MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

Application of Non Parametric and Semi-Parametric Models on Survival Time of Patients with Cardiovascular Disease: Case Study of Barau Dikko Teaching Hospital, Kaduna, Nigeria

Enoch Yabkwa Yanshak*, Aliyu Yakubu[1], and Kolawole Daramola[1]

[1]*Department of Statistics, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Nigeria.*

KEYWORDS

Kaplan Meier curve
Log rank Test
Survival function
Hazard function
Cox proportional model

ABSTRACT

This study examines the survival times of cardiovascular patients using Kaplan-Meier (KM) survival curves and the Cox Proportional Hazards model. Analysis of data from Barau Dikko Teaching Hospital, Nigeria, revealed shorter survival times among male patients and those consuming alcohol or smoking. KM curves and Log-rank tests identified significant risk factors such as high blood pressure and irregular pulse rates. Cox model hazard ratios highlighted alcohol consumption as the highest risk factor. These findings demonstrate the utility of non-parametric and semi-parametric models in identifying survival determinants among cardiovascular patients.

ARTICLE HISTORY

Received 9 December 2024
Received in revised form 6 August 2025
Accepted 30 August 2025
Available online 17 September 2025

1. INTRODUCTION

Cardiovascular diseases (CVDs) represent a significant global health challenge, accounting for substantial morbidity and mortality worldwide. These conditions, including deep vein thrombosis, cardiac embolism, peripheral arterial disease, cerebrovascular illnesses, and congenital heart disease, collectively affect millions of individuals and impose immense economic and societal burdens [1]. According to the World Health Organization (WHO), approximately 17.9 million deaths were attributed to CVDs in 2016, representing 31% of all global deaths [4].

According to the World Heart Federation's 2023 report, cardiovascular diseases (CVDs) remain the leading cause of death globally, accounting for approximately 20.5 million deaths in 2021. This represents a significant increase from the 17.9 million CVD-related deaths reported in 2016. Notably, over 75% of these deaths occur in low- and middle-income countries, highlighting persistent global health disparities. The World Health Organization emphasizes that a substantial proportion of these deaths are due to heart attacks and strokes, with many being premature and preventable [20] . In sub-Saharan Africa, including Nigeria, limited access to healthcare resources, coupled with rising urbanization and aging populations, has contributed to the growing prevalence of CVD-related mortality.

The burden of CVD is not uniform across countries. While developed nations benefit from advanced diagnostic and therapeutic measures, LMICs face systemic challenges in prevention, early detection, and treatment. These disparities are reflected in the disproportionate mortality rates observed in LMICs, which often lack the infrastructure to address the rising prevalence of chronic diseases. Understanding these regional differences and their underlying causes is crucial for tailoring interventions. For instance, Ethiopia's experience

*Corresponding author:
E-mail address: Enoch Yabkwa Yanshak < enochyanshak@gmail.com >.
https://doi.org/10.56532/mjsat.v5i3.440
2785-8901/

with globalization and demographic transitions underscores the need for context-specific strategies to manage CVD [5].

Addressing CVD mortality requires a twofold approach: preventive measures to reduce new cases and strategies to improve survival outcomes for affected individuals. Survival analysis is a pivotal tool in this regard, providing insights into the timing of critical events, such as death or disease recurrence, and identifying significant risk factors. Non-parametric methods, such as the Kaplan-Meier (KM) curve, allow for the visualization of survival probabilities over time. Semi-parametric models, like the Cox proportional hazards model introduced by Cox (1972), enable the estimation of hazard ratios while accounting for covariates [6]. The Cox model's flexibility in handling censored data and time-varying covariates makes it a cornerstone of survival analysis in medical research.

Despite the global application of these models, there is limited research employing both KM and Cox models to analyze survival data in Nigeria, particularly for CVD patients. This study addresses this gap by applying non-parametric and semi-parametric survival analysis techniques to data from patients with cardiovascular diseases at Barau Dikko Teaching Hospital, Kaduna, Nigeria. By identifying critical risk factors and assessing their influence on survival time, this research aims to contribute to the growing body of knowledge necessary for informed clinical decision-making and policy interventions in resource-constrained settings.

1.1 Problem Statement

Cardiovascular diseases (CVDs) remain a leading cause of morbidity and mortality globally, with a disproportionate impact on healthcare systems in low- and middle-income countries like Nigeria. In Kaduna State, the increasing prevalence of CVDs has placed additional strain on medical resources, particularly at Barau Dikko Teaching Hospital, a key healthcare institution in the region. Despite this growing burden, critical gaps remain in understanding the survival outcomes and associated risk factors of CVD patients in this locality.

Existing research in Nigeria has largely overlooked the application of advanced statistical methods, such as the Cox Proportional Hazards (Cox PH) model, which is widely regarded for its robustness in survival analysis. This model allows for the simultaneous evaluation of multiple risk factors and accommodates censored data, making it invaluable for medical research and decision-making. A lack of local studies utilizing such methodologies has limited the evidence available to inform clinical decision-making, resource allocation, and treatment planning for CVD patients in the region.

This study addresses this gap by employing both non-parametric and semi-parametric models to analyze the survival time of cardiovascular patients at Barau Dikko Teaching Hospital. Non-parametric methods, including Kaplan-Meier survival curves and log-rank tests, will provide insights into survival patterns among patient subgroups. In contrast, the Cox PH model will evaluate the effects of critical risk factors such as age, gender, comorbidities, lifestyle behaviors, and treatment regimens on survival time.

The findings from this study aim to provide healthcare professionals and policymakers with actionable, evidence-based insights to optimize resource allocation and improve prevention, treatment, and management strategies for CVDs. By focusing on survival analysis using advanced statistical tools, this research will fill a significant knowledge gap and contribute to reducing the burden of cardiovascular diseases in Kaduna State and beyond.

1.2 Objectives

The objectives are as follows:

- Kaplan-Meier Curves: To describe survival patterns among cardiovascular patients based on categorical variables such as age, gender, comorbidities, and treatment regimens.
- Log-Rank Test: To statistically compare survival distributions across different subgroups of cardiovascular patients.
- Cox Proportional Hazards Model: To identify and quantify the impact of various risk factors, including age, gender, lifestyle behaviors, comorbidities, and treatment regimens, on the survival time of cardiovascular patients.

1.3 Literature Review

1) Non-Parametric Approaches

Non-parametric methods, such as Kaplan-Meier curves and the log-rank test, are widely used for descriptive and comparative survival analysis. These methods are advantageous for their simplicity and minimal assumptions about the underlying survival distributions. For instance, [12] analyzed heart failure survival in 228 patients from Ecuador's Andean population, reporting one-year and five-year survival rates of 86% and 46%, respectively. Prognostic factors included heart failure etiology and age (HR = 1.035; $p = 0.04$). While effective for summarizing survival patterns, non-parametric methods lack the ability to incorporate multiple covariates or quantify their effects.

2) Semi-Parametric Approaches

The Cox proportional hazards model remains a cornerstone in survival analysis due to its flexibility and interpretability. [18] highlighted its utility in estimating hazard ratios while adjusting for confounders, emphasizing the importance of the proportional hazards assumption. Similarly, [19] noted its accessibility for clinicians, making it a preferred choice in medical research. However, violation of the proportional hazards assumption can compromise its validity. For example, [14] modeled university students' performance using Cox models and found predictors like GPA and faculty to be significant but observed violations of the proportional hazards assumption, recommending the log-normal Accelerated Failure Time (AFT) model instead.

To address scalability in high-dimensional data, [16] proposed a scalable algorithm for semi-parametric AFT models, demonstrating improved computational speed and accuracy over existing methods. Additionally, [11] developed a privacy-preserving approach to train Cox models on distributed datasets using secure multi-party computation. While ensuring data security, this method faced computational challenges, including extended processing times.

3) Parametric Approaches

Parametric survival models, such as Weibull and log-normal models, offer precise estimation of survival times and are preferred when the underlying survival distribution is known or can be approximated. [13] compared Weibull and Cox proportional hazards models, observing that Weibull outperformed Cox when its shape parameter was correctly specified. Similarly, [15] assessed survival models for 332 cardiac patients in Ethiopia and found that male patients faced higher risks (HR = 1.93; p = 0.019). The Weibull AFT model outperformed other models in predictive accuracy. Moreover, [17] evaluated lifetime distributions, highlighting Weibull, Erlang, Gamma, and Generalized Exponential (GE) as flexible options, with log-normal distributions recommended for shorter time-to-failure data.

4) Comparison of Methods

While non-parametric methods are effective for descriptive purposes, semi-parametric and parametric models provide greater analytical depth by quantifying covariate effects. Angela (2008) emphasized the Cox model's robustness in hazard ratio estimation, whereas Samar et al. (2021) underscored the benefits of parametric models like Weibull in predictive accuracy and handling time-to-event distributions. Despite their strengths, parametric models rely on assumptions about the survival distribution, which may limit their applicability in some scenarios.

5) Recent Studies

To complement these findings, recent studies like Belaynesh & Zeytu (2021) have demonstrated the application of survival models in similar settings, providing updated insights into cardiovascular risk factors and survival trends in Sub-Saharan Africa. Incorporating such recent research enhances the relevance and applicability of the current study.

2. METHODOLOGY

2.1 Study Design

This study utilized secondary data collected from cardiovascular patients' records and information sheets at Barau Dikko Teaching Hospital in Kaduna State, Nigeria. The dataset includes patients who underwent either pre- or post-operative care and were under follow-up from January 2014 to December 2022.

2.2 Ethical Approval and Data Anonymization

The research protocol was reviewed and approved by the Barau Dikko Teaching Hospital Research Ethics Committee. Informed consent was waived due to the retrospective nature of the study. Data anonymization was ensured by removing patient identifiers, such as names and contact information, and replacing them with unique codes to protect patient privacy.

2.3 Data Description

The dataset comprises variables expected to influence the mortality of cardiovascular patients. The dependent variable, average time to survival, is defined as the duration between the admission date for treatment and either the date of death or censoring. Censoring was applied to patients who were alive during the study period or lost to follow-up before experiencing the event of interest (death).

The explanatory variables include:

- Demographics: Age, sex, and region.
- Medical conditions: Hypertension/high blood pressure, diabetes mellitus, anemia.
- Lifestyle factors: Smoking and alcohol usage.
- Clinical measures: Body mass index (BMI), ejection fraction, serum creatinine, creatinine phosphokinase, pulse rate.
- Patient status: Alive or deceased.

2.4 Data Overview

- Total number of cases: 299.
- Number of censored cases: 203.
- Number of uncensored cases: 96.
- Preprocessing: Missing data were imputed using mean/median imputation and continuous variables were scaled to ensure uniformity.

2.5 Model Justification

This study employed the Kaplan-Meier (KM) curves and Cox proportional hazards model for survival analysis.

Kaplan-Meier Curves

The KM estimator is a non-parametric method used to estimate survival probabilities and visualize survival patterns. Its stepwise nature is particularly useful for handling censored data. While parametric models assume a specific distribution for survival times, KM curves do not, making them suitable for datasets with unknown or complex survival distributions.

Cox Proportional Hazards Model The Cox model was chosen for its semi-parametric nature, allowing the inclusion of covariates without assuming a specific baseline hazard distribution. It is particularly advantageous when the proportional hazards assumption holds, as it enables the identification and quantification of risk factors. In comparison to parametric models like Weibull or log-normal, the Cox model is more flexible and robust for datasets with heterogeneous survival times or incomplete follow-up.

2.6 Software

The analysis was conducted using Python (version [insert version]). The following libraries were employed:

- Lifelines: For survival analysis (KM and Cox models).
- Pandas: For data preprocessing and manipulation.
- Matplotlib/Seaborn: For visualization of survival curves.
- Scikit-learn: For handling missing data and scaling continuous variables.

3. RESULTS AND DISCUSSION

The response variable of interest for this study was survival time of cardiovascular patients until death in days and the independent variables considered are in two groups, the continuous group and the categorized group. The continuous variables are Age, serum_sodium, ejection_fraction, Creatinine_phosphokinase and serum_creatinine. The categorized variables are anemia, Diabetes mellitus, alcoholic usage, smoking, high blood pressure, body mass index , sex, region and pulse rate . Table 1, presents the description and the summary of covariates.

Table 1. Descriptive Statistics of Categorical Covariates in the Study Population, Including Frequency and Proportions.

Covariates	Categories	Event (%)	Censored (%)	Total (%)
Anaemia	0	50 (52.083333)	120 (59.1133)	170 (56.856187)
	1	46 (47.916667)	83 (40.887)	129 (43.143813)
Diabetes	Yes	69 (71.875)	156 (76.847291)	225(75.250836)
Militus	No	27 (28.125)	47(23.152709)	74(24.749164)
High_blood_pressure	0	39 (40.625)	66 (32.512315)	105 (35.117057)
	1	57 (59.375)	137 (67.487685)	194(64.882943)
Body mass index	Normal	13 (13.541667)	82 (40.394089)	95 (31.772575)
	Over-weight	66 (68.750000)	62 (30.541872)	128 (42.809365)
	Under-weight	17(17.70833)	59 (29.064039)	76 (25.418060)
Sex	0	34 (35.416667)	71 (34.975369)	105(35.117057)
	1	62 (64.583333)	132 (65.024631)	194(64.882943)
Smoking	0	30(31.25)	66 (32.512315)	96(32.107023)
	1	66(68.75)	137(67.487685)	203(67.892977)
Alcohol usage	Yes	88(91.666667)	16(7.881773)	104 (34.782609)
	No	8(8.333333)	187(92.118227)	195(65.217391)
Pulse rate	Irregular	84 (87.5)	37(18.226601)	121(40.468227)
	Regular	12 (12.5)	166(81.773399)	178(59.531773)
Region	Kaduna Central	39(40.624997)	75(36.945813)	111(37.123746)
	Kaduna North	33(34.375000)	71 (34.975369)	104(34.782609)
	Kaduna South	24 (25.000000)	54 (26.600985)	78 (26.086957)

The median survival time of cardiovascular disease (CVD) patients in this study was 214.0 days, indicating that half of the patients experienced the event (death) with a probability of 0.5. The study involved 300 CVD patients with a mean age of approximately 60.83 years. The proportion of male patients was 64.88% (n = 194), while females comprised 35.12% (n = 105).

3.1 Patient Demographics:

- Age: The mean age of the cohort was approximately 60.83 years.
- Sex Distribution: Among the patients, 64.88% were male, and 35.12% were female. Male patients experienced a higher proportion of events (64.58%) compared to females (35.42%).
- Regional Distribution: Patients were distributed across the three senatorial zones of Kaduna: Kaduna Central (37.12%), Kaduna North (34.78%), and Kaduna South (26.09%). Patients from Kaduna Central had the highest proportion of events (40.62%), followed by Kaduna North (34.38%) and Kaduna South (25.00%).

3.2 Prevalence of Comorbidities:

- Anemia: The prevalence of anemia among the patients was 43.14% (n = 129). Patients without anemia had a higher event proportion (52.08%) compared to those with anemia (47.92%).
- Diabetes Mellitus: The prevalence of diabetes was 75.25% (n = 225). Among them, 71.88% experienced the event, compared to 28.13% of patients without diabetes.
- High Blood Pressure: High blood pressure was observed in 64.88% (n = 194) of the patients. Among those with high blood pressure, 59.38% experienced the event compared to 40.63% without high blood pressure.
- Body Mass Index (BMI): Overweight patients accounted for 42.81% of the cohort, followed by normal BMI (31.77%) and underweight patients (25.42%). Overweight patients experienced the highest proportion of events (68.75%) compared to underweight (17.71%) and normal BMI (13.54%).
- Smoking: The prevalence of smoking was 67.89% (n = 203). Patients who smoked experienced a higher proportion of events (68.75%) than nonsmokers (31.25%).
- Alcohol Usage: Among the cohort, 34.78% (n = 104) reported alcohol usage. These patients experienced a significantly higher proportion of events (91.67%) compared to non-alcohol users (8.33%).
- Pulse Rate: Irregular pulse rates were observed in 40.47% (n = 121) of the patients, with 87.50% experiencing the event compared to 12.50% of patients with regular pulse rates.

Figure.1 to Figure.8 presents the Kaplan-Meier curve for the categorical variables (Non Parametric Approach)

For a more elaborate descriptive analysis, the study uses a non-parametric approach to provide a summary of the distribution of the variables. Comparison of Survival Estimates of Different Categories of Covariates Using Kaplan-Meier Survival Curve the Kaplan-Meier survival curve used to compare the survival of cardiac patients under different categories of categorical covariates. The mean for the overall survival of the cardiovascular patients is 130days the maximum and minimum survival time of the cardiovascular patients were 285 and 4 days respectively.

Using KM estimate, we plot eight of the categorical variables with time recorded in days. The categorical variables considered for the study are:

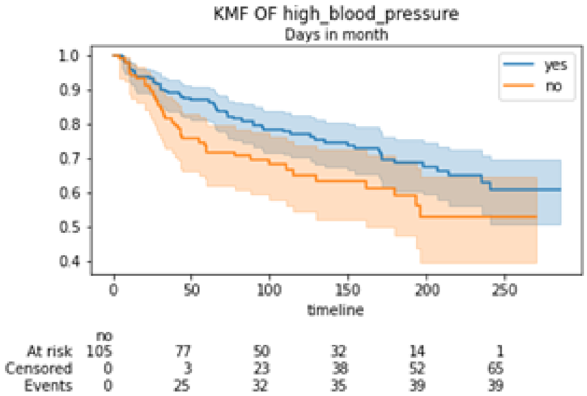

Fig. 1. Kaplan-Meier Survival Curves for Cardiovascular Patients Based on HBP

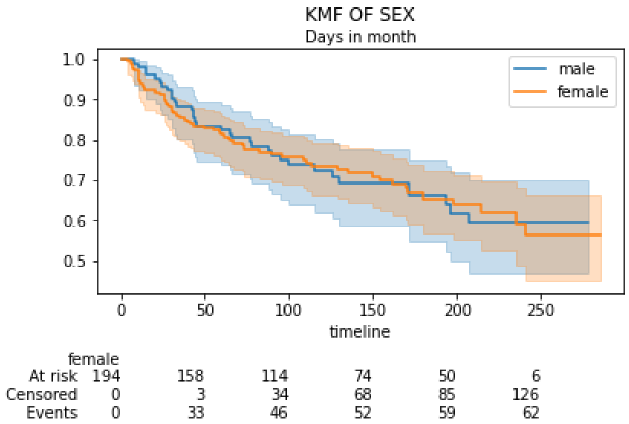

Fig. 2. Kaplan-Meier Survival Curves for Cardiovascular Patients Based on Sex

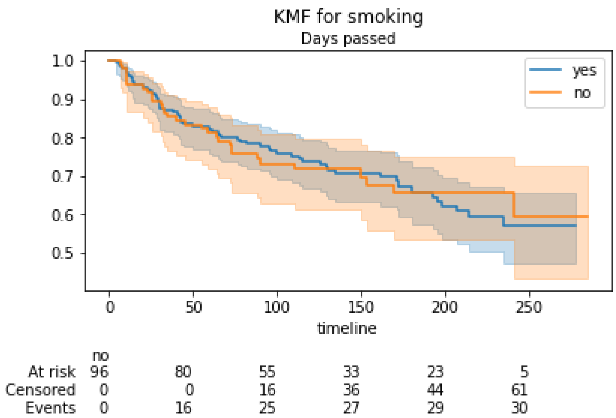

Fig. 3. Kaplan-Meier Survival Curves for Cardiovascular Patients Based on Smoking

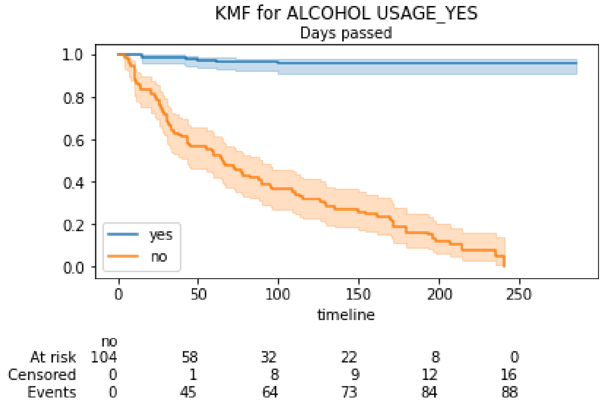

Fig. 4. Kaplan-Meier Survival Curves for Cardiovascular Patients Based on Alcohol Usage

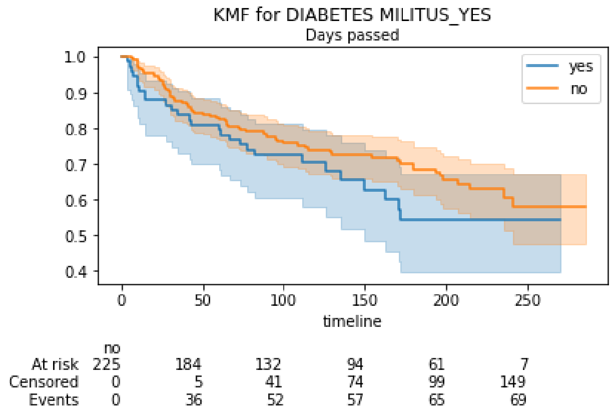

Fig. 5. Kaplan-Meier Survival Curves for Cardiovascular Patients Based on DIABETES

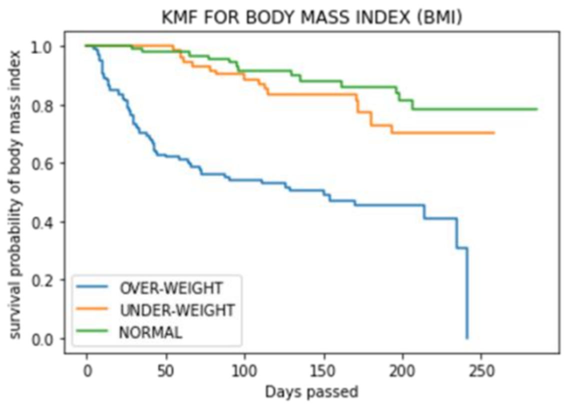

Fig. 6. Kaplan-Meier Survival Curves for Cardiovascular Patients Based on BMI

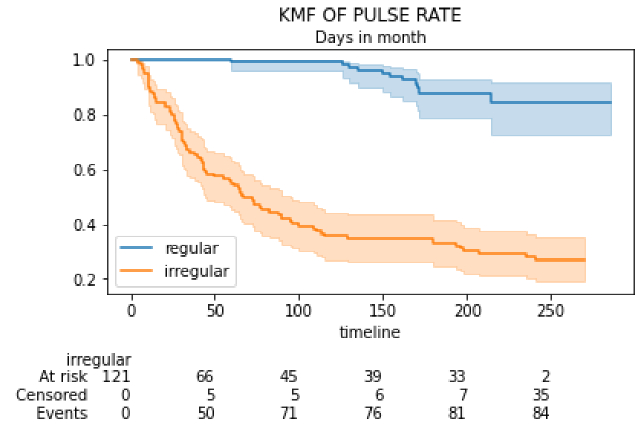

Fig. 7. Kaplan-Meier Survival Curves for Cardiovascular Patients Based on PULSE RATE

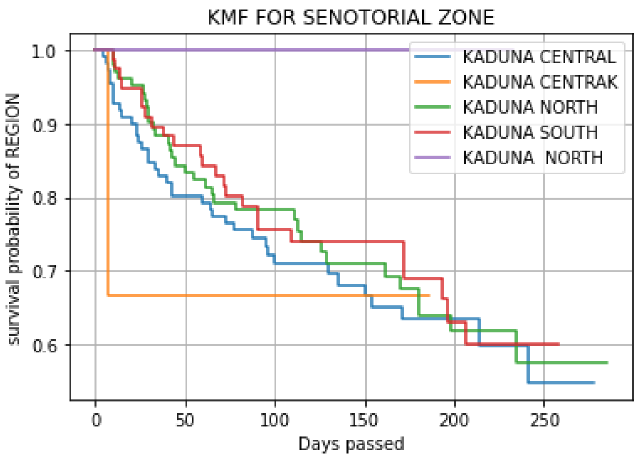

Fig. 8. Kaplan-Meier Survival Curves for Cardiovascular Patients Based on REGION

Figure 1 to figure 8 is the Comparison of Survival Estimates of Different Categories of Covariates Using Kaplan-Meier Survival Curve. The Kaplan-Meier survival curve used to compare the survival of cardiac patients under different categories of categorical covariates. In general patients belong to the categories whose survival curve lays above the survival curve of the other category has a better survival time.

In this regard, the Kaplan-Meier survival curve revealed that male patients have less survival time to the female counterpart from the curve in figure 2, the difference in the survival curves for the two groups are not wide enough. It can also be noticed from Figure 1 that patients without high blood pressure have a higher probability of surviving than patients with high blood pressure. From the curve in Figure 7 the difference in the survival curves for the two groups that is irregular and regular pulse rate are wide enough suggesting good significance difference. By looking at the number of patients who smoked with those that don't smoke in Figure 3 the graph shows that patients who don't smoke have higher survival probability compared to those that smoke. As observed in Figure 4 there is wide survival curve between patients who drink alcohol and those that don't take alcohol , those that take alcohol have less survival compare to those that don't take alcohol. Figure 5 shows the survival curve of present diabetes in the Patients or not and the revealed that with diabetes have less chance of survival compare to those without diabetes. Patients with overweight in figure 6 have less survival time compare to those with normal weight and underweight and maintain a very wide survival curve as indicated in the figure which suggesting a good significance difference. Finally, as we noticed from figure 4.8 survival curve between the three senatorial zones is not wide enough. Generally, it could be observed that graphically most of the KM curves indicate a difference between the covariate categories but a statistical test is able to reveal whether the differences we observe are significantly different.

A formal test was carried out using the Log rank test to compare difference between each categorical variable. The general hypothesis states that there is no difference between the groups. Thus we wish to test that:

Table 2. Comparison of Kaplan-Meier Survival Curves for Key Covariates Using the Log-Rank Test to Assess Differences in Cardiovascular Patient Survival.

Covariates	Chi-square value	Df	p-value
Sex	5	2	0.06527
Region	1.2	3	0.04
Smoke use	6.5	2	0.0105
Alcohol use	13.5	2	0.000238
Blood pressure	16.1	2	0.00032
Pulse rate	22.3	1	2.27e-06
Over-weight	12.7	2	0.00175

H_0: The survival times of patients between the groups are not different

H_1: The survival times of patients between the groups are different

Figures 1 to 8 illustrate Kaplan-Meier survival curves, comparing the survival probabilities of cardiac patients across different categories of categorical covariates. The KM curves provide a graphical representation of survival probabilities over time, while the Log-rank test complements these findings by statistically assessing the significance of the differences observed in the curves. Below is an enhanced interpretation linking visual findings to chi-square values:

1) *Sex (Figure 2):*

- The KM curve for sex shows that male patients have slightly lower survival probabilities than female patients. However, the curves are close to each other, indicating minimal differences in survival times.
- The Log-rank test yielded a chi-square value of 5 with a p-value of 0.06527, which is above the 5% significance level. This suggests that the observed difference in survival between male and female patients is not statistically significant.

2) *Blood Pressure (Figure 1):*

- The KM curves for patients with and without high blood pressure reveal that patients without high blood pressure have higher survival probabilities. The

separation of the curves suggests a meaningful difference in survival times.
- The Log-rank test supports this observation with a chi-square value of 16.1 and a p-value of 0.00032, indicating a highly significant difference between the two groups.

3) *Smoking (Figure 3):*
- The KM curves show that non-smokers have higher survival probabilities compared to smokers. While the difference in curves is noticeable, it is less pronounced than for other covariates.
- The Log-rank test resulted in a chi-square value of 6.5 and a p-value of 0.0105, indicating statistical significance, though less strong compared to other covariates.

4) *Alcohol Use (Figure 4):*
- The KM curves for alcohol use demonstrate a significant difference, with non-alcohol users having higher survival probabilities than alcohol users. The wide separation of the curves suggests a substantial impact of alcohol use on survival.
- The Log-rank test confirmed this with a chi-square value of 13.5 and a p-value of 0.000238, indicating strong statistical significance.

5) *Pulse Rate (Figure 7):*
- The KM curves for irregular and regular pulse rates are markedly separated, with patients having a regular pulse rate exhibiting significantly better survival probabilities.
- The Log-rank test result, with a chi-square value of 22.3 and a p-value of 2.27e-06, provides strong statistical evidence supporting the observed difference in survival times.

6) *Body Mass Index (Figure 6):*
- The KM curves reveal that overweight patients have the lowest survival probabilities, followed by underweight patients, with normal-weight patients having the highest survival probabilities. The wide separation of the curves highlights a clear difference between the groups.
- The Log-rank test supports this observation with a chi-square value of 12.7 and a p-value of 0.00175, indicating a statistically significant difference in survival times across body mass index categories.

7) *Region (Figure 8):*
- The KM curves for patients from the three senatorial zones (Kaduna Central, Kaduna North, and Kaduna South) show minimal separation, indicating similar survival probabilities across regions.
- The Log-rank test, with a chi-square value of 1.2 and a p-value of 0.04, indicates that while the difference is statistically significant, it is less pronounced compared to other covariates.

8) *General Observation:*

The KM curves visually demonstrate differences in survival probabilities across the covariates. The Log-rank test complements these findings by statistically quantifying the significance of these differences. Covariates such as smoking, blood pressure, alcohol use, pulse rate, and body mass index showed both visual and statistical significance, while sex though visually suggestive, was not statistically significant at the 5% level. This dual analysis ensures a comprehensive understanding of the factors influencing survival among cardiac patients.

Table 3. Estimated Hazard Ratios (HR) and 95% Confidence Intervals (CI) from the Cox Proportional Hazard Model for Key Risk Factors Affecting Cardiovascular Survival Time.

	Coef	**exp(coef)**	**se(coef)**	**coef lower 95%**	**coef upper 95%**	**exp(coef) lower 95%**	**exp(coef) upper 95%**	**cmp to**	**Z**	**P**	**-log2(p)**
Age	0.02	1.02	0.01	0.00	0.04	1.00	1.04	0.00	1.97	0.04	4.35
Anaemia	-0.21	0.81	0.24	-0.67	0.25	0.51	1.29	0.00	-0.88	0.38	1.40
serum_sodium	-0.02	0.98	0.02	-0.07	0.03	0.93	1.03	0.00	-0.82	0.41	1.29
ejection_fraction	0.00	1.00	0.01	-0.02	0.03	0.98	1.03	0.00	0.33	0.74	0.44
high_blood_pressure	0.45	1.56	0.23	-0.01	0.90	0.99	2.47	0.00	1.91	0.06	4.15
serum_creatinine	0.05	1.05	0.08	-0.12	0.21	0.89	1.23	0.00	0.56	0.58	0.79
Smoking	-1.11	0.33	0.27	-1.65	-0.58	0.19	0.56	0.00	-4.07	<0.005	14.36
Region_kaduna north	-1.09	0.34	0.28	-1.64	-0.55	0.19	0.58	0.00	-3.92	<0.005	13.46
Region_kaduna central	-0.61	0.54	0.29	-1.18	-0.04	0.31	0.96	0.00	-2.10	0.04	4.81
Alcohol usage_yes	3.07	21.49	0.45	2.20	3.94	8.98	51.41	0.00	6.89	<0.005	37.40
Body mass index_over-weight	1.65	5.21	0.29	1.07	2.23	2.93	9.26	0.00	5.62	<0.005	25.62
Pulse rate_irregular	2.21	9.08	0.35	1.51	2.90	4.54	18.19	0.00	6.23	<0.005	30.98
Diabetes militus_yes	-1.58	0.21	0.29	-2.14	-1.01	0.12	0.36	0.00	-5.47	<0.005	24.41

The results of the multivariable cox proportional hazard model in Table 3 showed that diabetes mellitus, smoking, alcoholic usage, irregular pulse rate, age, body mass index and region were significant covariates at 0.05 level of significance as well the Hazard Ratio (exp(coef)). If the hazard ratio is greater than 1, it suggests an increased risk of the event occurring. If less than 1, it suggests a decreased risk.

The table presents results from a Cox proportional hazards model, which evaluates the relationship between predictors and the hazard of experiencing an event over time. Each variable's effect is summarized by the coefficient (Coef), hazard ratio (exp(coef)), and significance (P-value). For Age Coefficient (Coef) is 0.02 (positive), Hazard Ratio (exp(coef)) is 1.02 (2% increase in hazard per year), P-value: 0.04 (significant). It observed that as age increases, the hazard slightly increases. The effect is small but statistically significant. For Anaemia Coef is -0.21 (negative), Hazard Ratio is 0.81 (19% reduction in hazard, P-value is 0.38 (not significant) which implies Anaemia appears to reduce the hazard, but this effect is not statistically significant. For Serum Sodium Coef is -0.02 (negative), Hazard Ratio is 0.98 (2% reduction in hazard), P-value: 0.41 (not significant). It is observed Serum sodium has no significant impact on the hazard. For Ejection Fraction Coef is 0.00 (neutral), Hazard Ratio is 1.00 (no effect on hazard), P-value: 0.74 (not significant). Which implies that Ejection fraction does not influence the hazard in this analysis. For High Blood Pressure Coef is 0.45 (positive), Hazard Ratio: 1.56 (56% increase in hazard), P-value: 0.06 (marginally non-significant). It is observed High blood pressure shows a trend toward increasing the hazard but does not reach statistical significance. For Serum Creatinine Coef is 0.05 (positive), Hazard Ratio: 1.05 (5% increase in hazard), P-value: 0.58 (not significant). Implies Serum creatinine does not significantly affect the hazard. For Smoking Coef is -1.11 (negative), Hazard Ratio: 0.33 (67% reduction in hazard), P-value: <0.005 (highly significant). It is observed Smoking significantly reduces the hazard. This result is counterintuitive and might reflect confounding factors or survivor bias. For Region (Kaduna North) Coef is -1.09 (negative), Hazard Ratio: 0.34 (66% reduction in hazard), P-value: <0.005 (highly significant). It is observed Living in Kaduna North significantly reduces the hazard. Region (Kaduna Central) Coef is -0.61 (negative), Hazard Ratio: 0.54 (46% reduction in hazard), P-value: 0.04 (significant). Its observed Living in Kaduna Central is associated with a significant reduction in hazard, though the effect is less pronounced than Kaduna North. For Alcohol Usage (Yes) Coef is 3.07 (positive), Hazard Ratio: 21.49 (21-fold increase in hazard), P-value: <0.005 (highly significant). It is observed Alcohol usage dramatically increases the hazard, indicating it is a strong risk factor. For Body Mass Index (Overweight) Coef is 1.65 (positive), Hazard Ratio: 5.21 (5.2-fold increase in hazard), P-value: <0.005 (highly significant).this implies Being overweight significantly increases the hazard, suggesting it is a major risk factor. For Pulse Rate (Irregular) Coef is 2.21 (positive), Hazard Ratio: 9.08 (9-fold increase in hazard), P-value: <0.005 (highly significant). This implies that Irregular pulse rate is a significant and strong risk factor for increased hazard. For Diabetes Mellitus (Yes) Coef is -1.58 (negative), Hazard Ratio: 0.21 (79% reduction in hazard), P-value: <0.005 (highly significant).this implies that Diabetes mellitus significantly reduces the hazard, potentially reflecting the effects of medical management or other confounders.

In Summary, Strong Risk Factors are Alcohol usage, being overweight, and irregular pulse rate are the most significant predictors of increased hazard, Protective Factors are Living in Kaduna regions, smoking, and diabetes mellitus are associated with reduced hazard. Marginal Effects are Age and high blood pressure which show trends but have weaker significance. The Non-Significant Predictors are Anaemia, serum sodium, ejection fraction, and serum creatinine do not show significance.

The Cox proportional hazards model provides insights into how various covariates influence the hazard of the event occurring over time. Below is a focused discussion on the significant predictors and their implications:

3.3 *Strong Risk Factors*

1) Alcohol Consumption (HR = 21.49):

- The hazard ratio of 21.49 implies that patients who consume alcohol have a 21-fold increase in the risk of the event occurring compared to those who do not consume alcohol, holding all other factors constant. This finding highlights alcohol consumption as a critical and strong risk factor.
- Implications: This dramatic increase in hazard suggests that alcohol use likely exacerbates underlying health conditions, accelerates disease progression, or worsens the prognosis in cardiac patients. Intervention strategies targeting alcohol cessation could significantly improve patient outcomes and reduce mortality risks.

2) Irregular Pulse Rate (HR = 9.08):

- A hazard ratio of 9.08 indicates that patients with an irregular pulse rate have a 9-fold higher risk of the event occurring compared to those with a regular pulse rate.
- Implications: This result underscores the importance of monitoring cardiac rhythm abnormalities. Irregular pulse rate likely reflects underlying cardiac dysfunctions such as arrhythmias, which are strongly associated with poor outcomes. Prompt diagnosis and management of arrhythmias are crucial to mitigate this risk.

3) Body Mass Index (Overweight, HR = 5.21):

- Being overweight is associated with a 5.2-fold increase in the hazard compared to patients with normal or underweight BMI.
- Implications: This highlights obesity as a major risk factor for adverse cardiac events. Overweight individuals may face heightened cardiovascular strain, metabolic dysregulation, and inflammation, contributing to worse outcomes. Weight management interventions are essential to improve survival probabilities in this population.

3.4 *Protective Factors*

1) Smoking (HR = 0.33):

- Surprisingly, smoking is associated with a 67% reduction in hazard (HR = 0.33). While counterintuitive, this finding may reflect survivor bias or the presence of confounding factors, such as smokers who might have received aggressive medical interventions or other protective health measures.
- Implications: This result warrants cautious interpretation and further investigation. It may indicate that while smoking cessation is vital for long-term health, immediate survival outcomes in certain populations may depend on additional factors.

2) Region (Kaduna North, HR = 0.34; Kaduna Central, HR = 0.54):

- Patients residing in Kaduna North and Kaduna Central have a 66% and 46% reduction in hazard, respectively, compared to patients from Kaduna South.
- Implications: The protective effect of these regions could reflect differences in healthcare access, socioeconomic factors, lifestyle, or environmental exposures. Further studies are needed to identify the underlying factors contributing to these regional disparities.

3) Diabetes Mellitus (HR = 0.21):

- Diabetes mellitus is associated with a 79% reduction in hazard (HR = 0.21). This protective effect may result from medical management or survivor bias, as diabetic patients often receive more frequent medical attention and interventions.
- Implications: This finding highlights the potential benefits of intensive management for chronic diseases. It also suggests that well-controlled diabetes might mitigate the risks of adverse outcomes in cardiac patients.

3.5 Marginally Significant Factors

4) Age (HR = 1.02):

- A hazard ratio of 1.02 indicates a 2% increase in hazard per year of age, which is statistically significant (p = 0.04).
- Implications: While the effect of age is small, it reinforces that older patients face slightly higher risks. Age remains a background risk factor that interacts with other comorbidities to influence outcomes.

5) High Blood Pressure (HR = 1.56):

- A hazard ratio of 1.56 indicates a 56% increase in hazard, though the result is marginally non-significant (p = 0.06).
- Implications: This finding suggests that high blood pressure may play a role in worsening outcomes, though further research is needed to confirm its impact.

3.6 Summary of Implications

- Clinical Interventions: The results emphasize the need for targeted interventions, such as reducing alcohol consumption, managing irregular pulse rates, addressing obesity, and ensuring comprehensive management of diabetes and hypertension.
- Public Health Strategies: Regional differences highlight the importance of equitable access to healthcare and tailored interventions addressing local needs.
- Future Research: Unexpected findings, such as the protective effects of smoking and diabetes mellitus, warrant further investigation to identify potential confounders or biases.

Broader Implications and Practical Recommendations The findings from this study provide actionable insights into the practical measures that healthcare providers, policymakers, and patients can adopt to mitigate risks associated with cardiovascular conditions. Below are the broader implications, with a particular focus on lifestyle interventions for high-risk factors:

3.7 Targeting High-Risk Behaviors: Alcohol Consumption

- **Findings:** Alcohol consumption significantly increases the hazard (HR = 21.49), making it the strongest risk factor identified in this study.

1) Practical Implications:

- Clinical Practice: Healthcare providers should prioritize screening for alcohol use among cardiovascular patients and provide tailored counseling on its detrimental effects. Referrals to substance abuse programs or support groups could be beneficial for individuals struggling to reduce alcohol intake.
- Public Health Campaigns: Community-based awareness programs highlighting the risks of alcohol consumption, particularly in regions with high prevalence rates, could help reduce its impact on cardiovascular outcomes.
- Policy Recommendations: Enforcing stricter regulations on alcohol sales and advertisements, alongside increasing access to treatment for alcohol use disorders, could contribute to long-term public health improvements.

3.8 Addressing Overweight and Obesity

- Findings: Being overweight (HR = 5.21) is a significant predictor of increased hazard, emphasizing the need to address obesity as a critical risk factor.

1) Practical Implications:

- Lifestyle Interventions: Physicians and nutritionists should collaborate to design individualized weight management programs, including dietary modifications, exercise regimens, and behavioral therapies.
- Community Initiatives: Workplaces, schools, and local organizations can promote healthy lifestyles by creating opportunities for physical activity, such as fitness challenges or community exercise programs.
- Policy Recommendations: Policymakers should prioritize strategies to combat obesity, such as implementing taxes on sugary beverages, subsidizing healthier food options, and ensuring access to fitness facilities in underserved communities.

3.9 Strengthening Cardiac Monitoring for Irregular Pulse Rate

- Findings: Irregular pulse rate significantly increases the hazard (HR = 9.08), indicating the need for close monitoring and timely management of cardiac rhythm abnormalities.

1) Practical Implications:

- Clinical Monitoring: Regular electrocardiograms (ECGs) and the use of wearable cardiac monitors can help identify and manage arrhythmias early.
- Patient Education: Educating patients about recognizing symptoms of irregular pulse rates and seeking medical attention promptly could improve early detection and management.

3.10. Leveraging Regional Insights for Resource Allocation

- Findings: Living in Kaduna North and Kaduna Central is associated with reduced hazard, suggesting regional disparities in outcomes.

1) Practical Implications:

- Equitable Resource Distribution: Policymakers should investigate factors contributing to better outcomes in these regions and replicate successful practices in other areas.

- Strengthening Healthcare Access: Regions with higher hazards should receive increased funding for healthcare infrastructure and accessibility to ensure equitable care delivery.

3.11 Holistic Lifestyle Interventions

- Beyond alcohol and obesity, comprehensive lifestyle interventions addressing smoking, physical activity, and diet are essential to mitigate cardiovascular risks.
- Collaboration between healthcare providers, dieticians, fitness experts, and community leaders can create a supportive ecosystem for individuals aiming to adopt healthier lifestyles.

3.12 Comparison with Other Studies

The findings of this study align with and expand upon the research conducted by Musa et al. (2023), who applied both parametric and semi-parametric survival models to obstetric fistula data from low-resource settings. Similar to the current study, Musa et al. focused on identifying significant covariates influencing survival outcomes and evaluating the performance of different survival models.

2) Methodological Comparison:

While Musa et al. (2023) employed both Proportional Hazard (PH) and Accelerated Failure Time (AFT) models, this study utilized the Cox proportional hazard model, a semi-parametric approach, to analyze cardiovascular data. Both studies used survival analysis to identify covariates with significant impacts on patient outcomes, but Musa et al. also evaluated the performance of various parametric models (e.g., Weibull, Gompertz), concluding that the Weibull AFT model provided the best fit for their data.

3) Key Findings

- Risk Factors: Both studies identified critical risk factors associated with increased hazards. For instance, Musa et al. reported that prolonged labor and lack of antenatal care were significant risk factors for obstetric fistula, whereas this study identified alcohol consumption, irregular pulse rate, and obesity as the strongest predictors of increased hazard for cardiovascular events.
- Protective Factors: Musa et al. highlighted the protective effects of formal education and antenatal care, which mirror the protective role of factors such as living in certain regions (Kaduna North and Central) and diabetes mellitus in this study. Both findings underscore the importance of preventive healthcare measures and regional variations in survival outcomes.
- Model Selection and Performance: Musa et al. emphasized the importance of model selection by comparing AIC and BIC values across different approaches, concluding that the Weibull AFT model was the most suitable for obstetric fistula data. While this study did not directly evaluate parametric models, the Cox model's results provide a robust foundation for future research to explore models like Weibull, as recommended.
- Broader Implications: Both studies underscore the need for targeted interventions in low-resource settings. Musa et al. recommended antenatal care to reduce obstetric fistula risks, while this study emphasizes the importance of addressing lifestyle factors such as alcohol consumption and obesity to mitigate cardiovascular risks

4. CONCLUSION

This study employed the Cox proportional hazard model to analyze cardiovascular data sourced from Barau Dikko Teaching Hospital, Kaduna, with the goal of identifying significant covariates associated with patient survival. The findings revealed that diabetes mellitus, smoking, alcohol use, irregular pulse rate, age, body mass index, and region were significant predictors of survival at a 0.05 level of significance. Notably, alcohol usage posed the highest risk (Hazard Ratio: 21.49), followed by irregular pulse rate (Hazard Ratio: 9.08) and overweight body mass index (Hazard Ratio: 5.21), as corroborated by Figures 4, 6, and 8.

These results underscore the urgent need to prioritize interventions targeting high-risk groups, particularly individuals with alcohol consumption habits, smoking behavior, or irregular pulse rates.

Future Directions: To enhance predictive accuracy and broaden the applicability of the findings, future research could explore parametric models such as Weibull to provide insights into long-term survival trends. Additionally, expanding the dataset to include other regions would offer a comparative perspective and potentially generalize the findings across diverse populations.

REFERENCES

[1] Mendis, S., Puska, P. & Norrving, B. (2011). Global Atlas on Cardiovascular Disease Prevention and Control (PDF). World Health Organization url: http://apps.who.int/iris/handle/10665/44701

[2] Naghavi, M., Wang, H., Lozano, R., Davis, A., Liang, X. and Zhou, M. (2013). Mortality and Causes of Death Collaborators Global, regional, and national age-sex specific all-cause and cause specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study. Lancet. 385(9963);117–171. doi:https://doi.org/10.1016/s0140-6736(14)61682-2)

[3] Wang, Haidong, Mohsen Naghavi, Christine Allen, Ryan M. Barber, Zulfiqar A. Bhutta, Austin Carter, Daniel C. Casey et al. "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015." The lancet 388, no. 10053 (2016): 1459-1544. doi: https://doi.org/10.1016/S0140-6736(16)31012-1

[4] World Health Organization, author. 2023 report. doi: https://doi.org/10.1089/acm.2019.29077.jjw

[5] Adem, A., Demis, T. & Feleke, Y. (2011). Trend of diabetic admissions in Tikur Anbessa and St. Paul's University Teaching Hospitals from January 2005-December 2009, Addis Ababa, Ethiopia. Ethiopia Med J.; 49 (3);231-8. doi: https://doi.org/10.1210/jendso/bvae208

[6] Spoto, B., D'Arrigo, G., Tripepi, G., Bolignano, D. & Zoccali, C. "Serum gamma-glutamyltransferase, oxidized LDL and mortality in the elderly," Aging Clinical and Experimental Research, 33(5); 1393–1397. doi: https://doi.org/10.1007/s40520-019-01391-4

[7] Cox, D. R. (1972). Regression Models and Life Tables, Journal of the Royal Statistical Society, Series B, (Methodological), 34(2); 187-220. doi: https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[8] Musa A., Doguwa.,S & Yahaya A. (2023) comparative study on survival models for obstetric fistula patients in Nigeria. Sule lamido university journal of science & technology,7(3). doi: https://doi.org/10.56471/slujst.v7i.478

[9] Edward, H., Achmad, E., Geert, M. & Alain, G. B. (2015). Comparison of risks of cardiovascular events in the elderly using standard survival analysis and multiple-events and recurrent-events methods.BMC Medical Research Methodology. 15(15); 156-167 doi: https://doi.org/10.1186/s12874-015-0004-3.

[10] Kundu J.P. & Moeschberger, M. L. (2021). Survival Analysis: Techniques for Censored and Truncated Data, 2nd ed. New York: Springer. doi: https://doi.org/10.1007/s42081-021-00115-1

[11] Bart, K., Thomas, R., Thijs, V., Matteo, C. & Daan K. (2022). Accurate training of the Cox proportional hazards model on vertically-partitioned data while preserving privacy. BMC Medical Informatics and Decision Making 22(49). 186-193. doi: https://doi.org/10.1186/s12911-022-01771-3

[12] Alvaro, F. G. & René, B. (2021) Survival analysis of patients with heart failure in the Ecuadorian Andean population. Medwave; 21(07);84-94. url: https://repositorio.puce.edu.ec/handle/123456789/4790

[13] Angela, M. C. (2008). Comparison between Weibull and Cox Proportional Hazards Models. A Report submitted in partial fulfillment of the requirements for the degree. Master of Science Department of Statistics College of Arts and Sciences Kansas State University Manhattan, Kansas.

[14] Azme K., Che A., Che H., Mohd A. & Affend A. (2020). Modeling Students' Performance using Cox and Parametric Accelerated Failure Time Models. Scientific Research Journal (SCIRJ), 8(7); 201-279. doi: http://dx.doi.org/10.31364/SCIRJ/v8.i7.2020.P0720785

[15] Belaynesh, Y.E. & Zeytu, G.A. (2021). Comparison of survival models and assessment of risk factors for survival of cardiovascular patients at Addis Ababa Cardiac Center, Ethiopia: a retrospective study. National Library of Medicine. National Centre of Biotechnology Information. doi: https://doi.org/10.4314/ahs.v21i3.29

[16] Piotr, M. S. & Aaron, J. M. (2022). Scalable algorithms for semiparametric accelerated failure time models in high dimensions. Department of Statistics and Genetics Institute University of Florida, Gainesville, FL. doi: https://doi.org/10.1002/sim.9264

[17] Rana A. W. & Shuaib, M.K. (2006). Comparative Distributions of Hazard Modeling Analysis. Pak. j. stat. oper. res. L(2); 127-134. doi: https://doi.org/10.18187/pjsor.v2i2.96

[18] Samar, A. E., Graziella, D., Daniela, L., Maria, F., Giovanni, T. & Stefanos, R. (2021). Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. Hindawi Oxidative Medicine and Cellular Longevity. 13(11); 6. doi: https://doi.org/10.1155/2021/1302811

[19] Salil, V.D., Vaishali, D. & Varun, S. (2021). Cox proportional hazards model. Indian Journal of Thoracic and Cardiovascular Surgery 37(2); 229–233.