IJEMD-BMCR, 3 (1) (2025) https://doi.org/10.54938/ijemdbmcr.2025.03.1.421

International Journal of Emerging Multidisciplinaries:
Biomedical and Clinical Research

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2957-8620 ISSN (online): 2960-0731

Comparative Analysis of Accelerated Failure Time Models and the Assessment of Risk Factors Influencing Survival Time of Cardiovascular Patients: A Case Study in Kaduna, Nigeria

Enoch Yabkwa Yanshak [1*], Emmanuel Lekwot [1] , Kefas Rimamnuskeb Galadima [2]

1. Ahmadu Bello University, Zaria, Nigeria

2. Federal University, Wukari , Nigeria

Abstract

A primary focus of Survival analysis in medicine is modeling time to surviving of a particular disease. In this paper, survival analysis was carried out on the cardiovascular disease data modeling time to survive the disease. The data was gotten from Barau-Dikko teaching hospital Kaduna, Nigeria. Accelerated Failure Time (AFT) models like Weibull AFT model, Logistic AFT model, Log-normal AFT model,Log-logistic AFT model and Exponential AFT model are considered to be used for modeling the time to surviving cardiovascular diseases . Models selection criteria were used as a guide to unravel the best model for modeling cardiovascular diseases. The test for assumption of proportionality was conducted; the result revealed that the data violated the assumption of proportionality. Hence guarantee the use of accelerated failure time models. Based on the result from accelerated failure time models, the lognormal AFT model out-performed the other models since it has the lowest AIC and the highest log-likelihood value with 1022.23 and -47.82 respectively.

Keywords: Cardiovascular Disease: Accelerated Failure Time (AFT) Models: Weibull AFT: Exponential AFT: Survival Time

1. INTRODUCTION

Email addresses: enochyanshak@gmail.com (Enoch), lekwote@gmail.com (Emmanuel), kefasmanu3@gmail.com (Kefas)

Cardiovascular disease (CVDs) refers to a collection of heart and blood vessel disorders that poses a significant global health challenge. It encompasses a group of conditions affecting the vessels and the heart and exerts a major impact on global health, with 82% of fatalities related to CVD happening in countries with lower to middle income levels like Nigeria. Various cardiac medical conditions, such as deep vein thrombosis, cardiac embolism, peripheral arterial disease, cerebrovascular illness, and congenital heart disease contribute to this health burden [9]. In the present day, CVD has surpassed all other causes of death, emerging as the foremost global mortality factor [11]. However, the impact of CVD is not distributed equally; its types and prevalence vary between developed and developing countries. Surprisingly, in 2016, approximately 17.9 million individuals worldwide lost their lives due to cardiovascular diseases, constituting 31% Out of all deaths worldwide, over 75% of this fatalities occurred in low and middle-income nations [12]. Ethiopia has been affected by globalization, aging, and urbanization, leading to a rise in CVD occurrence and making it the primary cause of death in the country [2]. In order to address the mortality rate attributed to CVD, it is imperative to collect extensive data about how risk factors are distributed among different geographical and socio-economic segments within the population, indeed, the prevention of CVDs has always been a top priority. Nevertheless, it is equally essential for scholars and medical professionals to focus on strategies that can extend the lifespan of individuals already affected by CVD. Implementing appropriate interventions is crucial to decrease both the mortality and morbidity rates associated with CVD. Diligent efforts should be made by all potential stakeholders to point out the most significant risk variables that contribute to the death of Individuals with heart-related ailments. However, it is important to acknowledge that there will be variations between patients due to differences in biology, environment access to healthcare facilities, and variations in physician experience and commitment levels. These heterogeneities must be taken into account while developing tailored approaches to manage CVD effectively.

1.1 Problem Statement

Cardiovascular diseases remain a critical global health challenge, emphasizing the need for accurate survival predictions and the identification of key risk factors to enhance treatment strategies and patient care. Parametric survival models, such as [7], have been widely applied in medical research through time-to-event data analysis. In Nigeria, researchers have employed proportional hazard models and Accelerated Failure Time (AFT) models to study survival times for diseases like liver cirrhosis, breast cancer, lung cancer, kidney transplants, and obstetric fistula. However, their application to cardiovascular disease data remains largely unexplored. [14] highlighted risk factors such as age, gender, smoking status, diabetes mellitus, comorbidities, and treatment approaches in cardiovascular diseases. Similarly, Aishat et al. (2023) conducted a comparative survival analysis on obstetric fistula patients in Nigeria. This underscores the need to identify the most suitable Accelerated Failure Time models to investigate the factors influencing the survival time of cardiovascular patients effectively.

1.2 Objective

The aim of this study is to compare some survival models and assess the risk factors influencing the survival time of cardiovascular patients. The aforementioned aim will be achieved through the following objectives to:

i fit Weibull, loglogistic, and lognormal accelerated failure time models
ii assess the risk factors and investigate some associated factors with the survival time of cardiovascular patients.

Literature Review

This is the reviews of some widespread literature that is related to the area of study, including results and contributions made by several authors in the field of survival analysis that will serve as guide in carrying out this research.

[13] Analyzed cardiac arrest patients hospitalized at Pakistan's Institute of Cardiology and Allied Hospital (April–December 2015). Focusing on patients aged 40+ with left ventricular systolic dysfunction (NYHA classes III & IV), Cox regression was used to identify mortality risk factors. Key predictors included age, ejection fraction, renal impairment, anemia, and blood pressure. Kaplan-Meier plots showed high early mortality, with Martingale residuals and nomograms further refining predictions. [14] Assessed cardiovascular disease (CVD) risk factors in Nigerians with impaired fasting glucose (IFG) and diabetes mellitus, analyzing data from 2447 individuals aged 18–89. It found IFG had a higher prevalence (5.8%) than diabetes (3.1%) and highlighted comorbidities like dyslipidemia, obesity, and high blood pressure. Cholesterol and triglycerides emerged as significant risk factors. IFG was more strongly associated with CVD risks than diabetes. [6] At Addis Ababa Cardiac Center, compared survival models for 332 cardiac patients (2010–2018). Using parametric, semi-parametric, and non-parametric models, they found the Weibull AFT model outperformed others. Male patients had a 1.9 times higher mortality risk than females. The median survival time was 1925 days. [3] Conducted in Ecuador, the study analyzed heart failure survival among 228 patients (2015–2019). Using actuarial and Cox regression methods, they identified a five-year survival rate of 46%. Key predictors included age (HR: 1.035) and heart failure etiology. [1] Investigated hypertension survival in Ethiopia, the study analyzed 430 patients (2013–2019) using parametric AFT models, with the Weibull model performing best. Factors like age, residence, family history, hypertension stage, and cholesterol were significant, while gender, alcohol, and diabetes showed no association. [4] Compared Weibull and Cox models, the study found the Weibull model superior when its shape parameter was known. Both models yielded similar outcomes when the parameter was unknown. [8] In gastric cancer patients, Weibull and Cox models were compared. The Weibull model outperformed Cox in prognostic accuracy, identifying factors like tumor grade, stage, and surgical extent as critical predictors.

In addition, [15] Introduced the Weibull regression model, discussing its implementation using R for survival analysis. It emphasized variable selection, model adequacy checks, and visual presentations to enhance interpretability. [5] Modeled student performance via survival analysis, the study found the log-normal AFT model superior to the Cox model due to proportional hazards violations. Factors like GPA,

course, and entry qualifications were significant predictors. [2] studied the advanced scalable algorithms for high-dimensional semi-parametric AFT models. Their method improved computational efficiency and predictive accuracy in penalized rank-based criteria scenarios, outperforming traditional approaches.

2.. METHODOLOGY

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.

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.

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.

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.

Model Justification

The choice of Accelerated Failure Time (AFT) models over other survival models, such as the Cox Proportional Hazards (PH) model or non-parametric approaches like the Kaplan-Meier estimator, is driven by several key considerations:

1. **Focus on Survival Time:** AFT models directly model the survival time by quantifying how covariates accelerate or decelerate it. This makes them particularly useful when the interest is in understanding and predicting time-to-event outcomes, rather than hazard ratios, which are emphasized in PH models.
2. **Handling Non-Proportional Hazards:** AFT models do not rely on the proportional hazards assumption. This makes them an ideal choice when the effects of covariates on hazards are not constant over time, as was observed in this study with violations in the proportionality assumption for the Cox PH model.
3. **Interpretability:** AFT models provide a straightforward interpretation of covariate effects in terms of their multiplicative impact on survival time, which is often easier to communicate to practitioners and stakeholders compared to hazard ratios.
4. **Flexibility of Parametric Distributions:** AFT models allow for the specification of various parametric distributions (e.g., lognormal, Weibull, exponential), enabling a better fit to data with specific survival patterns. This flexibility can result in more accurate modeling and improved predictive performance.
5. **Model Comparison and Fit:** In this study, the lognormal AFT model demonstrated better performance in terms of model fit compared to the Cox PH model, as indicated by smaller AIC values.

Scenarios Where AFT Models Outperform Other Survival Models

1. **When Proportional Hazards Assumption is Violated:** Unlike the Cox PH model, AFT models are robust to non-proportional hazards, making them a preferred choice when this assumption does not hold.
2. **Focus on Time-Based Predictions:** AFT models are advantageous when the primary interest lies in quantifying how covariates affect the actual survival time rather than the hazard rate.
3. **Censored Data with Time-Dependent Effects:** AFT models can accommodate censored survival data while effectively modeling the influence of covariates with time-dependent effects.
4. **Small to Moderate Sample Sizes:** With appropriately chosen distributions, AFT models can provide efficient estimates even in studies with limited sample sizes.

Need for Parametric Assumptions: In scenarios where specific survival distributions (e.g., lognormal or Weibull) are believed to underlie the data, AFT models are better suited than semi-parametric models like the Cox PH model

Software

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

- **Lifelines**: For survival analysis (Weibull AFT, Exponential AFT,Loglogistics AFT and LogNormal AFT).

- **Pandas**: For data preprocessing and manipulation.
- **Matplotlib/Seaborn**: For visualization of survival curves.

Scikit-learn: For handling missing data and scaling continuous v

The hazard function of accelerated failure time model is expressed as:

$$h(t/x) = h_0[t\exp(-X'\beta^*)]\exp(-X'\beta^*) \quad (1)$$

The survival function is given as:

$$s(t/x) = \exp\{-H_0[t\exp(-X'\beta^*)]\} \quad (2)$$

The probability density function is given as:

$$f(t;x) = h_0[t\exp(-X'\beta^*)]\exp(-X'\beta^*)\exp\{-H_0[t\exp(-X'\beta^*)]\} \quad (3)$$

The time scale in AFT regression models is determined by the impact of variables in a technique that if $\exp(X'\beta^*) > 1$ the covariate vector has the effect of slowing the survival process, and if the $\exp(X'\beta^*) < 1$. The effect is to accelerate it [20].

This paper considered exponential, weibull, lognormal and log-logistics accelerated failure time models on secondary data sourced from the records and information sheets of cardiovascular patients treated at Barau Dikko Teaching Hospital in Kaduna state

2.1 The Exponential Accelerated Failure Time Regression Model

The exponential failure time model is a statistical model commonly used in reliability and survival analysis. It describes the time until an event of interest (such as failure or death) occurs for a particular unit or system. The key assumption of this model is that the hazard rate, which represents the instantaneous failure rate at any given time, is constant over time [18], [17].The hazard function is given as:

$$h(t/x) = h_0[\exp(X'\beta)] \quad (4)$$

Survival function is expressed as:

$$S(t:x,\beta^*) = \exp[-\exp(y - X'\beta^*)] \quad (5)$$

In terms of the extreme value distribution, which is provided by, the density function may be stated as:

$$f(t;x,\beta^*) = \exp[(y - X'\beta^*) - \exp(y - X'\beta^*)] \quad (6)$$

2.2 Weibull Accelerated Failure Time Regression Model

In this section, the idea of [21] has been applied using Weibull Accelerated Failure Time Regression Model**.** Thus, Survival function of weibull accelerated failure time model may be represented as;

$$s(t;x,\beta^*,p^*) = \exp[-\exp(\frac{y - X'\beta^*}{p^*})] - \infty < y < \infty \quad (7)$$

p^* is the scale parameter

the Weibull hazard function as an inverse in term of the AFT;

$$h(t:x,\beta^*,p^*)=(p^*)^{-}\exp(\frac{y-X^{'}\beta^*}{p^*}) \quad -\infty<y<\infty \tag{8}$$

The AFT density functions of the Weibull regression model may be directly stated as:

$$f(t;x,\beta^*P^*)=\exp[-\exp(\frac{y-X^{'}\beta^*}{p^*})](p^*)^{-1}\exp(\frac{y-X^{'}\beta^*}{p^*})(p^*)^{-1}\exp[(\frac{y-X^{'}\beta^*}{p^*})-$$

$$\exp(\frac{y-X^{'}\beta^*}{p^*})],-\infty<y<\infty \tag{9}$$

2.3 Lognormal Accelerated Failure Time Regression Model

To describe a monotonic risk function process, the lognormal distribution often used parametric function. Since the logarithm of a lognormal distribution is utilized when assuming that AFT survival times follow a log-normal distribution, it is simple, which contributes to its wide usage [19], Piotr & Aaron 2022*).* The baseline survival function and hazard function are provided by

$$S_0(t)=1-\phi(\frac{\log t-\mu}{\sigma}) \tag{10}$$

$$h_0(t)=\frac{\varphi(\frac{\log t}{\sigma})}{[1-\varphi(\frac{\log t}{\sigma})]\sigma t} \tag{11}$$

μ is an intercept, σ is scale parameter and is a random variable;

ϕ (x) is the cumulative density function of the standard normal distribution.

The i^{th} individual's survival function is

$$S_i(t)=s_0\left(\frac{t}{\eta_i}\right)=1-\phi(\frac{\log t-\alpha^i x_i}{\sigma}) \tag{12}$$

when $\eta_i=\exp(\alpha_1x_1+\alpha_2x_2+\ldots+\alpha_px_p)$. Consequently, i^{th} individual's log survival time has been normal $(\mu+\alpha^Ix_i,\sigma)$. The AFT property applies to the log-normal distribution

3. RESULT AND DISCUSSION

This section we test of PH assumptions for variables of the models, present analysis on AFT (WeibullAFT, LognormalAFT, and LoglogisticsAFT) . and as well present the models selection process and the model used for interpretation in this study.

Table 1: Proportional Hazard (PH) Assumption Test

		test_statistic	P	-log2(p)
Alcohol usage_yes	KM	9.01	<0.005	8.54
	Rank	7.07	0.01	7.00
Body mass index_over-weight	KM	8.67	<0.005	8.27
	Rank	8.71	<0.005	8.30
Diabetes militus_yes	KM	0.28	0.60	0.75
	Rank	0.39	0.53	0.91
Pulse rate_irregular	KM	7.22	0.01	7.11
	Rank	5.67	0.02	5.86
Region_kaduna north	KM	0.32	0.57	0.80
	Rank	0.27	0.60	0.74
Region_kaduna south	KM	0.47	0.49	1.02
	Rank	0.51	0.48	1.07
Age	KM	1.12	0.29	1.79
	Rank	1.15	0.28	1.82
Anaemia	KM	0.06	0.81	0.30
	Rank	0.09	0.76	0.39
ejection_fraction	KM	0.22	0.64	0.64
	Rank	0.32	0.57	0.80
high_blood_pressure	KM	0.32	0.57	0.81
	Rank	0.41	0.52	0.94
serum_creatinine	KM	0.35	0.55	0.85
	Rank	0.49	0.48	1.05
serum_sodium	KM	2.99	0.08	3.58
	Rank	3.67	0.06	4.17
Smoking	KM	0.11	0.74	0.43
	Rank	0.10	0.75	0.41

1. Variable 'ALCOHOL USAGE_YES' failed the non-proportional test: p-value is 0.0027.
2. Variable 'BODY MASS INDEX_OVER-WEIGHT' failed the non-proportional test: p-value is 0.0032.
3. Variable 'PULSE RATE_IRREGULAR' failed the non-proportional test: p-value is 0.0072.
As the software clearly state them with their respective p values

The test for assumption of proportionality in Table 1 clearly shows that some variables violated proportionality test:

- Variable 'ALCOHOL USAGE_YES' failed the non-proportional test: p-value is 0.0027.

- Variable 'BODY MASS INDEX_OVER-WEIGHT' failed the non-proportional test: p-value is 0.0032.

- Variable 'PULSE RATE_IRREGULAR' failed the non-proportional test: p-value is 0.0072.

Table 2 : Estimate from WeibullAFT

		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)
lambda_	Alcohol usage_yes	-2.090	0.124	0.305	-2.687	-1.493	0.068	0.225	0.000	-6.862	<0.0005	37.096
	Body mass index_over-weight	-1.139	0.320	0.184	-1.500	-0.779	0.223	0.459	0.000	-6.201	<0.0005	30.728
	Diabetes militus_yes	1.081	2.947	0.180	0.729	1.433	2.073	4.191	0.000	6.018	<0.0005	29.078
	Pulse rate_irregular	-1.560	0.210	0.228	-2.008	-1.113	0.134	0.329	0.000	-6.830	<0.0005	36.776
	Region_kaduna north	0.758	2.134	0.175	0.416	1.101	1.515	3.006	0.000	4.338	<0.0005	16.086
	Region_kaduna central	0.382	1.466	0.193	0.003	0.762	1.003	2.142	0.000	1.977	0.048	4.380
	Age	-0.013	0.987	0.006	-0.025	-0.000	0.975	1.000	0.000	-1.992	0.046	4.431
	Anaemia	0.148	1.160	0.155	-0.156	0.452	0.856	1.572	0.000	0.957	0.339	1.563
	ejection_fraction	-0.003	0.997	0.008	-0.018	0.012	0.982	1.012	0.000	-0.417	0.676	0.564
	high_blood_pressure	-0.310	0.734	0.152	-0.607	-0.012	0.545	0.988	0.000	-2.041	0.041	4.599
	serum_creatinine	-0.022	0.978	0.056	-0.131	0.087	0.877	1.091	0.000	-0.396	0.692	0.530
	serum_sodium	0.008	1.008	0.016	-0.024	0.039	0.977	1.040	0.000	0.489	0.625	0.678
	Smoking	0.768	2.156	0.171	0.433	1.104	1.542	3.016	0.000	4.487	<0.0005	17.080
	Intercept	7.023	1121.667	2.169	2.772	11.273	15.987	78695.093	0.000	3.238	0.001	9.698

The Table 2 has resulted fitted from Weibull accelerated failure time model as some of the predictor's variables such as alcoholic usage, high blood pressure, over body mass index, irregular pulse rate and many other are significant at 0.05 level while others are not.

Table 3: Estimate from LogLogistics

		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)
alpha_	Alcohol usage_yes	-1.64	0.19	0.22	-2.07	-1.20	0.13	0.30	0.00	-7.34	<0.005	42.12
	Body mass index_over-weight	-1.30	0.27	0.17	-1.64	-0.96	0.19	0.38	0.00	-7.57	<0.005	44.58
	Diabetes militus_yes	0.98	2.66	0.18	0.62	1.34	1.85	3.82	0.00	5.31	<0.005	23.13
	Pulse rate_irregular	-1.53	0.22	0.21	-1.93	-1.12	0.14	0.32	0.00	-7.42	<0.005	42.94
	Region_kaduna central	-0.49	0.61	0.63	-1.72	0.74	0.18	2.10	0.00	-0.78	0.04	1.20
	Region_kaduna north	0.37	1.44	0.18	0.02	0.71	1.02	2.03	0.00	2.09	0.04	4.78
	Region_kaduna south	0.33	1.39	0.19	-0.04	0.70	0.96	2.02	0.00	1.74	0.08	3.61
	Age	-0.01	0.99	0.01	-0.02	0.00	0.98	1.00	0.00	-1.44	0.15	2.73
	Anaemia	-0.01	0.99	0.15	-0.31	0.29	0.73	1.34	0.00	-0.07	0.94	0.08
	ejection_fraction	-0.00	1.00	0.01	-0.01	0.01	0.99	1.01	0.00	-0.14	0.89	0.17
	high_blood_pressure	-0.37	0.69	0.16	-0.67	-0.06	0.51	0.94	0.00	-2.36	0.02	5.77
	serum_creatinine	-0.05	0.95	0.06	-0.16	0.06	0.85	1.06	0.00	-0.89	0.38	1.41
	serum_sodium	0.02	1.02	0.02	-0.01	0.05	0.99	1.05	0.00	1.32	0.19	2.41
	Sex	-0.00	1.00	0.17	-0.34	0.34	0.71	1.40	0.00	-0.03	0.98	0.03
	Smoking	0.48	1.61	0.19	0.11	0.84	1.12	2.32	0.00	2.57	0.01	6.63
	Intercept	4.62	101.74	2.19	0.33	8.92	1.39	7462.97	0.00	2.11	0.03	4.84

Table 3 shows result for fitting loglogistic accelerated failure time model is interpreted as: Individuals with alcohol usage have a significantly lower odds (exp(coef) = 0.20) of the outcome compared to those without alcohol usage. Being overweight is associated with significantly lower odds (exp(coef) = 0.28) . Presence of diabetes is associated with significantly higher odds (exp(coef) = 2.66) of the outcome. Irregular pulse rate is associated with significantly lower odds (exp(coef) = 0.21) of the outcome. Belonging to the "Kaduna central" region is not statistically significant ($p > 0.05$). Belonging to the "kaduna north" region is associated with significantly higher odds (exp(coef) = 1.51) of the outcome. Belonging to the "kaduna south" region is not statistically significant ($p > 0.05$). Older age is associated with lower odds (exp(coef) = 0.99) of the outcome. Presence of anemia is not statistically significant ($p > 0.05$). Ejection fraction does not significantly impact the odds ($p > 0.05$). Having high blood pressure is associated with significantly lower odds (exp(coef) = 0.68) of the outcome. Serum creatinine level is not statistically significant ($p > 0.05$). Serum sodium level is associated with slightly higher odds (exp(coef) = 1.02) of the outcome. Gender is not statistically significant ($p > 0.05$). Smoking is associated with significantly higher odds (exp(coef) = 1.69) of the outcome.

Significance and Confidence Intervals:

The 'z' and 'P' columns provide the z-statistic and p-value for each coefficient, respectively. The 95% confidence intervals for the coefficients are given by "coef lower 95%" and "coef upper 95%. A p-value less than 0.05 is often considered statistically significant

Table 4: Estimate from LogNormalAFT

		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)
mu_	Alcohol usage_yes	-1.63	0.20	0.20	-2.02	-1.23	0.13	0.29	0.00	-8.12	<0.005	50.88
	Body mass index_over-weight	-1.27	0.28	0.17	-1.61	-0.93	0.20	0.39	0.00	-7.32	<0.005	41.92
	Diabetes militus_yes	0.98	2.66	0.17	0.64	1.32	1.89	3.74	0.00	5.61	<0.005	25.53
	Pulse rate_irregular	-1.55	0.21	0.20	-1.95	-1.16	0.14	0.32	0.00	-7.69	<0.005	45.91
	Region_kaduna central	-0.49	0.61	0.72	-1.90	0.92	0.15	2.51	0.00	-0.68	0.04	1.01
	Region_kaduna north	0.41	1.51	0.17	0.07	0.75	1.07	2.13	0.00	2.37	0.02	5.81
	Region_kaduna south	0.30	1.35	0.19	-0.06	0.67	0.94	1.94	0.00	1.62	0.10	3.26
	Age	-0.01	0.99	0.01	-0.02	0.00	0.98	1.00	0.00	-1.48	0.04	2.86
	Anaemia	0.04	1.04	0.15	-0.26	0.34	0.77	1.40	0.00	0.24	0.81	0.30
	ejection_fraction	-0.00	1.00	0.01	-0.01	0.01	0.99	1.01	0.00	-0.17	0.87	0.20
	high_blood_pressure	-0.39	0.68	0.15	-0.69	-0.10	0.50	0.91	0.00	-2.61	0.01	6.78
	serum_creatinine	-0.04	0.96	0.05	-0.15	0.06	0.86	1.06	0.00	-0.81	0.42	1.26
	serum_sodium	0.02	1.02	0.02	-0.01	0.05	0.99	1.05	0.00	1.28	0.20	2.32
	Sex	0.02	1.02	0.17	-0.32	0.36	0.73	1.43	0.00	0.11	0.91	0.13
	Smoking	0.53	1.69	0.18	0.17	0.88	1.19	2.40	0.00	2.94	<0.005	8.24
	Intercept	4.82	124.23	2.04	0.82	8.83	2.27	6813.86	0.00	2.36	0.02	5.77

The 4.is result fitted from LogNormal accelerated failure time model as some of the predictors variables such as alcoholic usage, high blood pressure, over body mass index, irregular pulse rate and many other are significant at 0.05 level while others are not which will be discuss in details in comparison.

3.1 Comparison of the AFT models using QQ plot, Log-likelihood and AIC

Quantile-Quantile Plot: A quantile-quantile plot was made to check whether the accelerated failure time model provide an adequate fit to the data set or not. We also checked the adequacy of the accelerated failure-time model by comparing the various categories, the result from Fig1 to Fig 4. Shown are

approximately linear for all covariates which is an indication that accelerated failure time models provide an adequate fit to the data set. Hence, the LogNormal accelerated failure time model has better performance since it is approximately more linear compare to other AFT models.

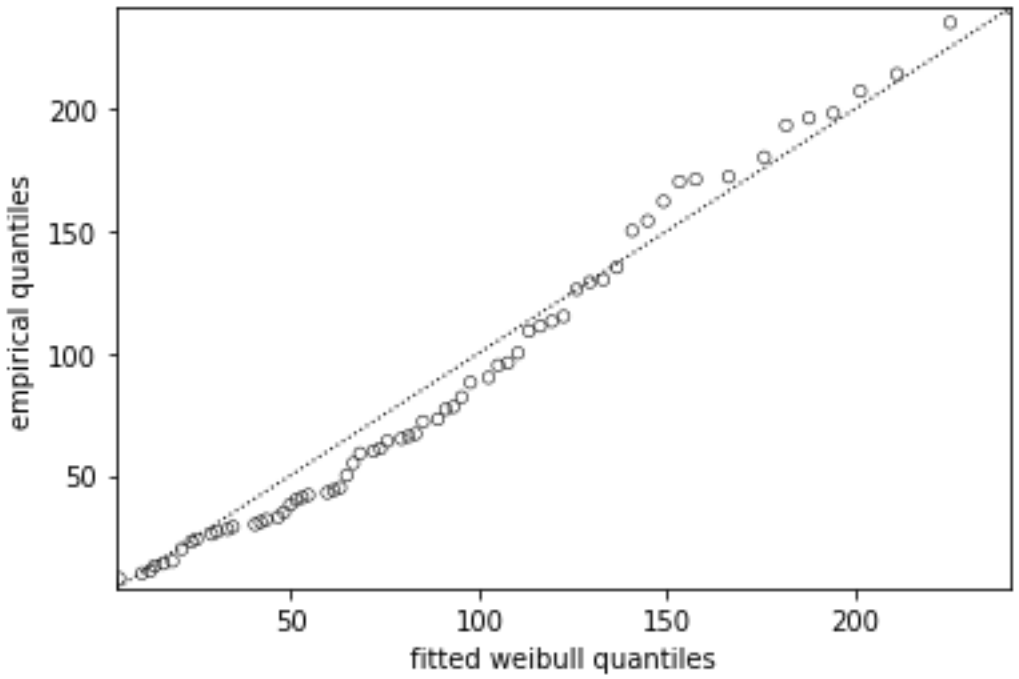

Fig.1. Weibull Quantiles

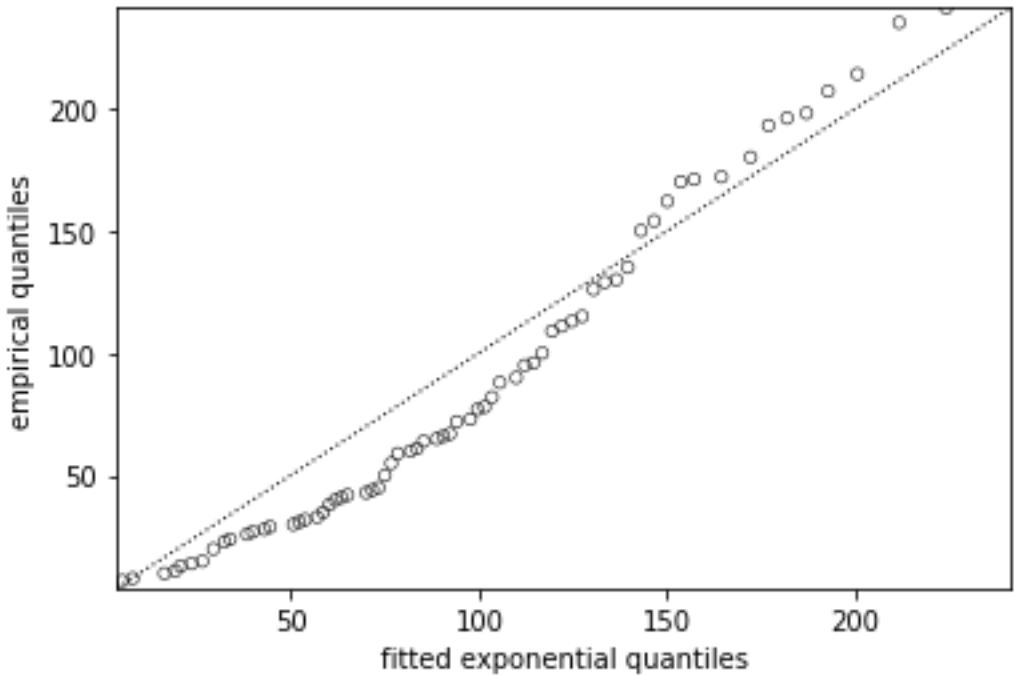

Fig.2. Exponential Quantiles

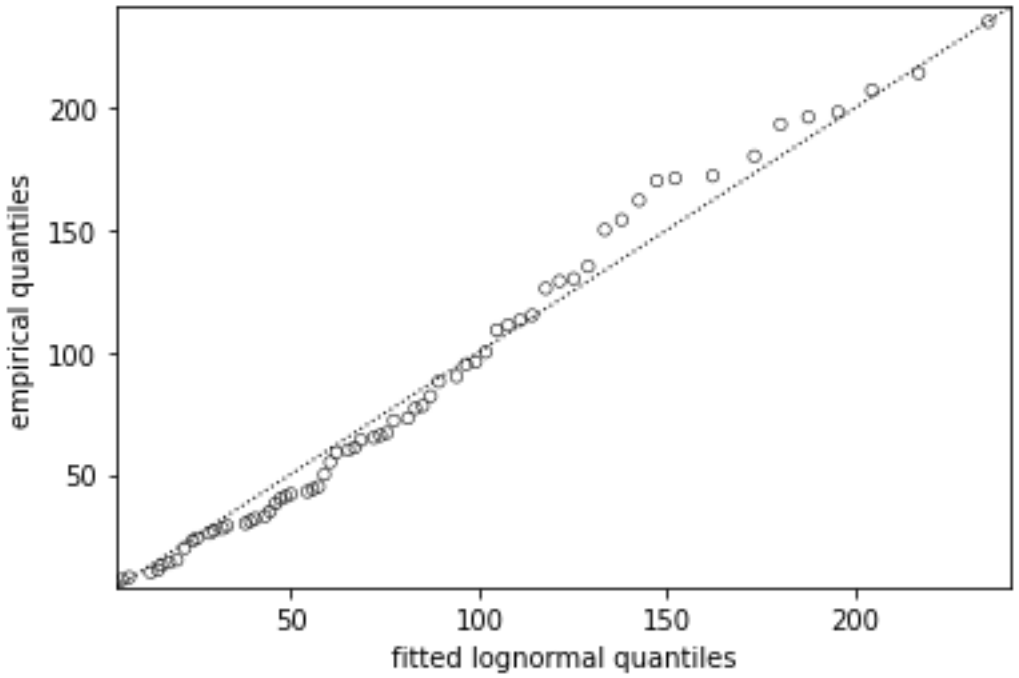

Fig.3.Lonormal Quantiles

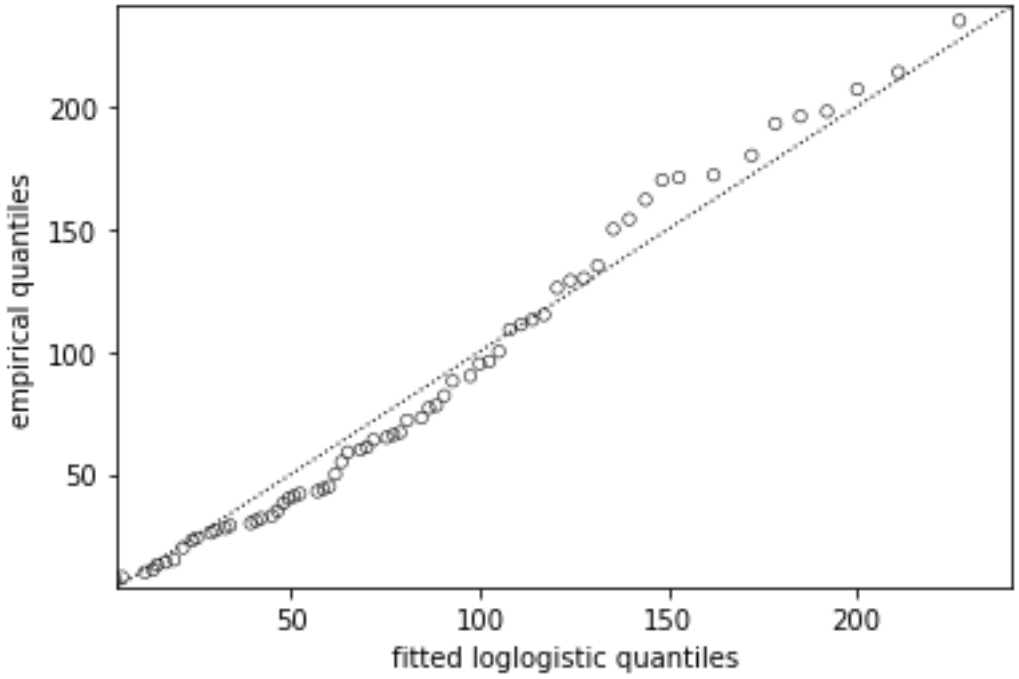

Fig.4.Loglogistic Quantiles

This sub-section also compares the Parametric AFT Models (Weibull, Exponential, Log-logistic and Log-Normal). The likelihood ratio, AIC and values are used to judge the best-fitting model and finally selected model will be used for interpretation.

Table 4.10: Selection of best fitted model

Model Type	Model	Obs	Loglik	Df	AIC
Accelerated Failure Time	Exp	299	-672.54	13	1346.08
	Weibull	299	-502.44	13	1034.89
	Log-Logistic	299	-496.11	15	1026.00
	Lognormal	299	-47.82	15	1022.23

Parametric Survival Models; QQ plot, AIC and Log-likelihood were used to identify the appropriate Accelerated failure time models and proportional hazard models among the widely considered survival models. Since the data violated the assumption of proportionality, cox proportional hazard model will not be used for any interpretation. Thus, we used LogNormal AFT survival model to determine predictors of CVD patient since it has smaller AIC and higher Loglikelihood among all the accelerated failure time models considered as shown in Table 4.10

We selected the appropriate parametric AFT model for the dataset among the exponential, Weibull, log-logistic, lognormal. For this conclusion, we selected the LN AFT model over the other fitted AFT models as we considered ICS in Table 4.10 for selecting the most appropriate model. The smaller magnitude of the information criterion statistics (ICS) for LN AFT model led us to the conclusion that the LN AFT model is the best fit for the data. Therefore, we used the LN AFT model to represent the parametric AFT models for comparison.

3.2 Discussion of Log-Normal AFT modeling

The results of LogNormal AFT model presented in Table 4.9 showed that explanatory variables, region, smoking, high blood pressure, body mass index ,irregular pulse rate, alcohol use , Diabetes mellitus and age have significant effect on survival of CVDs patients at 5% levels of significance but in cox PH model p value for high blood pressure is 0.06 which is not significance though very close to significance level. The Cox PH model identified to have smaller partial AIC but with violation of proportional assumption hence, any conclusion made from the model may not be accurate. LogNormal AFT modeling: A unit increase in covariate indicates that the ean/median survival time will change by a factor of exp(coefficient).

Coefficients (coef): These values represent the log-linear change in the log-transformed survival time for a one-unit change in the predictor variable, holding all other variables constant. In other words, it tells you the direction (positive or negative) and the magnitude of the effect of each predictor on the survival time.

Exponentiated Coefficients (exp(coef)): These values are the exponential of the coefficients and can be interpreted as the multiplicative effect on the survival time. They represent how much the survival time is expected to change when the predictor variable increases by one unit.

From Log-Normal AFT regression model; Alcohol usage yes has coefficient of about -1.63. Note that the higher hazard means more at risk of the event occurring. Here, the value of exp(-1.63) is called the hazard ratio. It shows that, the coefficient of -1.63 suggests that for the corresponding predictor variable, as it increases by one unit, the log-transformed survival time decreases by approximately 1.63 units. **0.20**: The exponentiated coefficient (exp(coef)) is 0.20, which means that a one-unit increase in the predictor variable is associated with a 0.20 (or 20%) decrease in the survival time. This implies that as this variable increases, the survival time is expected to decrease by 80%. BODY MASS INDEX_OVER-WEIGHT has Coefficient -1.27 and Exponentiated Coefficient be 0.28, this interpret that, Patients classified as overweight (BODY MASS INDEX_OVER-WEIGHT) have a lower expected survival time (exp(coef) = 0.28) compared to those with normal BMI. DIABETES MILITUS_YES has Coefficient of 0.98 and Exponentiated Coefficient to be 2.66 , This indicate that, the risk (rate) of dying is 2.66 times for Patients with diabetes (DIABETES MILITUS_YES) as (exp(coef) = 2.66) compared to those without diabetes. PULSE RATE_IRREGULAR has Coefficient of -1.55 and Exponentiated Coefficient: 0.21 which indicated that Patients with an irregular pulse rate (PULSE RATE_IRREGULAR) have a lower expected survival time (exp(coef) = 0.21) compared to those with a regular pulse rate in the case of region patients that come from KADUNA CENTRAL has Coefficient of -0.49 and the Exponentiated Coefficient: 0.61 which indicated that Patients from Kaduna Central (REGION_KADUNA CENTRAL) have a lower expected survival time (exp(coef) = 0.61) compared to the other two regions. REGION_KADUNA NORTH has Coefficient of 0.41 and Exponentiated Coefficient: 1.51 which indicated that Patients from Kaduna North (REGION_KADUNA NORTH) have a higher expected survival time (exp(coef) = 1.51) compare to Kaduna central. REGION_KADUNA SOUTH has Coefficient of 0.30 and Exponentiated Coefficient to be 1.35 this also indicated that Patients from Kaduna South (REGION_KADUNA SOUTH) have a higher expected survival time (exp(coef) = 1.35) compared to both Kaduna central and Kaduna north. AGE has Coefficient of -0.01 and Exponentiated Coefficient to be 0.99, this indicated that, one-year increase in age is associated with a very slight decrease in expected survival time (exp(coef) = 0.99). ANAEMIA has Coefficient of 0.04 and Exponentiated Coefficient be1.04 , this indicated that,Patients with anaemia have a slightly higher expected survival time (exp(coef) = 1.04) compared to those without anaemia.

EJECTION_FRACTION has Coefficient of -0.00 and Exponentiated Coefficient to be 1.00 this indicted that, the ejection fraction does not significantly affect the expected survival time (exp(coef) is approximately 1.00). HIGH_BLOOD_PRESSURE has Coefficient is -0.39 and Exponentiated Coefficient to be 0.68 this indicated that, Patients with high blood pressure have a lower expected survival time (exp(coef) = 0.68) compared to those without high blood pressure. SERUM_CREATININE has Coefficient of 0.02 and Exponentiated Coefficient to be 1.02, this indicated that one-unit increase in serum sodium is associated with a very slight increase in expected survival time (exp(coef) = 1.02). SERUM_SODIUM has Coefficient of 0.02 and Exponentiated Coefficient to be 1.02, this indicated that one-unit increase in serum sodium is associated with a very slight increase in expected survival time (exp(coef) = 1.02). SEX has Coefficient of 0.02 and Exponentiated Coefficient to be 1.02 this interpret

that, there is a very slight difference in expected survival time between the two sexes (exp(coef) = 1.02). SMOKING has Coefficient is 0.53 and the Exponentiated Coefficient to be 1.69, this indicates that, the risk (rate) of dying is 1.69 times for Patients who smoke (exp(coef) = 1.69) compared to non-smokers. INTERCEPT has Coefficient: 4.82 and the Exponentiated Coefficient to be 124.23. The intercept represents the baseline survival time when all predictor variables are at their reference levels. The exponentiated intercept is 124.23, which is the baseline expected survival time.

Broader Implications

The findings of this study have significant practical implications for improving the survival of cardiovascular disease (CVD) patients. Targeting high-risk groups with tailored interventions is crucial. For instance, the results highlight the adverse impact of alcohol use, smoking, and being overweight on survival time, with alcohol use and high BMI associated with an 80% and 72% decrease in survival time, respectively. These findings emphasize the need for lifestyle interventions, such as programs to reduce alcohol consumption, smoking cessation campaigns, and support for healthy weight management.

Additionally, patients with comorbidities like diabetes mellitus and irregular pulse rates face significantly higher risks of mortality, necessitating closer clinical monitoring and more aggressive management of these conditions. Region-specific disparities in survival times also suggest the need for localized health strategies, particularly for patients in Kaduna Central, where survival outcomes were notably poorer compared to other regions. Age-related risks further underscore the importance of early intervention and continuous care for older patients to improve their long-term outcomes.

By addressing these high-risk factors through targeted health policies and individualized patient care, healthcare providers can significantly enhance survival outcomes for CVD patients.

4. CONCLUSION

The study identified high-risk groups that require targeted interventions to improve cardiovascular disease outcomes. Specifically, patients in advanced clinical stages, those who are older, overweight, or have comorbidities like diabetes mellitus and irregular pulse rates, as well as individuals who consume alcohol or smoke, were found to have significantly reduced survival times. These findings emphasize the need for health workers to prioritize lifestyle interventions, such as promoting healthy weight management, alcohol reduction, and smoking cessation, while closely monitoring patients with these risk factors.

Future Directions

Future researchers are encouraged to explore the application of Weighted Least Squares Estimation (WLSE) to datasets with heteroscedastic properties and a larger set of covariates to test its validity in diverse settings. Additionally, validating the lognormal AFT model's performance on homoscedastic real-world datasets could further refine its utility in survival analysis. These efforts would provide deeper insights into the effectiveness of parametric AFT models for predicting survival outcomes across varying populations and conditions.

5. REFERENCES

[1] Abtsega, S., Ayalew, M., Abiso, E. & Kabtamu, T. G. (2019). Survival Analysis of Factor Affects Survival Time of Hypertension Patients. *Open Journal of Modelling and Simulation,* 7, 177-189.

[2] 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.

[3] Alvaro, F. G. & René, B. (2021) Survival analysis of patients with heart failure in the Ecuadorian Andean population. *Medwave*;21(07);84-94.

[4] 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*.

[5] 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.

[6] 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*.

[7] Cox, D. R. (1972). Regression Models and Life Tables, *Journal of the Royal Statistical Society, Series B, (Methodological),* 34(2); 187-220.

[8] Hui, P. Z., Xin, X., Chuan, H. Y., Ahmed, A., Shun, F. L., Yu, K. D. (2011). Application of Weibull model for survival of patients with gastric cancer. *BMC Gastroenterology 2011, 11:1 http://www.biomedcentral. com/ 1471-230X/11/1*.

[9] Mendis, S., Puska, P. & Norrving, B. (2011). Global Atlas on Cardiovascular Disease Prevention and Control (PDF). *World Health Organization in*

mortality for 249 causes of death, 1980-2015: *a systematic analysis for the Global Burden of Disease Study. Lancet.* 388(10053); 1459–1544.

[10] Musa Aishat, O., Doguwa, S., & Yahaya, A. (2023). Comparative Study on Survival Analysis Models for Obstetric Fistula Patients in Nigeria.

[11] 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.

[12] Ndejjo, R. (2021). Cardiovascular disease prevention in Mukono and Buikwe districts in Uganda: *Evidence to implementation*.

[13] Tanvir, A., Assia ,M., Sajjad ,H. B., Muhammad ,A., Muhammad ,A. R. (2017) Survival analysis of heart failure patients: *A case study. PLoS ONE* 12(7); 567-603.

[14] Victor, M. O., Ezekiel, U. N., Ifeoma, I. U., Adeseye, A. A., Ekene, E. C., Phillip, T. B., Ross, S. R. & Timothy, C. S. (2017). Cardiovascular disease risk factors in a Nigerian population with impaired fasting blood glucose level and diabetes mellitus. *BMC Public Health*.

[15] Zhongheng, Z. (2016). Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translational Medicine*;4(24);48-56.
[16] H., Naghavi, M., Allen, C., Barber, R.M., Bhutta, Z.A., & Carter, A. (2015). Mortality and Causes of Death Collaborators Global, regional, and national life expectancy, all-cause mortality, and cause-specific
[17] Austin, P. C., Steyerberg, E. W., & Putter, H. (2021). Fine-Gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: Cumulative total failure probability may exceed 1. *Statistics in Medicine,* 40(19), 4200–4212.
[18] Lili, Y., Liang, L. & Ding-Geng, C. (2018). A homoscedasticity test for the accelerated failure time model. *Electronic Supplementary Material*.
[19] Javeria K.,Muhammad A., & Zahra A.(2021). Influence Diagnostics in Log-Normal Regression Model with censored data. *Mathematical problem in engineering*,2021,02-15.
[20] Barnwal, A., Cho, H., & Hocking, T. (2022). Survival regression with accelerated failure time model in XGBoost. *Journal of Computational and Graphical Statistics*, 31(4), 1292–1302.
[21] Yesuf, A. M. & Ding-Geng, C. (2021). Accelerated failure-time model with weighted least-squares estimation: application on survival of HIV positives. *Archives of Public Health*, 9(3); 79-88.