

Comparative Study of Cox Proportional Hazard and Accelerated Failure Time Models on Survival of Diabetes Mellitus and Hypertension Patients

Aliyu Sani Aliyu ¹; Ibrahim A. Loko²; Nweze N. O. ²; Abubakar M. Auwal ²; Rabi Babangida ²

¹National Agency for Science and Engineering Infrastructure (NASENI). ²Department of Statistics, Nasarawa State University, Keffi, Nigeria.

DOI: <https://doi.org/10.62154/qjastr.2024.016.010417>

Abstract

Comparative analysis of survival models for hypertension and diabetics carried out in this study. The study compared the performance of Cox Proportional Hazard (CPH) and Accelerated Failure Time (AFT) models with and without random effects in analyzing hypertension and diabetes survival data obtained from General Hospital Keffi and General Hospital Nasarawa for the period of five years (2018-2022). Models evaluated include CPH, AFT (Exponential, Weibull, Log-Logistic) and their corresponding random effect counterparts. The Weibull AFT random effect model outperformed all other models in terms of goodness-of-fit, predictive accuracy and interpretability. This model accounted for individual heterogeneity and clustering effects, providing a more nuanced understanding of risk factors' impact on survival rates. The Weibull AFT random effects model is the most suitable choice for analyzing hypertension and diabetes survival data, offering improved predictive power and insight into the complex relationships between risk factors and time-to-event outcomes. These findings have significant implications for healthcare professionals, policymakers, and researchers seeking to optimize survival analysis methodologies.

Keywords: Survival Analysis, Hypertension, Diabetes, Weibull AFT, Random Effects, Model Comparison.

Introduction

Chronic diseases, such as diabetes mellitus and hypertension are increasing worldwide and they are associated with poor quality of life and increased economic burden; therefore, development of preventive measures against chronic diseases is imperative (Lahham, 2019).

In a nation with unstable economic conditions, the weight of DM and HTN has serious socioeconomic repercussions and reduces productivity. The country's health care systems will be severely strained if evidence-based preventive measures are not put into practice as soon as possible. This would lead to a significant rise in mortality and morbidity. Two studies have demonstrated how socioeconomic position affects the management and control of HTN in different contexts and among the general public (Gulliford et al., 2024).

Three main methods, namely, non-parametric, semi-parametric and parametric survival models are used to analyze survival data. Previously non-parametric techniques such as Kaplan and Meier (K-M) estimate of survival function for estimating the survival probability, and to assess difference of the survival experience across different groups, log-rank test (Shankar, et al., 2019) were found frequently used in clinical set up since these techniques do not require any distributional assumptions. The parametric hazards models such as exponential, Weibul, Gompertz; parametric survival models such as Lognormal, Log-Logistic, Gamma, e.t.c. have been generally used to analyzed time-to-event. The methods give similar results but each has its own unique procedure usually under specific assumptions or no assumptions (Shankar, et al., 2019). Cox Proportional Hazards (CPH) model is a semi-parametric survival model used for the analysis of time to event data since the baseline hazard function embedded in this model, does not require to follow any probability distribution. It means this model does not require to specify any probability distribution in order to quantify the effects of independent variables (Cox, 1972; Cox, 1975). Literatures on which method is most accurate among the three methods of survival analysis is still under battle especially in the studies of clinical data sets. Many researchers have proposed which methods fits well in their research studies based on the underlying assumptions. Parametric method assumed that the time to event variable comes from a specific distributional family. Based on asymptotic results, under certain circumstances, parametric models lead to more efficient parameter estimates (Shankar, et al., 2019). Vincent and Ismaila (2020) performed parametric survival analysis of Tuberculosis data collected from Federal Medical Centre Bida. Three AFT parametric survival (Exponential, Weibull and Log logistic) were fitted. It was found that the Weibull model performed better. Previous studies have applied Cox Proportional Hazard (CPH) and Accelerated Failure Time (AFT) models separately to analyze survival data. The studies inadequately addressed individual heterogeneity and clustering effects. Failed to compare CPH and AFT models with and without random effects.

The present study aims to:

- i. Compare the performance of CPH and AFT models with and without random effect.
- ii. Evaluate the impact of covariates on survival rates using different models.
- iii. Identify the most suitable model for hypertension and diabetes survival analysis.

Materials and Methods

Source of Data

The data used for the analysis of this study was sourced from records of hypertension and diabetic patients at General Hospital Keffi and General Hospital Nasarawa, Nasarawa State, Nigeria. The data consist of clinical records of patients attended on hypertension and diabetics in the two hospitals from 2018-2022.

Model Implementation

CPH and AFT (Exponential, Weibull, Log-logistic) models were fitted. The CPH focuses on relative risk, while AFT focusses on survival time. The CPH assumes proportional hazards, whereas AFT assumes accelerated failure time.

Cox Proportional Hazard Model

The Cox Proportional Hazards model is considered as the semi-parametric model. Cox emphasize more on the hazard function than the survival time. Suppose that the data taken on n subject are denoted by (t_i, δ_i, z_i) , where t_i is the failure time of the i^{th} subject, δ_i is the failure indicator such that for the i^{th} subject $\delta_i = 1$ if a subject is observed to failure and $\delta_i = 0$ otherwise and $z_i = (z_1, \dots, z_p)$ is p-dimensional vector of covariates. We can estimate the parameters β which reflect the effects of treatment and other covariates without having to make any assumptions about the form of $\lambda_0(t)$. In other words, we don't have to assume that $\lambda_0(t)$ follows an exponential model, or a Weibull model, or any other particular parametric model.

$$\text{Let } \lambda(t, z) = \lambda_0 \Psi(z) \quad (1)$$

We can write the second term as: $\Psi(z) = \exp(\beta z)$

Suppose $z = 1$ for treatment subject and $z = 0$ for un-treatment subject.

Then this model compares increase by a factor $\exp(\beta)$ for treatment subject verses un-treatment subject. The Cox proportional hazard model assume that the hazard function $\lambda(t)$ for the i^{th} subject with covariate value z_i has the following form

$$h(t/z) = h_0(t) \exp(\beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p) = h_0(t) \exp(\beta' z) \quad (2)$$

Where;

$\lambda_0(t)$ is an unspecified baseline hazard function reflecting the underlying hazard for subjects with all covariate $z_1, \dots, z_i$ equal 0. (i.e reference group) and β is a p -vector of unknown regression coefficient. The model 3.2 is referred as a semi-parametric hazard function because the dependence function $\exp(\beta' z)$ can be modeled explicitly but no specific probability distribution is assumed for the survival time.

For illustration purpose, consider a categorical variable with two levels (say $X = 1$, for treatment group and $X = 0$ for placebo), then the hazard ratio (HR), for th two groups is define as:

$$HR = \frac{h(t|x=1)}{h(t|x=0)} \quad (3)$$

When $HR = 1$, then the individuals in the two categories are at the same risk of getting the event, $HR > 1$, the individuals in the first group ($X=1$) are at higher risk of getting the event and if $HR < 1$, the individuals in the second group ($X = 0$) are at a high risk of getting the event.

Accelerated Failure Time Model

The parametric models use the Accelerated Failure Time (AFT) model for estimating univariate survival censored data. In the AFT model more attention is given to the survival time than to hazard function. The AFT model means that the survival function of an individual with covariate X at time t is the same survival function of an individual with a baseline survival function at a time $\exp(\beta^t X)t$ where β^t is a vector of regression coefficients. It means:

$$S(t|X) = S_0[\exp(\beta^t X)t] \text{ for all } t \quad (4)$$

where $\exp(\beta^t X)$ is called an acceleration factor. The acceleration factor allows to evaluate the effect of predictor variables on survival time. That is, how a change in covariate alters the time scale from the baseline time scale. It also implies that the median time to event with covariate X is the baseline median time to event divided by its acceleration factor. Similarly, the hazard function is also calculated. It is related to the baseline hazard rate.

$$\lambda(t|X) = \lambda_0[\exp(\beta^t X) t] \exp(\beta^t X), \text{ for all } t \quad (5)$$

In general, for linear regression model, the response variable is given by

$$T_i = \mu + \beta' X_i + \sigma \epsilon_i \quad (6)$$

If the response variable T_i is now replace by $\log T_i$, the resultant model is called the AFT model. That is:

$$\log T_i = \mu + \beta' X_i + \sigma \epsilon_i \quad (7)$$

Where;

X_i 's are the independent variable's (covariates)

β 's are constants to be determined and

ϵ_i s the error term

σ is the scale parameter.

μ is the intercept parameter

The role of the covariates in the above equation is to accelerate (or decelerate) the time to failure. The error term ϵ_i , is assumed to be independent and identical distributed with mean zero. The various choice of ϵ_i distribution led to the regression version of different

parametric survival models. For the log transformation of T , its main purpose is to ensure that predicted values of survival time T , are positive, regardless of the values of the x 's and the β 's. If a parametric model has just n scale parameter λ , that parameter is replaced by:

$$\lambda = \exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n) = \exp(\alpha_0) \cdot \exp(\alpha_1 x_1) \cdot \exp\left(\frac{1}{p} \epsilon\right) \quad (8)$$

The general form of the AFT model in terms of the survival time T is

$$T = \exp(\alpha_0 + \alpha_1 x + \dots + \frac{1}{p} \epsilon) \quad (9)$$

AFT model is multiplicative in terms of T and additive in terms of $\log(T)$. One assumption of the parametric model is that the survival time is assumed to follow a distribution with density function $f(t)$. The AFT survival models considered in this study are: Exponential, Weibull, Log-Normal and Log-logistic survival distributions.

The Weibull AFT Model

Survival time t is a positive random variable with Weibull probability density function can be expressed as:

$$f(t; \mu, \alpha) = \frac{\alpha}{\mu} \left(\frac{t}{\mu}\right)^{\alpha-1} \exp\left[-\left(\frac{t}{\mu}\right)^\alpha\right] \quad (10)$$

where, $\mu > 0$ and $\alpha > 0$ and the baseline hazard function of the distribution becomes:

$$h(t; \mu, \alpha) = \frac{\alpha}{\mu} \left(\frac{t}{\mu}\right)^{\alpha-1} \quad (11)$$

This yields the following survivorship functions: $S(t) = \exp\left[-\left(\frac{t}{\mu}\right)^\alpha\right]$ and the cumulative hazards function becomes $H(t) = \left(\frac{t}{\mu}\right)^\alpha$

Depending on the value of α , the hazard function can increase or decrease with increasing survival time. Hence the Weibull model can yield an accelerated failure time model. Independent observations (t_i, δ_i) , $i = 1, \dots, n$ with survival time t_i and censoring indicator δ_i which has value of one if i^{th} observation is not censored and zero when the i^{th} observation is censored and let β be the unknown parameter. The likelihood function is:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n ([f_i(t_i)]^{\delta_i} [S_i(t_i)]^{1-\delta_i}) \\ &= \prod_{i=1}^n \left\{ \left(\frac{f_i(t_i)}{S_i(t_i)} \right)^{\delta_i} S_i(t_i) \right\} \end{aligned}$$

$$\prod_{i=1}^n \{h(t_i)^{\sigma_i} S_i(t_i)\}$$

$$\prod_{i=1}^n \left\{ \left(\frac{\alpha}{\mu} \left(\frac{t}{\mu} \right)^{\alpha-1} \right)^{\sigma_i} \exp \left[\left(-\frac{t}{\mu} \right)^{\alpha} \right] \right\} \quad (12)$$

Re-parameterizing the Weibull distribution using $\lambda = \mu^{-\alpha}$ then $h_0(t) = \lambda \alpha t^{\alpha-1}$ will be the baseline hazard function. Now incorporate covariates X in the hazard function, the Weibull regression models become:

$$h(t; \mu, \beta) = \lambda \alpha t^{\alpha-1} \exp(X\beta) \quad (13)$$

The model assumes that individual i and j with covariates x_i and x_j have proportional hazard function of the form: $\frac{h(t; x_i)}{h(t; x_j)} = \frac{\exp(x_i \beta)}{\exp(x_j \beta)} = \exp((x_i - x_j)' \beta)$

The Exponential AFT Model

The time data is skewed to the right with exponential distribution, the time of survival for a set of covariates X , which is called, accelerated failure time is expressed as:

$$T = \exp(\mu + \beta'X + \sigma\varepsilon) \quad (14)$$

Where ε is the error component.

The exponential model ($t \sim \exp(\alpha)$) is the simplest parametric model and assumes a constant risk or hazard over time, which reflects the property of the distribution appropriately called "lack of memory" because the hazard function $h(t) = \alpha$ does not depend on time. Hence the probability of failure in a time interval $[t, t+\delta]$ does not depend on previous interval. The survivorship function may be obtained by expressing in terms of time as: $(t, X_i) = e(-te^{-\beta'X})$ and the hazard function of the exponential regression model is $h(t, X, \beta) = e^{-\beta'X}$.

The Log-Logistic AFT Model

Multiple covariate log-logistic accelerated failure time may be expressed as:

$$\log T = \exp(\beta'X + \sigma\varepsilon) \quad (15)$$

Where σ is the scale parameter and ε is the residual (unexplained) variation in the transformed survival time. The survivorship function for the model in (3.17) is $s(t, X, \beta, \sigma) = ([1 + \exp(z)]^{-1})$

Where z is the standardized log-time outcome variable, that is; $z = \frac{(y - \beta_0 - \beta_i X)}{\sigma}$ and $y = \ln(t)$

The odds of a survival time of at least t are, $OR = \frac{s(t,x,\beta,\sigma)}{1-s(t,x,\beta,\sigma)} = \exp(-z)$, assumes that the covariate is dichotomous and coded 0 or 1. The odds- ratio at time t from the ratio the odds of a survival time evaluated at $x=0$ and $x=1$ is:

$$OR(x=1, x=0) = \frac{\exp\left[\frac{-(y-\beta_0-\beta_1x_1)}{\sigma}\right]}{\exp\left[\frac{-(y-\beta_0-\beta_1x_0)}{\sigma}\right]} = \exp\left(\frac{\beta_1}{\sigma}\right) \quad (16)$$

This is independent of time.

The Log-normal AFT Model

The log-normal model may take censored time dependent variable that allows the hazard rate to increase and decrease. The log-normal model assumes that $\varepsilon \sim N(0, 1)$. Let $h(t)$ be the hazard function of T for the model (3.17) when $\beta=0$ i.e. $\beta_0 = \beta_1 = \dots = \beta_p = 0$. Then, it can be shown that $h(t)$ has the following functional form:

$$h(t) = \frac{\phi\left(\frac{\log(t)}{\sigma}\right)}{[1-\Phi\left(\frac{\log(t)}{\sigma}\right)]\sigma t} \quad (17)$$

where $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$ is the probability density function, and $\Phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$ is the cumulative distribution function of the standard normal distribution. Obviously, we no longer have a proportional hazard model. If the baseline hazard function is desired, it can be obtained from equation (3.20) by setting $x=0$. The survival function (t/X) at any covariate x can be expressed as:

$$s(t/X) = \phi[\beta_0^* + \beta_1^*x_1 + \dots + \beta_p^*x_p - \alpha \log(t)] \quad (18)$$

Where $\alpha = \frac{1}{\sigma}$, $\beta_j^* = \frac{\beta_j}{\sigma}$ for $j = 0, 1, \dots, p$

This is the final survival model with intercept depending with t .

Results and Discussion

Analysis of Hypertension Dataset

Table 4.1: Results of Model Selection Criterion

Model	No RE		GAMMA RE		Inverse Gaussian RE	
	AIC	BIC	AIC	BIC	AIC	BIC
CPHM	3932.4	3999.3	3931.9	3998.7		
Exponential	870.2	940.9	872.158	946.9	872.158	946.9
Weibull	582.7	657.4	512.8	591.5	546.1	624.7
Lognormal	521.8	596.5	521.5	600.1	521.8	600.4
Log logistic	513.1	595.8	518.8	596.5	568.8	691.5

Source: STATA output

In order to compare the performance of the conventional cox proportional hazard model with cox proportional hazard model with random effect (RE). The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values were used. The result revealed that the cox proportional hazard model with random effect parameter has least value of AIC and BIC as compared to the conventional cox proportional hazard model (without random effect parameter). This suggest that the cox proportional hazard model with random effect parameter outperformance the conventional cox proportional hazard model. Hence, the researcher based the interpretation of the result based on the cox proportional hazard model with random effect parameter.

The performance of the AFT model with and without the random effect parameter were also compared using the AIC and BIC values of the respective models. The result also indicated that the Weibull AFT model with Gamma random effect distribution has the least AIC and BIC. This suggest that Weibull AFT model with Gamma random effect performance better than the exponential, log-logistics and lognormal models. Hence, it was considered as the best AFT model for predicting the survival time for Hypertension patient. This implies that Weibull Gamma random effect AFT model it is more efficient model to describe the determinant factors of time-to-event of hypertension patient. Hence, the study focused on the interpretation of the Weibull Gamma random effect AFT model. Lastly, on the general performance, between CPHM and AFT model, the AFT model with RE performed better than the CPHM.

Table 4.2: CPHM for HTN Data with RE Component

Variable/Category	B	HR	S. E	P-value
Diagnosis				
HTN only	<i>Ref.</i>			
HTN & Others	0.0831	1.0866	0.1610	0.575
Drugs				
No	<i>Ref.</i>			
Yes	-0.8538	0.4258	0.1134	0.001***
Occupation				
Student	<i>Ref.</i>			
Self-employed	-0.0008	0.9992	0.3130	0.998
Employed	0.0530	1.0544	0.4243	0.895
Alcohol Intake				
Yes	<i>Ref.</i>			
No	-0.0741	0.9286	0.1065	0.519
Family Type				
Negative	<i>Ref.</i>			
Positive	0.0515	1.0529	0.1129	0.630
Blood Group				
o ⁺	<i>Ref.</i>			
o ⁻	0.1172	1.1244	0.2041	0.518
A ⁺	0.0618	1.0638	0.2104	0.754
A ⁻	0.1181	1.1254	0.2099	0.527
B ⁺	0.0253	1.0256	0.1917	0.893
B ⁻	-0.1922	0.82514	0.1933	0.412
AB	0.2698	1.3097	0.2942	0.230
C ⁺	0.1156	1.1226	1.1412	0.909
D ⁻	0.1610	1.1747	0.8539	0.825
BMI				
Under weight				
Normal weight	0.2471	1.2803	0.3189	0.321
Overweight	0.0249	1.0252	0.1772	0.886
Obese	0.2650	1.3034	0.1659	0.037**
θ		0.0093	0.0192	
τ		0.0046		
Likelihood ratio (θ): $\chi^2 = 0.53$; prob. = 0.234				

Source: STATA output

As stated earlier that the cox proportional hazard model with random effect outperformance the conventional cox proportional model, the interpretation of the result was based on the selected model. The random effect in the cox proportional hazard model

is assumed to follow the Gamma distribution with mean 1 and variance equal to θ . The heterogeneity in the population of the study which is used as a cluster as estimated by the selected model is $\theta = 0.0093$ and the dependence within the clusters (hospitals) is measured by Kendall's tau is $\tau = 0.0046$. A variance of zero ($\theta = 0$) indicate that the random effect component does not contribute to the model. A likelihood ratio test for the hypothesis $\theta = 0$ indicates a chi-square χ^2 value of 0.53 resulting to an insignificant p value of 0.234. This implied that the random effect component had insignificant contribution to the model.

The Hazard Ratio (HR) which is the $\exp(B)$ gives the hazard ratio between two groups with $HR < 1$ indicating reduction in hazard, $HR > 1$ indicating increase in hazard and $HR = 1$ indicating no hazard.

The column containing the estimated parameters provides the estimate of the regression parameters with interpretation that the estimated coefficient of $B_{i/s}$ gives the expected changes of the log hazard ratio for every one unit increase in the corresponding covariates $X_{i/s}$ when all other covariates are held constant. In the selected model (Cox proportional hazard model with random effect), the categorical variables drugs and BMI were significant factor implying that drugs and BMI were the two factors in this study that significantly contributed to the hazard of hypertension patient. The estimated values for patient that were giving drugs is ($B = -0.8538$, $HR = 0.4258$, $S.E. = 0.1134$, $p\text{-value} = 0.001$). The negative sign of the coefficient indicates negative relationship between the covariates and the event (death) from hypertension. Hence, the negative coefficient for patient with drugs indicates that patients that were giving drugs were at less risk of dying from hypertension than those that were not giving drugs. Those with drugs were at lower risk of dying with hypertension by a factor 0.4258 (42.6%) times lower than hypertension patient that were not given drugs when other covariates are held constant.

Also, the estimated coefficient for the obese hypertensive patient was positive and significant ($B = 0.2650$, $HR = 1.3034$, $p\text{-value} = 0.037 < 0.05$). This implies that hypertension patient that are obese are at a higher risk of dying with hypertension than those that are under weight. The Hazard Ratio (1.3034) indicates that hypertension patient that are obese are at higher risk of dying with hypertension by a factor 0.3034 (30.3%) times higher than those with underweight.

However, the variable BMI, the patient with normal and overweight were not significant using underweight as the reference category ($B = 0.2220$, $HR = 1.2486$, $S.E. = 0.3104$, $p\text{-value} = 0.372$ and $B = -0.0172$, $HR = 0.9829$, $S.E. = 0.1676$, $p\text{-value} = 0.9190$). The estimated coefficient for patient that were giving drugs was -0.8638 indicating the expected change of the log hazard ratio for every one-unit increase in patient with drugs when other covariates are held constant. The other covariates Diagnosis, occupation, alcohol intake, family type and blood group were insignificant.

Table 4.3: Weibull AFT Models for HTN Data with Gamma RE component

Variable/Category	B	TR	S. E	P-value
Intercept	0.1908	1.2102	0.2382	0.423
Diagnosis				
HTN only				
HTN & Others	-0.0116	0.9885	0.0645	0.857
Drugs				
No				
Yes	0.7712	2.1624	0.1237	0.000***
Occupation				
Student				
Self-employed	0.1052	1.1109	0.1641	0.521
Employed	0.0906	1.0948	0.1977	0.647
Alcohol Intake				
Yes				
No	0.0338	1.0344	0.0482	0.484
Family Type				
Negative				
Positive	-0.0002	.9998	0.0469	0.997
Blood Group				
O ⁺				
O ⁻	-0.1112	0.8948	0.0751	0.139
A ⁺	-0.0530	0.9484	0.0861	0.538
A ⁻	-0.0370	0.9637	0.0809	0.647
B ⁺	0.0379	1.0386	0.0802	0.637
B ⁻	0.1267	1.1351	0.1027	0.217
AB	-0.1871	0.8294	0.0939	0.046**
C ⁺	0.0430	1.0439	0.3623	0.905
D ⁻	-0.0571	0.9445	0.2812	0.839
BMI				
Normal weight				
Under weight	-0.0930	0.9112	0.1130	0.411
Overweight	0.0251	1.0254	0.0754	0.739
Obese	-0.1079	0.8977	0.0541	0.046**
μ	1.4215		0.0897	0.000***
ρ	4.1434		0.3719	
θ	1.1259		0.2311	
τ	0.3602			
Likelihood ratio (θ): $\chi^2 = 71.86$; prob. = 0.000***				

Source: Extracted from STATA output

Above presents the summary results of analysis of the selected AFT models (Weibull Gamma random effect AFT model). As reported earlier that the Weibull Gamma random effect AFT model appears to be appropriate model compared with other models since the model had a minimum AIC and BIC values of 512.8 and 591.5. The implication of this is that the Weibull Gamma random effect AFT model is more efficient model to describe determinant factors of time-to-event of hypertension patient. The random effect parameter in this model is assumed to follow a Gamma distribution with mean 1 and variance equal to theta (θ). The heterogeneity in the population of the hospitals which were used as a cluster are estimated by the selected model is $\theta = 1.1259$ and the dependence within the clusters (hospitals) is measured by Kendall's tau is $\tau = 0.3602$. A variance of zero ($\theta=0$) would indicate that the random effect component does not contribute to the model. A likelihood ratio test for the hypothesis $\theta = 0$ indicates a chi-square χ^2 value of 71.86 with resulting to a highly significant p value of 0.000. the implication of this findings is that, the random effect component had significant contribution to the model. The estimated Kendall's tau ($\tau = 0.3602$) shows that there is strong dependence within the cluster for Weibull Gamma random effect model.

The estimate of shape parameter in the Weibull Gamma random effect AFT model is $\mu = 1.4215$. This value shows the shape of hazard function is unimodal because the value is greater than unity i.e., it increases up to some time and then decreases. The Weibull with Gamma random effect distribution shows that drugs, blood group and BMI are statistically significant (p-value < 0.05) for hypertension patient. Whereas the diagnosis, occupation, alcohol intake and family type were found to be statistically insignificant factor for hypertension. The hypertension patients that were giving drugs have had longer survival by a factor 2.16 than those that were not giving drugs ($TR > 1$). Patient with blood group AB had lesser survival than those with blood group o* ($TR < 1$). Furthermore, the result of the analysis also revealed that Obese hypertensive patient had lesser survival time as compared to those with normal weight ($TR < 0.05$). This implies that Drugs, blood group and BMI are the risk factors of hypertension.

Application to Diabetes Mellitus Dataset

Table 4.4: Model Selection Criteria for Diabetes Data

Model	No RE		GAMMA RE		Inverse Gaussian RE	
	AIC	BIC	AIC	BIC	AIC	BIC
CPH	2578.9	2691.8	2571.5	2681.6		
Exponential	1005.1	1131.6	1007.1	1138.1	1007.1	1138.1
Weibull AFT	1111.7	1242.7	864.5	999.9	864.3	999.8
Lognormal	1054.1	1185.1	1056.1	1191.6	1056.1	1191.6
Log logistic	1055.7	1186.8	1057.7	1193.2	1057.7	1193.2

Source: STATA output

In order to compare the performance of the Cox proportional hazard model with and without random effect, the AIC and BIC values for the two models were used. The cox proportional hazard model with random effect were had the smallest AIC and BIC. This suggest that the Cox proportional hazard model with random effect performance better than the conventional cox proportional hazard model (without random effect). Hence, it was considered as the best semi-parametric model in this study. Hence, the researcher focused on the interpretation of the cox proportional hazard model with random effect. In addition, the performance of the AFT model with and without the random effect was also considered. The result of the analysis revealed that the Weibull AFT Inverse Gaussian random effect model has the least AIC and BIC hence, it was concluded that it outperformed the exponential, log-logistics and lognormal models with and without the random effect when predicting the survival time of diabetes patient. Hence, the researcher based the interpretation of the results for AFT models on Weibull AFT Inverse Gaussian random effect model. Compare the overall performance of the CPH and AFT models using diabetic data set, it was found out that Weibull AFT Inverse Gaussian random effect model performed better than the cox proportional hazard model using the AIC and BIC values.

Table 4.5: CPH Model for Diabetes Data with Random Effect

Variable/Category	B	HR	S. E	P-value
Gender				
Male	<i>Ref</i>			
Female	-.1253	0.8822	0.1533	0.471
Residential Area				
Urban	<i>Ref</i>			
Peri-Urban	.7864	2.1955	0.6534	0.008***
Rural	.4901	1.6325	0.3759	0.033**
Education				
Tertiary	<i>Ref</i>			
Secondary	.9448	2.5722	0.7278	0.001***
Primary	.3445	1.4113	0.2864	0.050
Non-Formal	.4431	1.5576	0.5002	0.168
Occupation				
Unemployed	<i>Ref</i>			
Employed	-.2285	0.7957	0.2521	0.471
Trading	-.1300	0.8781	0.3138	0.716
Farming	-.2064	0.8135	0.2140	0.433
Retire	-.5073	0.6021	0.1998	0.126
Smoking Status				
Non-Smoker	<i>Ref</i>			
Smoker	.4565	1.5785	0.6248	0.249

Drinking Status				
Non-Drunker	Ref			
Drunker	.6666	1.9477	0.3321	0.000***
Exercise				
Regular	Ref			
Sometimes	.1297	1.1385	0.4574	0.747
Not at all	.1431	1.1538	0.1836	0.368
Salt Intake				
Not at all	Ref			
As in food	-.1536	0.8576	0.1896	0.487
At table	.2258	1.2533	0.3207	0.377
Age				
Below 50	Ref			
50 – 59	-.0638	0.9382	0.1730	0.729
60 – 69	.0508	1.0521	0.2376	0.822
Above 70	.1104	1.1167	0.3133	0.694
BMI				
Normal	Ref			
Underweight	.2226	1.2493	0.2483	0.263
Overweight	.1131	1.1197	0.1539	0.512
Obsessed	.9289	2.5317	0.1833	0.045**
SBP				
Low	Ref			
Normal	-.0946	0.9097	0.1903	0.650
Pre-hypertension	-.0622	0.9397	0.1962	0.766
High	.0103	1.0104	0.2237	0.963
DBP				
Low	Ref			
Normal	-.5090	0.6011	0.3151	0.100
Pre-hypertension	-.1179	0.8888	0.1583	0.508
High	-.0743	0.9284	0.7578	0.927
θ		0.0100	0.0247	
τ		0.0050		
Likelihood ratio (θ): $\chi^2 = 26.26$; prob. = 0.036**				

Source: Extracted from STATA output

Above present the estimated parameters for Cox Proportional Hazard model with random effect. The model selection criterion indicated that the Cox proportional Hazard model with random effect outperformed the conventional cox proportional model. Hence, the interpretation of the result was based on CPHM with random effect. The random effect in the cox proportional hazard model is assumed to follow the Gaussian distribution with mean 1 and variance equal to theta (θ). The heterogeneity in the population of the study

which is used as a cluster as estimated by the selected model is $\theta = 0.0100$ and the dependence within the clusters (hospitals) is measured by Kendall's tau is $\tau = 0.0050$. A variance of zero ($\theta = 0$) indicate that the random effect component does not contribute to the model. A likelihood ratio test for the hypothesis $\theta = 0$ indicates a chi-square χ^2 value of 26.26 resulting to a significant p-value of 0.036. The implication of this findings is that the random effect parameter had a significant contribution to the model.

The categorical variables: area of residence (peri-urban and rural residence), education level (primary and secondary), drinking status and BMI (Obsessed) significantly contributed to the hazard of diabetes mellitus. The estimated values ($B = 0.7804$, $HR = 2.1955$, $S.E. = 0.6534$, $p\text{-value} = 0.008$ and $B = 0.4901$, $HR = 1.6325$, $S.E. = 0.3759$, $p\text{-value} = 0.033$) for peri-urban and rural residence indicates that peri-urban and rural residence had a higher risk of dying with diabetes mellitus by a factor 0.1955(19.6%) and 0.6325(63.25%) respectively times higher than urban residence when other covariates are held constant. Similarly, the estimated values ($B = 0.3445$, $HR = 2.4113$, $S.E. = 0.2864$, $p\text{-value} = 0.050$ and $B = 0.9448$, $HR = 2.5576$, $S.E. = 0.7278$, $p\text{-value} = 0.001$) indicated that diabetes mellitus patients with primary and secondary school education level had higher risk of dying from diabetes by a factor 0.4223 (41.1%) and 0.5576 (55.8%) times higher as compared to those with tertiary level of education respectively. Furthermore, the estimated values ($B = 0.6666$, $HR = 1.9477$, $S.E. = 0.3321$, $p\text{-value} = 0.000$) suggested that diabetes mellitus patients that are drunker had a higher risk of dying from diabetes by a factor 0.9477 (94.8%) times higher as compared to those that do not drink. The estimated values ($B = 0.9289$, $HR = 2.5317$, $S.E. = 0.1833$, $p\text{-value} = 0.045$) revealed that diabetes patient that are obese had a higher risk of dying with diabetes by a factor 0.1833 (18.3%) as compared with those that have normal weight. However, gender, occupation, smoking status, exercise, salt intake and age were not significant contributing factors of patients dying from diabetes.

Table 4.6: Weibull AFT Models for Diabetes with Inverse Gaussian RE Component

Variable/Category	B	TR	S. E	P-value
Intercept	2.3753	10.7542	0.2414	0.000***
Gender				
Male	Ref			
Female	0.0191	1.0193	0.0884	0.829
Residential Area				
Urban	Ref			
Peri-Urban	-0.4502	.6375	0.1484	0.002***
Rural	-0.2498	.7790	0.1149	0.030**
Education				
Tertiary	Ref			
Primary	-0.1863	.8300	0.1022	0.048**
Secondary	-0.5270	.5904	0.1405	0.000***
Non-formal	-0.2534	.7762	0.1625	0.119

Occupation				
Unemployed	Ref			
Employed	0.1257	1.1339	0.1583	0.427
Trading	0.1377	1.1476	0.1793	0.442
Farming	0.1002	1.1054	0.1320	0.448
Retire	0.2307	1.2595	0.1667	0.166
Smoking				
Non-Smoker	Ref			
Smoker	-0.1155	.8909	0.2050	0.415
Drinking Status				
Non-Drunker	Ref			
Drunker	-0.3573	.6996	0.0854	0.000***
Exercise				
Regular	Ref			
Sometimes	-0.1155	.8909	0.1994	0.563
Not at all	-0.0452	.9558	0.0819	0.582
Salt Intake				
Not at all	Ref			
As in food	0.0229	1.0232	0.1994	0.834
At table	-0.1441	.8658	0.0820	0.256
Age				
Below 50	Ref			
50 – 59	0.0362	1.0369	0.0911	0.691
60 – 69	-0.0617	.9402	0.1134	0.586
Above 70	-0.0651	.9370	0.1420	0.647
BMI				
Norma weight	Ref			
Under weight	-0.1644	.8484	0.0996	0.049**
Overweight	-0.0418	.9591	0.0852	0.628
Obsessed	-0.4496	0.6379	0.2314	0.032**
SBP				
Low	Ref			
Normal	0.0605	1.0624	0.1055	0.566
Pre-hypertension	-0.0446	.9564	0.1092	0.684
High	-0.0425	.9584	0.2016	0.921
DBP				
Low	Ref			
Normal	0.4674	1.5958	0.5261	0.998
Pre-hypertension	0.07791	1.0810	0.0893	0.383
High	0.2041	1.2264	0.4010	0.611
μ	2.0409			
ρ	0.4900			
θ	0.0543		0.0574	
τ	0.0264			
	Likelihood ratio (θ): $\chi^2 = 4.08$; prob. = 0.022**			

Source: STATA output

Above present the results of data analysis of Diabetes Mellitus patients using the selected AFT model. The model selection criterion indicates that Weibull with Inverse Gaussian Random effect distribution which had a minimum AIC and BIC values of 864.3 and 999.8 appears to be appropriate model compared with other models. This implies that Weibull inverse Gaussian random effect AFT model is more efficient model to describe determinant factors of time-to-event of diabetes mellitus patient.

The random effect parameter in this model is assumed to follow an inverse Gaussian distribution with mean 1 and variance equal to theta (θ).

The heterogeneity in the population of the hospitals which were used as a cluster as estimated by the selected model is $\theta = 0.0543$ and the dependency within the clusters (hospitals) is measured by Kendall's tau is $\tau = 0.0264$. A variance of zero ($\theta=0$) would indicate that the random effect component does not contribute to the model.

A likelihood ratio test for the hypothesis $\theta = 0$ indicates a chi-square χ^2 value of 4.08 which resulted to a highly significant p value of 0.022. The implication of this findings is that, the random effect component had significant contribution to the model. The estimated Kendall's tau ($\tau = 0.0264$) shows that there is weak dependency within the cluster for Weibull Gamma random effect model.

The estimate of shape parameter in the Weibull inverse Gaussian random effect AFT model is $\mu = 2.0409$. This value shows the shape of hazard function is unimodal because the value is greater than unity i.e., it increases up to some time and then decreases.

The Weibull with inverse Gaussian random effect distribution shows that area of residence, level of education, drinking status and BMI are statistically significant (p-value < 0.05) factor for diabetes patient. Whereas the gender, occupation, smoking status, exercise, salt intake, age, SBP and DBP were found to be statistically insignificant factor for diabetes patient. The diabetes patients that reside in peri-urban and rural area had lesser survival by a factor 0.6375 and 0.7790 respectively than those that reside in urban area ($TR < 1$).

The result of the analysis also showed that diabetes patient with primary and secondary level of education had lesser survival time by a factor 0.8300 and 0.5904 as compared with those that had tertiary level of education ($TR < 1$).

Furthermore, it was discovered that diabetes patient with drinking status had lesser survival time by a factor 0.6996 ($TR < 1$) as compared to those that do not drink. In addition, diabetes patient with underweight and obese had lesser survival time by a factor 0.8484 and 0.6379 respectively ($TR < 1$) as compared to those with normal weight. This implies that BMI, area of residence, level of education and drinking status are the risk factors of diabetes.

Table 4.7: Likelihood Ratio Test

Dataset	Model	Likelihood ratio (θ): χ^2	P-value
Hypertension	CPH	0.53	0.234
	AFT	71.86	0.000***
Diabetes	CPH	26.26	0.036**
	AFT	4.08	0.022**

Source: Stata output

χ^2 likelihood ratio test for the random effect parameter (θ) shows evidence that AFT models outperform CPH model with minimum p-value in application to both hypertension and diabetes data.

The Global test for proportional for proportional hazard assumption for both Hypertension and diabetes data set used in this study revealed that the proportional hazard assumption was met ($p = 0.7922 > 0.05$; $p = 0.9460 > 0.05$) for hypertension and diabetes data set respectively.

AIC and BIC revealed that Weibull Gamma random effect AFT model and the Weibull inverse Gaussian random effect AFT models performed better than the other models consider in this study for predicting the risk factors of events from hypertension and diabetes respectively. These findings is in line with previous study by Isaac (2019) whose study revealed that parametric survival model performed better that then the semi-parametric. The finding from te current study support that of Susan (2011) whose study also found that gamma frailty model would be a practical choice in real data analysis with time to event end point when the regression parameters are of primary interest. Fagbamigbe, et al. (2021) also opined that flexible parametric model with frailty than other models performed better than non-frailty models. Th findings of the current study support that of Mardhiah et al (2022) whose study concluded that AFT survival analysis provides more reliable results and can be used as an alternative statistical analysis for determining the prognostic factors of mortality in melioidosis patients in certain situations.

The results from the Weibull Gamma random effect AFT model for the hypertension data set revealed that the random effect component had significant contribution to the model ($\chi^2 = 71.86$, $p < 0.05$). The estimated Kendall's tau ($\tau = 0.3602$) shows that there is strong dependence within the cluster (hospitals). The findings of the study revealed further that Patient with blood group AB had lesser survival than those with blood group o⁺ ($TR < 1$). Furthermore, the result of the analysis also revealed that Obese hypertensive patient had lesser survival time as compared to those with normal weight ($TR < 0.05$). This implies that Drugs, blood group and BMI are the risk factors of hypertension.

The results from the Weibull inverse Gaussian random effect AFT model for the diabetes data set revealed that the random effect component had significant contribution to the model ($\chi^2 = 4.08$, $p < 0.05$). The estimated Kendall's tau ($\tau = 0.0264$) shows that there is weak dependency within the cluster (hospitals). Based on this model, it was found that area

of residence, level of education, drinking status and BMI are statistically significant risk factor for event from diabetes ($p\text{-value} < 0.05$). The diabetes patient that reside in peri-urban and rural area had lesser survival by a factor 0.6375 and 0.7790 respectively than those that reside in urban area ($TR < 1$). The result of the analysis also showed that diabetes patient with primary and secondary level of education had lesser survival time by a factor 0.8300 and 0.5904 as compared with those that had tertiary level of education ($TR < 1$). Furthermore, it was discovered that diabetes patient with drinking status had lesser survival time by a factor 0.6996 ($TR < 1$) as compared to those that do not drink. In addition, diabetes patient with underweight and obese had lesser survival time by a factor 0.8484 and 0.6379 respectively ($TR < 1$) as compared to those with normal weight. Patients with overweight, and obesity were found to have a lesser survival time as compared with those with underweight patients.

Conclusion

This study compared the performance of semi-parametric model with and without random effect and, AFT models with and without random effect. Finally, the overall performance of semi-parametric and parametric models was compared. CPH model with random effect components outperformed other CPH models without random effect components in modelling both hypertension and diabetes data. AFT model with random effect components outperformed other CPH models without random effect components in modelling both hypertension and diabetes data. The Weibull AFT model with random effect outperformed other models considered in this study.

Recommendations

- i. Data analysts and healthcare professionals should adopt the Weibull AFT random effects model for survival analysis in hypertension and diabetes care.
- ii. Healthcare professionals should consider individual heterogeneity and clustering effects when interpreting patient data.
- iii. The use model estimates to inform personalized treatment strategies and regularly review and update patient prognoses based on changing risk factors.
- iv. Finally, healthcare professionals should incorporate survival analysis into clinical decision-making processes.

Conflicts of Interest

The researchers declared no conflict of interest.

References

- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34, 187-220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62, 269 – 276.

- Fagbamigbe, A. F., Karlsson, K., Derks, J., & Petzold, M. (2021). Performance evaluation of survival regression models in analysing Swedish dental implant complication data with frailty. *Plos one*, 16, 1 – 16.
- Gulliford M. C., Mahabir D., Rocke B. (2024). Socioeconomic inequality in blood pressure and its determinants: Cross-sectional data from Trinidad and Tobago. *Journal of Human Hypertension*. 18, 61–70.
- Isaac, N. (2019). *Robustness of Semi-Parametric survival model: Simulation studies and application to clinical data*. Unpublished Master's thesis, East Tennessee State University.
- Lahham, H. N. M., (2019). Cardiovascular Diseases and Risk Factors among Diabetic Patients. Nablus District, West Bank, Palestine.
- Mardhiah, K., Wan-Arfah, N., Naing, N. N., Hassan, M. R. A., & Chan, H. K. (2022). Comparison of Cox proportional hazards model, Cox proportional hazards with time-varying coefficients model, and lognormal accelerated failure time model: application in time to event analysis of melioidosis patients. *Asian Pacific Journal of Tropical Medicine*, 15, 128-134.
- Shankar, P. K., Screenivas, V. & Subrat, K. A. (2019). Comparison of Cox Proportional Hazards Model and Lognormal Accelerated Failure Time Model: Application in Time to Event Analysis of Acute Liver Failure Patients in India. *Nepalese Journal of Statistics*, 3, 21 -40.
- Susan, G. W. (2011). *Frailty models with applications in medical research: observed and simulated data*. Unpublished Master's Thesis, University of Nairobi.
- Vincent, A. M. & Ismaila, A. B. (2020). Parametric survival modeling of Tuberculosis data-A case study of Federal Medical Centre, Bida, Nigeria. *Journal of Modern Applied Science*, 14, 1-14.