia0 *Ann Oncol*, 17(1), 167-71.
[15]. Willemse, W. J and Koppelaar, H. (2000), "Knowledge elicitation of Gompertz' law of mortality", *Scandinavian Actuarial Journal*, 2, 168–79.