Bima Journal of Science and Technology, Vol. 6 (1) April, 2022 ISSN: 2536-6041

PARAMETRIC AND SEMI-PARAMETRIC SURVIVAL MODELS WITH APPLICATION TO HYPERTENSION DATA

MUSA, GANAKA KUBI[1], LASISI K. E.[2], BELLO, ABDULKADR RASHEED[3], USMAN, Y. B.[4], ABDULLAHI, LAWAN[5]

[1 & 2]Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi, Nigeria
[3]Department of Mathematics and Statistics, Federal Polytechnic, Bauchi, Bauchi Satate, Nigeria
[4&5]Department of Mathematics and Statistics, Federal Polytechnic, Nasarawa, Nasarawa Satate, Nigeria
Corresponding Author: ganakamusak@gmail.com

ABSTRACT

Correlated survival data with possible censoring are frequently encountered in survival analysis. When there are dependencies among observed survival times, conventional Cox proportional hazards model (CPHM) and Accelerated Failure Time (AFT) models that assumes independent responses may not be appropriate. In this study, we compare the performance of parametric and semi-parametric survival models with application to clinical data. Specifically, the AFT model and the CPHM with and without Random effect were compared. Data on hypertension was collected from Federal Medical Centre Keffi and General Hospital Nasarawa for the period of five years (2016 – 2020). The results from the analysis revealed that the Weibull AFT model with Gamma Random effect distribution had the least AIC and BIC values indicating that it outperformed the other models considered in this study. Hence, the interpretation of the results was based on the most efficient model. Based on our results, it was found that hypertension patient that were giving drugs on the visit to the hospital has longer survival time compared to those that were not giving drugs. Also, Hypertension patient with blood group AB and Obesed have lesser survival time as compared to those with blood group o+ and normal weight respectively. The study recommend that health expert can use the Weibull AFT model with Gamma Random effect for predicting the risk factors of Hypertension especially when the data are correlated.

Keywords: AFT, CPHM, Hazard, Hypertension, Survival,

INTRODUCTION

Survival analysis is a set of statistical methods for data analysis where the outcome variable is the time until the occurrence of an event of interest (Collet, 2003). For example, the event of interest could be mortality, the onset of a disease, the failure of a device, or the recovery from surgery. Survival analysis is the name that is most generally used and recognized, despite the fact that different fields may emphasize slightly different methodologies and procedures (Xin, 2009). In survival analysis, the time is called the survival time.

The methods for dealing with survival analysis differ from other statistical methods for the following reasons: time is always the response variable, there are possibilities of staggered entries (the units in the study have different times of entrance), but this does not affect the survival analysis method because

the method deals with the length of time of observation rather than the same entrance time, and finally, the assumptions of normality do not hold because survival data are generic (Isaac, 2019). Three main classes, namely, non-parametric, semi-parametric and parametric survival models are usually used to analyze survival data. The discrepancies across the types of models are primarily due to assumptions made about the survival time distribution (Kleinbaum & Klein, 2012). When discussing nonparametric approaches, the Kaplan Meier estimator comes to mind immediately. Its limitation is that it can only handle one variable at a time. A semi-parametric model is the Cox proportional hazards model. A parametric survival model is one in which survival times are expected to follow a certain statistical distribution. Exponential, Weibull, gamma, generalized gamma, lognormal, and log-logistic distributions have all been utilized in this way. Semi-parametric models are more widely utilized in medicine than parametric ones.

In medical science, survival models have been used extensively. Here are a few examples: Pocock, Clayton and Altman (2002), Nardi, & Schemper (2003), Susan (2011), Shankar, Screenivas, and Subrat (2019), Vincent and Ismaila (2020) and so on. Chronic diseases such as hypertension are on the rise around the world, and they are linked to a lower quality of life and a higher financial burden; hence, developing preventive treatments for chronic diseases is critical (Lahham, 2009). A growing burden of chronic diseases, in addition to infectious diseases and dietary issues, is burdening developing countries. Despite the fact that chronic diseases account for a significant share of the disease burden in African countries, adequate efforts are not devoted to their prevention and control (WHO and AFRO, 2005).

Cox Proportional Hazard Model (CPHM) and Accelerated failure time (AFT) has received much attention recently. For instance, Nardi, & Schemper (2003) used real data in their study of the comparison of Cox and parametric models in clinical studies; Qi (2009) compare the performance of Proportional Hazards and Accelerated Failure Time Models.; Shankar, Screenivas, and Subrat (2019) compare the performance of Cox Proportional Hazards Model and Lognormal Accelerated Failure Time Model with application in Time to Event Analysis of Acute Liver Failure Patients in India; Isaac (2019) study the robustness of Semi-Parametric Survival Model: Simulation Studies and Application to Clinical Data Studies; Susan (2011) conducted a study on frailty models with applications in medical research: Using observed and simulated data, Vincent and Ismaila (2020) conducted a study on parametric survival modeling of tuberculosis data using data from Federal Medical Centre, Bida, Niger State, Nigeria. However, less attention has been given to modelling correlated survival data.

In survival analysis, correlated survival data with probable censoring is commonly in occurrence.

This comprises a multi centered study where individuals are clustered by clinical or other environmental factors that influence the predicted survival time. The traditional cox proportional hazards and AFT models, which assume observations are independent, are unsuitable for data in this situation. For Instance, Vincent and Ismaila (2020) performed parametric survival analysis of Tuberculosis data collected from Federal Medical Centre Bida. Three AFT parametric survival (Exponential, Weibull, Log normal and Log logistic) were fitted. It was found that the Weibull model performed better. However, the study focused on studying the effect of the fixed covariate on the survival

time without considering the cluster-specific (Hospital) random effect on the survival time.

In this study, we compare the performance of the parametric and semi-parametric survival models with application to clinical data. Specifically, the study compared the performance of Acceleration Failure Time and Cox Proportional Hazard Model with and without random effect with application to hypertension clinical data sets.

MATERIALS AND METHODS

Source of Data

Secondary data was used in this study. Data on Hypertension was collected from Federal Medical Centre (FMC) Keffi and General Hospital Nasarawa, Nasarawa State, Nigeria for the period of five years (2016 – 2020).

Method of Data Analysis

Cox Proportional Hazards models with Random Effect

A semi-parametric model (Cox-proportional Hazards model) with random effect can be formulated as:

$$\lambda_i(t) = \lambda_0(t)exp(\beta_1 z_{1i}, \beta_2 z_{2i}, ..., \beta_p z_{pi} + \propto_j) \quad (1)$$

Where λ_0 is the baseline hazard function, β_i is a vector of fixed effects corresponding to the covariates vectors $z_i \; and \; \propto_j$ is the per-subject random effect denotes the random effect associated with the j[th] cluster. The random effect can be thought of as an intercept that modifiers the linear predictors. This approach retains the full flexibility of Cox regression while accommodating associations among individual response times.

AFT Parametric Survival models with Random Effect Parameter

Conventional AFT model that assumes independent responses may not provide reliable inferences for clustered data that is, where subjects are correlated within clusters such as hospital. In this study, we introduce a random effect component to the AFT model that account for lack of independencies by introducing a random effect component as:

$$\log T_i = \mu + \propto' X_i + \sigma\epsilon_i + b \quad (2)$$

Where $\propto'$ is a vector of unknown regression coefficient, σ is a scale parameter, μ is the intercept parameter, the ϵ_i is the independently and identically distributed random errors, and the b is the cluster-specific random effects which are assumed to be independent, identically distributed random variables with density function $p(b)$. Here we have assumed that the random effect b follows gamma and inverse Gaussian distribution with mean zero and variance θ , as defined in the density function in equation (17) and (18) respectively.

$$f(Z) = \frac{Z^{\left(\frac{1}{\theta}\right)^{-1}}}{\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)} \exp\left(\frac{-X_i}{\theta}\right), \theta > 0 \quad (3)$$

Where $\Gamma(.)$ is the gamma function, it corresponds to the gamma distribution $Gam(\mu, \theta)$ with μ fixed to 1 for identifiability and its variance is θ the associate Laplace transform is:

$$L(\mu) = \left(1 + \frac{\mu}{\theta}\right)^{-\theta}, \; \theta > 0 \quad (4)$$

Note that if $\theta > 0$, there is heterogeneity. So the large values of θ reflect a greater degree of heterogeneity among groups and a stronger association within groups. The conditional survival and hazard function of the gamma frailty distribution is given by Gutierrez (2002):

$S_\theta = \left[1 - \theta \ln\left(S(t)\right)\right]^{\frac{-1}{\theta}}$ and $h_\theta(t) = h(t)\left[1 - \theta \ln\left(S(t)\right)\right]^{-1}$ (5)

where S(t) and h(t) are the survival and the hazard functions of the baseline distributions. For the Gamma distribution, the Kendall's Tau (Hougaard, 2012), which measures the association between any two event times from the same cluster in the multivariate case. It is an overall measure of dependence and independent of transformations on the time scale and the frailty model used. The associations within group members are measured by Kendall's, which is given by:

$\tau = \frac{\theta}{\theta+2} \epsilon(0,1)$ (6)

Similar to the gamma random effect model, simple closed-form expressions exist for the unconditional survival and hazard functions, this makes the model attractive. The probability density function of an inverse Gaussian shared distributed random variable with parameter $\theta > 0$ is given by:

$f_Z(X_i) = \left(\frac{1}{2\pi\theta}\right)^{\frac{1}{2}} X_i^{-3/2} \exp\left(\frac{-(X_i-1)^2}{2\theta Z_i}\right), \theta > 0, X > 0$ (7)

For identifiability, we assume X has expected value equal to one and variance θ. The Laplace transformation of the inverse Gaussian distribution is:-

$L(s) = \exp\left[\frac{1-(1+2\theta s)^{\frac{1}{2}}}{\theta}\right], \theta > 0, s > 0$ (8)

For the inverse Gaussian frailty distribution, the conditional survival and hazard function is given by Gutierrez (2002) in (9) and (10) respectively:

$S_\theta(t) = \exp\{\frac{1}{\theta}(1 - [1 - 2\theta \ln\{S(t)\}]^{\frac{-1}{2}}\}, \theta > 0$ (9)

and

$h_\theta(t) = \exp\{\frac{1}{\theta}\left([1 - 2\theta \ln\{S(t)\}]^{\frac{-1}{2}}\right\}, \theta > 0$ (10)

where S(t) and h(t) are the survival and the hazard functions of the baseline distributions. With multivariate data, an Inverse Gaussian distributed frailty yields a Kendall's Tau given by:

$\tau = \frac{1}{2} - \frac{1}{\theta} + 2\frac{\exp\left(\frac{2}{\theta}\right)}{\theta^2} \int_{\frac{2}{\theta}}^{\infty} \frac{\exp(-\mu)}{\mu} d\mu, \tau(0,\frac{1}{2})$ (11)

On the log survival time scale, the random effect can be thought of as an unobserved covariate that describes certain decreases or increases in event timings for distinct clusters. Within a cluster, all observations have a same unobserved random effect. The log of the survival time has a location-scale distribution in several survival time distributions, such as the Log-normal, Weibull, and Log-logistic distributions. Conditional on the random effects, the survivor function in (12) can be written in the form

$S_{ij}(t/b_i) = s_0(\frac{loT-\mu-\propto_i Z_i - b}{\sigma}|b).$ (12)

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.

1. The Weibull AFT Model

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

$f(t;\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \exp\left[\left(-\frac{t}{\mu}\right)^{\alpha}\right]$ (13)

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

$h(t;\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1}$ (14)

This yield 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 ,...,n with survival time t_i and censoring indicator δi which has value of one if ith observation is not censored and zero when the ith observation is censored and let α be the unknown parameter. The likelihood function is:

$$L(\alpha) = \prod_{i=1}^{n} \{f_i(t_i)\}^{\sigma_i}\{S_i(t_i)\}^{1-\sigma_i}\}$$
$$= \prod_{i=1}^{n} \{\left(\frac{f_i(t_i)}{S_i(t_i)}\right)^{\sigma_i} S_i(t_i)$$
$$\prod_{i=1}^{n} \{h(t_i)\}^{\sigma_i} S_i(t_i)\}$$
$$\prod_{i=1}^{n} \{\left(\frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1}\right)^{\sigma_i} \exp\left[\left(-\frac{t}{\mu}\right)^{\alpha}\right]\} \quad (15)$$

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:

$$h(t; \mu, \propto) = \lambda\alpha t^{\alpha-1}\exp(X \propto) \quad (16)$$

2. The Exponential AFT Model

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

$$T = \exp(\mu + \propto' X + \sigma\varepsilon) \quad (17)$$

Where ε is the error component

The survivorship function may be obtained by expressing in terms of time as:

$S(t, X, \propto) = exp(-t\, e^{-\propto' X})$ and the hazard function of the exponential regression model is

$h(t, X, \propto) = e^{-\propto' X}$.

3. The Log-Logistic AFT Model

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

$$logT = \exp(\mu + \propto' X + \sigma\varepsilon) \quad (18)$$

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 (11) is $s(t, X, \propto, \sigma) = \left([1 + \exp(z)\right)^{-1}$

Where z is the standardized log-time outcome variable, that is;

$$z = \frac{(y-\mu-\propto_i X)}{\sigma} \quad (19)$$

and $y = \ln(t)$

The odds of a survival time of at least t are,

$$OR = \frac{s(t,x,\propto,\sigma)}{1-s(t,x,\propto,\sigma)} = \exp(-z), \quad (20)$$

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:

$$OR(x = 1, x = 0) = \frac{\frac{\exp[-(y-\mu-\propto_1 x_1)]}{\sigma}}{\frac{\exp[-(y-\mu-\propto_1 x_0)]}{\sigma}}$$
$$= exp\left(\frac{\propto_1}{\sigma}\right) \quad (21)$$

This is independent of time.

4. The Lognormal AFT Model

The log-normal model assumes that ε ~N (0, 1). Let h(t) be the hazard function of T for the model (11) when β=0 i.e. $\beta 0 = \beta_1 = ... = \beta_p = 0$. Then, h(t) has the following functional form:

$$h(t) = \frac{\phi\left(\frac{\log(t)}{\sigma}\right)}{[1-\phi\left(\frac{\log(t)}{\sigma}\right)\sigma t} \quad (22)$$

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. The survival function $s(t/X)$ at any covariate x can be expressed as:

$s(t/X) = \phi[\mu + \propto_1^* x_1 + ... + \propto_p^* x_p - \propto \log(t)]$ (23)

Where $\propto = \frac{1}{\sigma}, \propto_j^* = \frac{\propto_j}{\sigma}$ for $j = 0,1, ..., p$

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

Method of Estimation

According to Gutierrez (2002), given the covariates information under assumptions of non-informative right -censoring and of independence between the censoring time and the survival time random variables, the marginal log-likelihood of the observed data is given by:

$l_{marg}(\varphi, \propto, \theta; z, X) = \prod_{i=1}^{s} [(\prod_{j=1}^{n_i} (h_0(y_{ij}) exp(X_{ij}^T \propto)^{\delta_{ij}}) X \int_0^{\infty} z_i^{d_i} exp\, z_i \sum_{j=1}^{n_i} h_0(y_{ij}) exp(X_{ij}^T \propto) f(z_i) dz_i]$

$= \prod_{i=1}^{s} [(\prod_{j=1}^{n_i} (h_0(y_{ij}) exp(X_{ij}^T \propto)^{\delta ij}) \times (-1)^{di} L^{di}(\sum_{j=1}^{n_i} h_0(y_{ij}) exp(X_{ij}^T \propto)]$ (24)

Taking the logarithm, the marginal likelihood is:

$l_{marg}(\varphi, \propto, \theta; z, X) = \sum_{i=1}^{s} \{[\sum_{j=1}^{n_i} \delta_{ij} (\log(h_0(y_{ij})) + X_{ij}^T \propto)] + \log[(-1)^{di} L^d([\sum_{j=1}^{n_i} h_0(y_{ij}) \exp(X_{ij}^T \propto)])]\}$ (25)

Where $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ is the number of event in the i^{th} cluster, and $L^{(q)}(.)$ is the q^{th} derivative of the Laplace transform of the random effect distribution defined as:

$L^{(q)}(s) = E[\exp(-Zs)] = \int_0^{\infty} \exp(-Z_{is}) f(Z_i) dz_i,\ s \geq 0$

where φ represents a vector of parameters of the baseline hazard function, $\propto$ the vector of regression coefficients and θ the variance of the random effect. Estimates of φ, $\propto$, θ are obtained by maximizing the marginal log-likelihood of the above; this can be done if one is able to compute higher order derivatives $L^{(q)}(.)$ of the Laplace transform up to $q = \max\{d_1, ..., d_s\}$.

Model Diagnosis

For the parametric regression problem, analogs of the semi parametric, residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates (Klein and Moeschberger 2003). The first of such residual is the Cox–Snell residual that provides a check of the overall fit of the model. The Cox–Snell residual, r_j, is defined by:

$$r_j = \widehat{H}(T_j | X_j)$$

Where $\widehat{H}$ is the cumulative hazard function of the fitted model. If the model fits the data, then the r_j 's should have a standard (λ=1) exponential distribution, so that a hazard plot of r_j versus the Nelson-Aalen estimator of the cumulative hazard of the r_j 's should be a straight line with slope 1. The best model will have the plots of the cumulative hazard close to the line of the residuals.

RESULTS

Application to Hypertension Data

Table 1: Descriptive Summary of Hypertension Data

Covariates	Categories	Status		Total
		Censored (%)	Event (%)	
BMI	Under weight	134(32.6)	14(3.4)	148(36.0)
	Normal weight	164(39.9)	19(4.6)	183(44.5)
	Overweight	55(13.4)	4(1.0)	59(14.4)
	Obese	16(3.9)	5(1.2)	21(5.1)
Blood Group	0^+	81(19.8)	6(1.5)	87(21.2)
	0^-	55(13.4)	6(1.5)	61(14.9)
	A^+	65(15.9)	7(1.7)	72(17.6)
	A^-	47(11.5)	7(1.7)	54(13.2)
	B^+	32(7.8)	4(1.0)	36(8.8)
	B^-	23(5.6)	5(1.2)	28(6.8)
	AB	62(15.1)	7(1.7)	69(16.8)
	C^+	1(0.2)	0(0.0)	1(0.2)
	D^-	2(0.5)	0(0.0)	2(0.5)
Family Type	Negative	35(49.3)	4(5.6)	39(54.9)
	Positive	27(38.0)	5(7.0)	32(45.1)
Alcohol intake	Yes	235(57.6)	28(6.9)	263(64.5)
	No	131(32.1)	14(3.4)	145(35.5)
Occupation	Student	19(4.7)	0(0.0)	19(4.7)
	Self-employed	332(82.0)	41(10.1)	373(92.1)
	Employed	13(3.2)	0(0.0)	13(3.2)
Diagnosis	HTN only	294(71.9)	36(8.8)	330(80.7)
	HTN and Others	73(17.8)	6(1.5)	79(19.3)
Drugs	Yes	351(86.7)	37(9.1)	388(95.8)
	No	12(3.0)	5(1.2)	17(4.2)

Source: Computed using SPSS

Table 2: Proportional Hazard Assumption

	Chi-square	P-value
Global test for Hypertension Data Set CPH model	12.13	0.7922

Source: Computed using STATA

The statistical approach (hypothesis testing) which tests the null hypothesis that proportional hazard assumption is met was used in testing the proportionality assumption. This approach is done by estimating the test of the proportional hazards assumption using the global test. Table 2 provides the results for the test. The "Global" gives the global test of probability for the model at once. The Chi-square values and p-values gives the test statistics and probability values under which the null hypothesis is tested. From the table 2, the p-value was found to be greater than 0.05. Hence, the null hypothesis was accepted.

This suggests that PH assumption is met at α = 5% significant level.

Table 3: AIC and BIC for Hypertension Data

	No RE		**GAMMA RE**		**Inverse Gaussian RE**	
Model	AIC	BIC	AIC	BIC	AIC	BIC
CPHM	3932.4	3999.3	***3931.9***	***3998.7***	-	-
Exponential	870.2	940.9	872.2	946.9	872.2	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: Computed using STATA

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for each model is presented in Table 5. In order to compare the performance of the conventional cox proportional hazard model with cox proportional hazard model with random effect. The AIC and BIC values were used. The result as presented in Table 3 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 focused on the interpretation of the result based on the cox proportional hazard model with random effect parameter.

In addition, the performance of AFT model with random effect and the conventional AFT model were also asses using the AIC and BIC values. The result of the analysis as presented in Table 3 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 outperformance 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 interpreting the result of Weibull Gamma random effect AFT model.

Table 4: Cox Proportional Hazard Models

		CPH (No RE)				CPH (With RE)			
Parameter	**Category**	**B**	**HR**	**S. E**	**P-value**	**B**	**HR**	**S. E**	**P-value**
Diagnosis	HTN only	Ref							
	HTN & Others	0.1180	1.1253	0.1645	0.419	0.0831	1.0866	0.1610	0.575
Drugs	No	**Ref**							
	Yes	-0.8635	0.4217	0.1121	0.001	-0.8538	0.4258	0.1134	0.001
Occupation	Student	**Ref**							
	Self-employed	0.0495	1.0507	0.3271	0.874	-0.0008	0.9992	0.3130	0.998
	Employed	0.0815	1.0849	0.4359	0.839	0.0530	1.0544	0.4243	0.895
Alcohol Intake	Yes	**Ref**							
	No	-0.0883	0.9155	0.1047	0.440	-0.0741	0.9286	0.1065	0.519
Family Type	Negative	**Ref**							
	Positive	0.0689	1.0713	0.1144	0.519	0.0515	1.0529	0.1129	0.630
Blood Group	0^+	**Ref**							
	0^-	0.1220	1.1298	0.2053	0.502	0.1172	1.1244	0.2041	0.518
	A^+	0.0755	1.0784	0.2131	0.702	0.0618	1.0638	0.2104	0.754
	A^-	0.1232	1.1311	0.2112	0.509	0.1181	1.1254	0.2099	0.527
	B^+	0.0388	1.0396	0.1943	0.835	0.0253	1.0256	0.1917	0.893
	B^-	-0.2084	0.8119	0.1902	0.374	-0.1922	0.82514	0.1933	0.412
	AB	0.2914	1.3383	0.3003	0.194	0.2698	1.3097	0.2942	0.230
	C^+	0.1973	1.2181	1.2367	0.846	0.1156	1.1226	1.1412	0.909
	D^-	0.1870	1.2056	0.87603	0.797	0.1610	1.1747	0.8539	0.825
BMI	Normal weight	Ref							
	Under weight	0.2220	1.2486	0.3104	0.372	0.2471	1.2803	0.3189	0.321
	Overweight	-0.0172	0.9829	0.1676	0.9190	0.0249	1.0252	0.1772	0.886
	Obese	0.3206	1.3780	0.16880	0.009	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: Computed using STATA

The estimated parameters for Cox Proportional Hazard model with and without random effect was presented in Table 4. Since CPHM with Random effect outperformed the conventional CPHM, the interpretation of the result was based on the selected model. The random effect in the CPHM is assumed to follow the Gamma distribution with mean 1 and variance equal to theta (θ). The heterogeneity in the population of the study which is used as a clusters 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 significantly to the model. A likelihood ratio test for the hypothesis $\theta = 0$ is shown at the bottom of Table 4 and indicates a chi-square $\chi 2$ 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. In the selected model (CPHM with random effect), the categorical variables drugs, BMI (Obese) were significant factor implying that drugs and BMI were the two factors in this study that significantly contributed to the hazard of hypertension patient. However, the variable BMI (underweight and overweight) were not significant using normal weight as the reference category (B = 0.2471, HR = 1.2803, S.E. = 0.3189, p-value = 0.321 and B = 0.0249, HR = 1.0252, S.E. = 0.1772, p-value = 0.886).

The estimated coefficient for patient that were giving drugs was -0.8538 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 estimated values for patient that were giving drugs is (B = -0.8538, HR = 0.4258, S.E. = 0.1134, p-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-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 normal 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. The other covariates Diagnosis, occupation, alcohol intake, family type and blood group were statistical insignificant.

Table 5 presents the summary results of analysis of Weibull AFT models with and without the random effect. As stated earlier, the Weibull Gamma random effect AFT model with a minimum AIC and BIC values of 512.8 and 591.5 appears to be appropriate model compared with other models (Table 3). The implication of this findings is that the Weibull Gamma random effect AFT model is more efficient model to describe determinant factors of time-to-event of hypertension. From Table 5, 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 clusters are estimated by the selected model is θ = 1.1259 and the dependence within the clusters (hospitals) is measured by Kendall's tau is τ = 0.3602. A variance (heterogeneity) of zero (θ=0) would indicate that the random effect component does not contribute to the model (no association among failure time). A likelihood ratio test for the hypothesis θ = 0 was presented at the bottom of Table 5 and indicates a chi-square ($\chi 2$) value of 71.86 which resulted 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. This suggest that there is a possible correlation in the survival time of patients within the hospital. The estimated Kendall's tau (τ = 0.3602) shows that there is strong dependence (association) within the cluster for Weibull Gamma random effect model. The estimate of shape parameter in the Weibull Gamma random effect AFT model is μ = 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.

Table 5: Weibull AFT Models for Hypertension

		Weibull (No RE)				Weibull (Gamma)				Weibull (Inverse Gaussian)			
Parameter	**Category**	**B**	**TR**	**S. E**	**P-value**	**B**	**TR**	**S. E.**	**P-value**	**B**	**TR**	**S. E**	**P-value**
Intercept		1.0903	2.9752	0.1875	0.000	0.1908	1.2102	0.2382	0.423	0.7108	2.0356	0.1950	0.000
Diagnosis	HTN only	**Ref.**											
	HTN & Others	-0.1381	.8710	0.0671	0.039	-0.0116	0.9885	0.0645	0.857	-0.0865	.9171	0.0662	0.192
Drugs	No	**Ref.**											
	Yes	0.4615	1.5865	0.1207	0.000	0.7712	2.1624	0.1237	0.000	0.538	1.7126	0.1185	0.000
Occupation	Student	**Ref.**											
	Self-employed	-0.0987	.9060	0.1422	0.488	0.1052	1.1109	0.1641	0.521	-0.0377	.9630	0.1433	0.793
	Employed	-0.1566	.8550	0.1875	0.404	0.0906	1.0948	0.1977	0.647	-0.0645	.9375	0.1836	0.725
Alcohol Intake	Yes	**Ref.**											
	No	0.1050	1.1107	0.0521	0.044	0.0338	1.0344	0.0482	0.484	0.0626	1.0646	0.0520	0.229
Family Type	Negative	**Ref.**											
	Positive	-0.0594	.9423	0.0490	0.224	-0.0002	.9998	0.0469	0.997	-0.0328	.9677	0.0490	0.503
Blood Group	0^+	**Ref.**											
	0^-	0.0480	1.0492	0.0848	0.572	-0.1112	0.8948	0.0751	0.139	-0.0212	.9790	0.0821	0.796
	A^+	-0.0723	.9303	0.0903	0.423	-0.0530	0.9484	0.0861	0.538	-0.0691	.9332	0.0910	0.447
	A^-	-0.1146	.8917	0.0859	0.182	-0.0370	0.9637	0.0809	0.647	-0.0892	.9147	0.0861	0.301
	B^+	-0.0559	.9456	0.0853	0.512	0.0379	1.0386	0.0802	0.637	-0.0189	.9813	0.08589	0.826
	B^-	0.1896	1.2088	0.1065	0.075	0.1267	1.1351	0.1027	0.217	0.1435	1.1543	0.1076	0.182
	AB	-0.1829	.8329	0.1027	0.075	-0.1871	0.8294	0.0939	0.046	-0.1838	.8321	0.1026	0.073
	C^+	-0.2164	.8054	0.4611	0.639	0.0430	1.0439	0.3623	0.905	-0.1057	.8997	0.4738	0.823
	D^-	-0.1135	.8927	0.3300	0.731	-0.0571	0.9445	0.2812	0.839	-0.0904	.9136	0.3375	0.789
BMI	Normal weight	**Ref.**											
	Under weight	-0.1847	.8314	0.1139	0.105	-0.0930	0.9112	0.1130	0.411	-0.1565	.8551	0.1147	0.173
	Overweight	-0.0442	.9568	0.0802	0.582	0.0251	1.0254	0.0754	0.739	-0.0189	.9813	0.07729	0.807

Obese	-0.2618	.7697	0.05565	0.000	-0.1079	0.8977	0.0541	0.046	-0.1999	.8188	0.0555	0.000
μ	0.7884		0.0376	0.000	1.4215		0.0897	0.000	1.1436		0.0707	
ρ	2.1998				4.1434		0.3719		3.1381		0.2218	
θ					1.1259		0.2311		1.3142		0.0225	
τ					0.3602				0.3965			
					Likelihood ratio (θ): $\chi^2 = 71.86$; prob. =0.000				Likelihood ratio (θ): $\chi^2 = 38.60$; prob. = 0.000			

Where: ρ is the shape parameter, μ is the scale parameter and θ is the random effect term

Source: Extracted using STATA

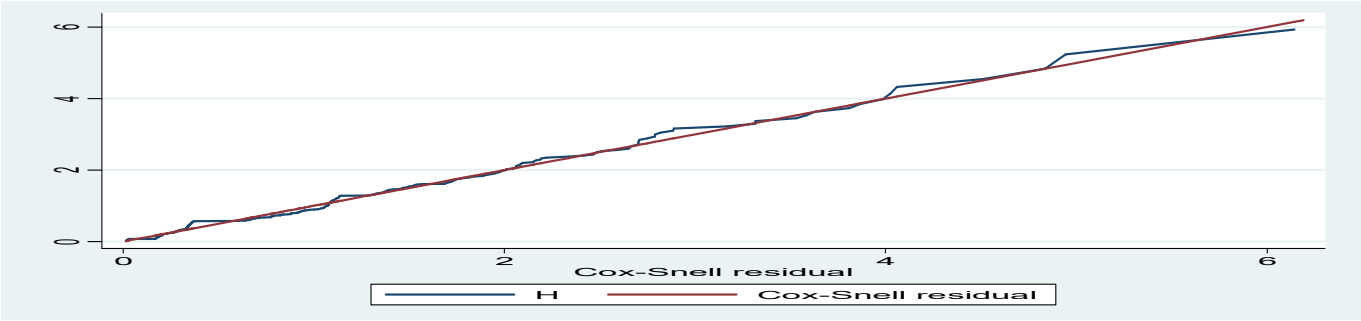

FIGURE 1: AFT Weibull Model with Gamma RE for HTN Data Sets

The estimated values, standard error, time ratio (TR), estimated parameters of baseline distributions and random effect parameter (θ) were presented in Table 5. 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 patient 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 0^+ (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 Cox–Snell residuals are one way to investigate how well the model fits the data. The plots of residuals for the selected models Weibull AFT with Gamma Random effect via maximum likelihood estimation with cumulative hazard functions are given in Fig. 1. If the model fits the data, the plot of cumulative hazard function of residuals against Cox–Snell residuals should be approximately a straight line. The plot for both model makes straight lines through the origin suggesting that the selected model is appropriate for time-to-Event of Hypertension.

CONCLUSION

This study aimed at comparing the performance of parametric and semi-parametric survival models with application to clinical data sets. Specifically, the study compared the performance of conventional semi-parametric model with extended semi-parametric model (CPHM with random effect). Also, the performance of conventional AFT models and AFT models with random effect were compared. Finally, the performance of semi-parametric and parametric AFT model with and without random effect were compared. Based on the results of the analysis as presented in the previous section, it was concluded that the Weibull AFT model with random effect (parametric AFT model) outperformed all other models considered in this study. The study recommend that the health expert can used AFT model with random effect in predicting the risk factors of Hypertension especially when the data arises from this chronic disease are correlated.

REFERENCES

Collett, D. (2003). Modeling survival data in medical research. Boca Raton Florida: Chapman & Hall/CRC.

Gutierrez, R.G. (2002). Parametric frailty and shared frailty survival models. *Stata Journal*. 2(1), 22–44

ısaac, N. (2019). *Robustness of Semi-Parametric survival model: Simulation studies and application to clinical data* (Master's thesis, East Tennessee State University).

Klein, J. P. & Moeschberger, M. L. (2003). Survival Analysis: Techniques for Censored and Truncated Data. Springer, Berlin (2003)

Kleinbaum, D. G., & Klein, M. (2012). Parametric survival models. In *Survival analysis* (pp. 289-361). Springer, New York.

Lahham, H.N.M., (2009). Cardiovascular diseases and risk factors among diabetic Patients. Nablus District, West Bank, Palestine.

Nardi, A. & Schemper. M. (2003). Comparing cox and parametric models in clinical studies. Statistics in Medicine, 2003:3597–3610, 2003.

Qi, Z. (2009). *Comparison of Proportional Hazards and Accelerated Failure Time Models* (Master's thesis, University of Saskatchewan Saskatoon, Saskatchewan).

Pocock, S. J., Clayton, T. C., & Altman, D. G. (2002). Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *The Lancet*, *359*(9318), 1686-1689.

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* (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*(7): 1-14.

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*(7): 1-14.

Xin, J. (2009). Frailty models for the between center variation in survival following rectum cancer diagnosis. Masters Dissertaion, Ghent University, Belgium.

WHO & AFRO (2005). Cardiovascular Diseases in African Region: Current Situation and Perspectives Report of the Regional Director. Fifty Fifth Sessions. Maputo, Mozambique.

WHO (2002). The Cost of Diabetes. Fact Sheet No. 236.