Survival Analysis for Hypertension among Adults Aged 18 Years and Above: A Case Study of Kerugoya Level 4 Hospital from 2018 to 2021

Peter Gachoki*, Lucas Macharia, Kinyanjui Jeremiah Ndung'u and Moreen Nderitu
[1]*Department of Pure and Applied Sciences, Kirinyaga University, P.O Box 143-10300, Kerugoya, Kenya*
**Corresponding*

Abstract: **Hypertension is among the several chronic diseases in human life and people have learned to live with it but ensure it is in good control. Hypertension can either be defined as systolic blood pressure (SBP) of 140mm of Hg or greater and/or diastolic blood pressure (DBP) of 90mm of Hg or greater, in simple terms hypertension is any blood pressure that is greater than 140/90mm of Hg. This research was set to investigate time to control hypertension among adults. Survival analysis of 66 patients with hypertension, adults aged 18 years and above, was carried out to estimate the time taken to control hypertension and assess the association as well as the impacts of the covariates namely: age, sex, hypertension stage, and diabetes status of the patients. Descriptive research through a retrospective cohort study was used to review the records of 66 patients sampled from the total number of hypertensive adults in Kerugoya Hospital by the use of a simple random sampling technique. Descriptive statistics analysis was done using a non-parametric method, Kaplan Meier (KM) method to estimate overall time to control by fitting survival function on the event, while Cox Proportional hazards model was used for multivariate analysis to assess the impact of the variables and to determine the adjusted hazard ratio. Log-rank test also compared time to control curves for different variables. Data analysis was done by the use of statistical software (R – Studio and SPSS) whose output shown that hypertension stage was the only significant factor for time to control hypertension with a p-value of (1.77E-05) while the rest of the covariates sex, age and diabetes status were insignificant. Log rank test output showed that there was no difference in the control curve for both sex and age while as there is a significant difference in the control curve for both hypertension stage and diabetes status.**

***Key words*: Hypertension, Kaplan Meier, Cox Proportional hazards model and Log-rank test**

I. INTRODUCTION

Hypertension is a common condition in which the long-term force of the blood against your artery walls is high enough that it may eventually cause health problems, such as heart disease (Nshimirimana, et al., 2019). The prevalence of hypertension in Kenya is 24.5%. Systemic hypertension is the leading cause of morbidity and mortality also known as a silent killer among non-communicable diseases. The World Health Organization (WHO) estimates that 36 million of the 57 million deaths that occurred worldwide in 2008 were due to Non-Communicable Diseases (NCD) making this the leading cause of death with a 63% mortality rate (Nshimirimana, et al., 2019).

Several chronic diseases have increased mortality at a very alarming rate, some of these chronic diseases include; diabetes, heart failure, cancer among others. In the list, High Blood Pressure (HBP) is in the frontline in contributing to the increment of mortality rate. According to Ogola et al. (2019) hypertension is the leading cause of global mortality accounting for 10.5 million deaths. According to the (World Health Organization, 2014), Hypertension Deaths in Kenya reached 1,995 or 0.60% of total deaths with an age-adjusted death rate of 12.49 per 100,000 population and this ranks Kenya position 120 in the world. Study research in Siaya County reported by (Awino, 2016) hypertension is major public health affecting approximately 26% of the adults' population worldwide. It is a significant risk factor for cardiovascular diseases, stroke, and renal failure and causes about 7.1 million deaths per year worldwide. The prevalence of hypertension is suggested to be increasing, particularly in developing countries due to nutritional extension.

Binary logistic regression has been used to show that age above 30 years (OR=12.0; 95% CI=2.7 to 52.3), marital status (Widowed OR= 16; 95% CI=6.34 to 40.52), smoking, BMI (overweight OR 2.98; 95% CI= 1.85 to 4.79 Obese OR= 4.23; 95% CI= 2.15 to 8.30 $p < 0.001$) were independent predictors of hypertension. However, increasing level of education was inversely related to hypertension (Awino, 2016). Even though hypertension is a worldwide chronic disease and the leading cause of cardiovascular diseases, stroke, diabetes among others the awareness of individuals with hypertension is low hence leading to untreated cases whose end product maybe die. For individuals with hypertension, only 15.6% are aware of their elevated blood pressure, among those aware only 26.9% are treated, and among those on treatment, only 51.7% have achieved blood pressure control, (Mohamed et al., 2018).

Research findings have shown that the overall prevalence of hypertension is 17%, with 21.4% in the urban population and 14.8% in the rural population. Increasing age, parental history of hypertension, tobacco smoking, tobacco chewing, physical inactivity, high estimated per capita salt consumption, and BMI ≥27.5 kg/m2 are the independent predictors for hypertension in the urban population, while in the rural population, increasing age, physical inactivity, central obesity, tobacco chewing, and tobacco smoking are the predictors (Erango, 2019).

Researchers have focused on survival time of the hypertensive patients bearing in mind hypertension cannot be cured but need to be in good control. This study aimed at examining the duration to achieve blood pressure (BP) control after the diagnosis of hypertension so as to minimize the increased prevalence of cardiovascular diseases due to high blood pressure. The study examined the factors affecting time to control hypertension among adults by use of the Kaplan Meier estimator and Hazard model. With the help of the Cox Proportional Hazard Model, the association and the impact of covariates was determined.

II. METHODOLOGY

2.1 Research Design

A retrospective cohort study of hypertensive patients was carried out at a public primary health care clinic at Kerugoya Hospital. The clinic provides medical care for acute cases as well as chronic diseases. Data were obtained from the reviews of the clinic's registry and their corresponding medical records. The inclusion criteria were diagnosed hypertensive patients with a date of diagnosis, age 18 years and older since 2018 and the study period was for 3 years. All hypertensive cases seen in this clinic were registered, and from this registry, we identified all cases that fulfilled the criteria from January 1, 2018 to December 30, 2021. Thus, among the 310 newly registered cases, only 66 cases were eligible.

The BP control target was defined as less than 140/90 mm Hg. The starting point of the retrospective cohort was at the date of diagnosis, and the end point was the date of controlled BP measurement or the closing date of the study, that is, December 30, 2021. The time to achieve BP control was defined as the duration from the date of diagnosis to the date of controlled BP measurement. Those who failed to turn up for the follow-up examinations or did not achieve the target BP were considered as censored data. Based on the BP reading at diagnosis, the cases were classified into stage 1 hypertension (SBP of 140-159 mm Hg and/or DBP of 90-99 mmHg), stage 2 hypertension (SBP of 160-179 mm Hg and/or DBP of 100-109 mm Hg), and stage 3 hypertension (SBP of ≥180 mm Hg and/or DBP of ≥110 mm Hg).

2.2 Survival Analysis Finding

The objects of interest in this study are the time to control hypertension and the association of the covariates. Survival analysis allows us to focus on the minimal time that can be taken to control hypertension among adults and estimate the relationship between time to control hypertension and the covariates i.e. the factors influencing time to control. In this study, two survival tools were used to analyze the data. First, the Kaplan Meier estimator to examine the general pattern of the time to control hypertension. Survival function gives the probability of time to control hypertension in a given time t. Secondly, Cox Proportional Hazards Model (CPHM) to determine the degree of influence of the factors on time to control.

2.3 Determination of Time to Control Hypertension Patterns (Kaplan Meier estimator)

Kaplan Meier estimator is a non-parametric statistic that gives the probability that individual subjects will "survive" past a particular time t (Kaplan & Meier, 1958). KM estimator was used to fitting survival function usually denoted as S (t) and it is given by the below, in this case, fit survival curve to determine the time to control patterns for different groups of hypertensive patients.

$$S\ (t) = \prod_{j/tj\leq t}\left(\frac{n_j-d_j}{n_j}\right)$$

Where: n_j = the number of hypertensive patients at time t_j.

d_j= the number of the patients who managed to control hypertension att_j.

Kaplan Meier estimates was used to develop a time to control curve which describes the probability to control for a different group of hypertensive patients to determine which of these groups take lesser time to control hypertension. The median time to control hypertension, as well as the percentiles, were obtained from these curves.

2.4 Factors Influencing Time to Control Hypertension (Cox Proportional Hazards Model)

There various factors associated with hypertension were namely: demographic factors example age, gender, marital status; health-related factor consists as the stages of HBP, alcoholism, stroke, heart failure, diabetes, and lastly economic factor to the likes as education level, exercise, alcoholic and smokers. The study determined these factors likely to influence the time taken to control hypertension. In particular, the study aimed at the degree of influence of these covariates on time to control hypertension which will be age, sex, hypertension stage, and diabetes. Cox Proportional Hazards Model is a semi-parametric model that allows to examine how covariates influence the rate of survival at a particular time t (Cox, 1972). The CPHM also produces covariates adjusted survival curve and hazards ratios. This model is characterized by hazard function also known as hazard rate. The hazard function is the probability of an event, usually denoted as $h\ (t)$,

$$h(t) = \lim_{\Delta t\to 0}\frac{P(t\leq T<t+\Delta t|T\geq t)}{\Delta t}$$

Cox Proportional Hazards Model that defines the hazard function by

$$h(t,x)\ = h_0(t)\ exp(\ \beta_i x)$$

Since there are more than one factor the Cox Proportional Hazards Model will be

$$h(t,x)\ = h_0(t)\ exp(\sum \beta_i\ x)$$

Where: $h_0(t)$= is called the baseline hazard (which is the intercept for the model)

$h(t)$ = hazard function

exp = the base of natural log
β_i= regressive coefficients, for $i = 1,2,3,4$
x_1= hypertension stages
x_2= age
x_3= sex
x_4= diabetes

The Cox Regression was done into two sections;

2.4.1 Univariate Cox Regression Analysis

This will be used to assess each variable concerning time to control, regression beta coefficients, effect size(given as hazard ratios), and statistical significance of each variable.

$$h(t, \text{hypertension stages}) = h_0(t)exp^{\Sigma\,\beta(\text{hypertension stages})}$$

$$h(t, \text{age}) = h_0(t)exp^{\Sigma\,\beta(\text{age})}$$
$$h(t, \text{sex}) = h_0(t)exp^{\Sigma\,\beta(\text{sex})}$$
$$h(t, \text{diabetes}) = h_0(t)exp^{\Sigma\,\beta(\text{diabetes})}$$

2.4.2 Multivariate Cox Regression analysis

This regression describes how the variables jointly impact time to control, which accounted for by performing a multivariate cox regression analysis. This model consisted of all the variables found to be significant to time control skipping those found not significant.

$$h(t, hypertension\ stage, \text{age}, \text{sex}, \text{diabetes}) = h_0(t)exp^{\Sigma\ \beta_0+\beta_1(hypertension\ stage)+\beta_2(age)+\beta_3(sex)+\beta_4(diabetes)}$$

Three statistical aims for the Cox Proportional Hazard model was to:

i. Test for significance of the age, sex, hypertension stage, and diabetes.
ii. Point estimate of age, sex, hypertension stage, and diabetes.
iii. Confidence interval for the age, sex, hypertension stage, and diabetes.
iv. Determine the hazards ratios for the age, sex, hypertension stage, and diabetes.

2.4.3 Computing the Hazard Ratio

In general, a hazard ratio (HR) is defined as the hazard for one individual divided by the hazard for a different individual. The two individuals being compared can be distinguished by their values for the set of predictors, that is, the X's. When division is done both sides by of the Cox Proportional Hazards Model $h_0(t)$ we obtain Hazard Ratio (HR) which is expressed as;

$$HR = (h(t, x*i))/(h(t,x)) = exp(\textstyle\sum \beta_i\ (x*i - xi))$$

Hazard ratio defined as the relative risk of failure to control hypertension at a time was used to show which groups are more likely to control hypertension first (Barracloug et al., 2011).

2.5 Comparison of survival curves (Log-rank test)

This describes how to test whether two or more survival curves are estimating a common curve. The most popular such test is called the log-rank test. The study compared different groups that are in gender (male and female), diabetes (diabetes and non-diabetes), stage of hypertension (elevated, stage 1, 2 and 3), and age. Log-rank test also called Mantel-Haenszel to log rank test, is a non-parametric hypothesis test according to that survival curves of two populations do not differ. The three groups were used to plot the time to control curve and compare their time to control hypertension.

III. RESULTS AND DISCUSSION

3.1 Determination of time to control hypertension patterns (Kaplan Meier estimator)

From the 66 hypertensive patients eligible for the study, data cleaning was done before conducting the actual analysis. A total of 66 hypertensive patients' records fulfilled the inclusion criteria and were included in the analysis. The median age was 54.5 (IQR = 38.5) years. More than half of the patients (60.6%) were female, 35.1% were Diabetic, and 44.8% were non-diabetic (Table 1). At diagnosis, 68.2% of the patients were at stage 2 or 3 of hypertension (Table 1) and rest were in stage 1 (Table 1). More than three-fourths (78.5%) of the patients reached the treatment target, while 47% of the patients reached the target at 6-month follow-up and 65% at 1-year follow-up. The median time to achieve BP control was 58 weeks (95% confidence interval [CI] = 45-64). A third of the patients were censored due to loss to follow-up, death, and withdrawal from the study, relocation, and the study ending before the occurrence of the event. The table below shows descriptive statistics of the study participants that is age, hypertension stages, sex, diabetes status, and control time.

Table 1: Characteristics of the study participants

Characteristics	Number of Patients	% of Patients
Age in Years		
Mean	55.11	
Range	19-94	
Control time		
Mean	59.5	
Median	58	
Mode	34	
Range	10-149	
Gender		
Male	40	60.4
Female	26	39.4
Hypertension		
Stage 1	21	31.8
Stage 2	27	40.9
Stage 3	18	29.3
Diabetes status		

Diabetic	37	56.1
Non-diabetic	29	43.9
Event		
Controlled	40	60.4
Censored	26	39.4

The comparison of the control probabilities of the subjects according to sex was done by plotting the Kaplan –Meier (KM) graph. The KM graph displays the cumulative control function on a linear scale by sex (Figure 1). The control curve of females was lower than that of males, which means that females have a higher probability of controlling (hypertension) than males. The overall median control time was 66 weeks for males and 34 weeks for females.

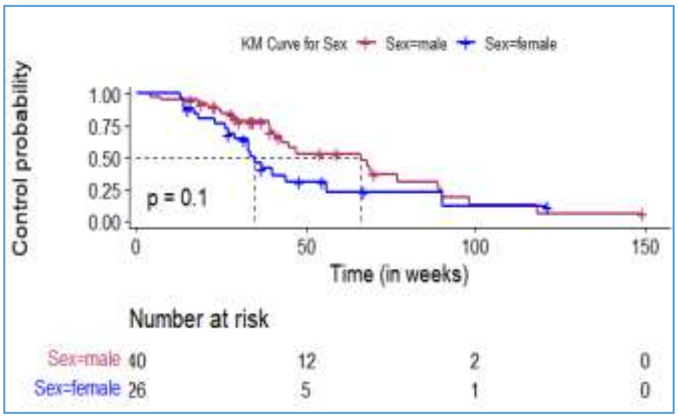

Figure 1: KM curve for the sex (male, female)

Among the 66 hypertensive patients chosen for the study 41 of them were found to be diabetic while the rest were non-diabetic. A KM curve for this strata clearly shows that non-diabetic patients have a higher probability of controlling hypertension, unlike the diabetic with a median control time of 33.5 weeks as presented in Figure 2.

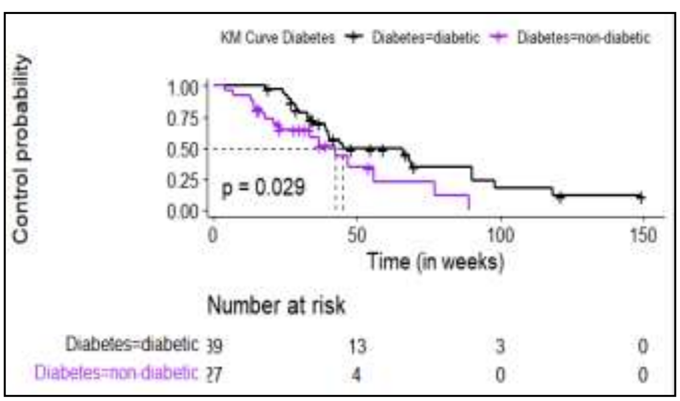

Figure 2: KM curve for diabetic and non-diabetic patients

Hypertension stage was among the factors under investigation by the use of the Figure 3 categorization procedure was done and the total study group yielded a total of (20, 27, 19) were confirmed to belong to the stage (1, 2, 3) respectively. A total of 17 patients were censored from the three stages.

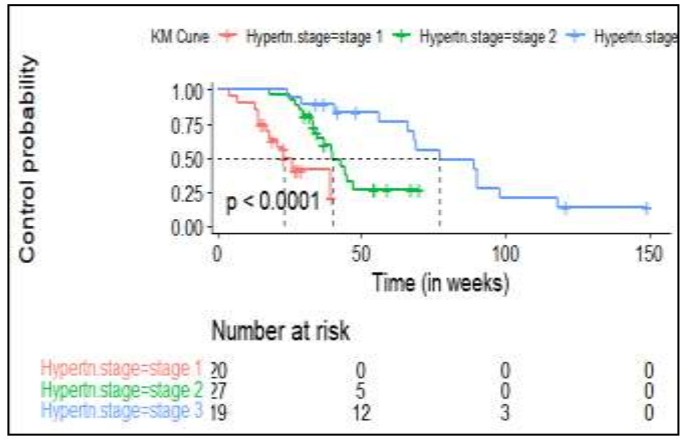

Figure 3: KM curve for hypertension stages

Since age was a continuous variable there was need to be categories for easier comparison of curves and for ease in hypothesis testing. For this reason age was categorized into two groups of less than 50 years classified as young and 50 years and above classified as old. A Kaplan Meier curve was used to show the pattern for the two categories (Figure 4).

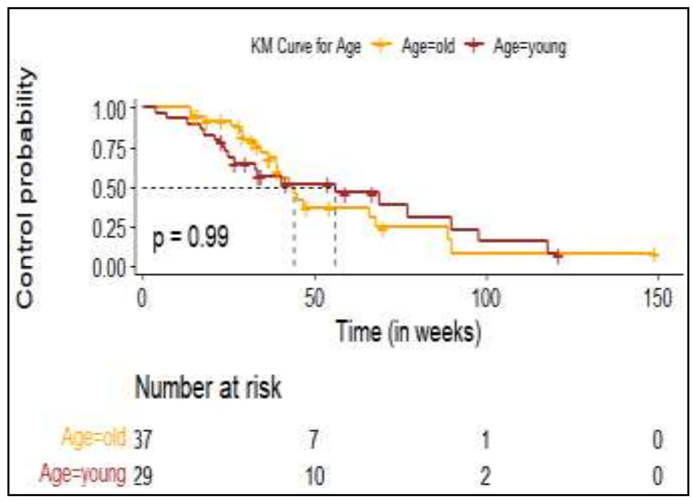

Figure 4: Kaplan Meier curve for age

3.2 Computation of the hazard ratios for the data

Hazard ratios indicate instantaneous reduction of risk to failure to control hypertension. The hazard ratio showed the risk of failure to control BP at any particular point in time. In hypertension stage the reference group was taken to be stage 1. A hazard ratio of 0.22 for for stages showed that patients with stage 2 have a reduced risk of failing to control compared to patients with stage 1. Similarly, for diabetes a hazard ratio of 1.64 for non-diabetic patients, means that non-diabetic patients are doing better by 0.64 in controlling compared to diabetic patients. For the case of age the young groups is doing by 0.21 as compared to the old. Finally. Female are controlling comparing to males by 0.818 (Figure 5).

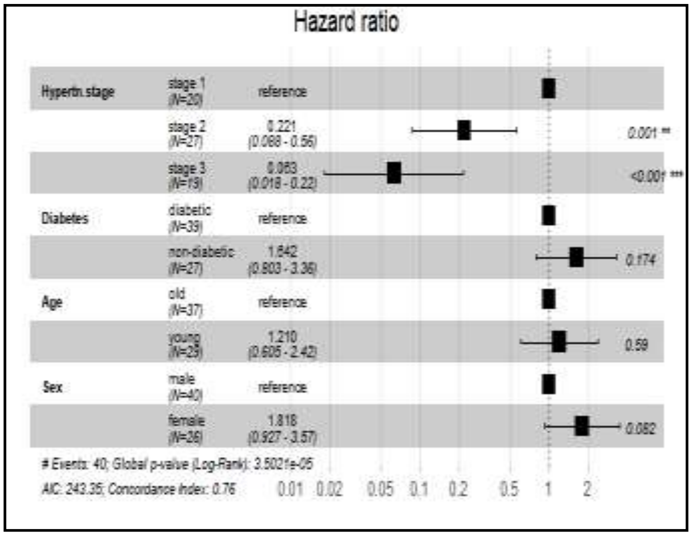

Figure 5: Hazard Ratio Representation for the Data

3.3 Comparison of the KM curves (log-rank test)

A log-rank test was implemented to show the differences between curves of the given strata of the study. There was a significant difference in the duration to achieve BP control between hypertension stage, sexes and diabetic patients. Female patients took a shorter time to achieve the BP target than their male counterparts (Figure 1). However, after the use of Log-rank test, the results showed otherwise (p-value= 0.1) and thus it was concluded that the curves had no difference (Figure 6). Furthermore, there was also a significant difference in the duration to achieve BP control for the hypertensive stage of the patient and diabetic hypertensive patients as confirmed by the use of log-rank test (p-value= 2e-06, p-value =0.03) respectively as in Figure 7 and Figure 8. However, there were no significant differences in the duration to achieve BP control between different age groups (Figure 9). Log-rank test was used to the hypothesis that;

Null Hypothesis: There is no difference in survival curves

Alternative Hypothesis: There is a difference in survival curves

```
Call:
survdiff(formula = s1 ~ Sex, data = hypertension)

       N Observed Expected (O-E)^2/E (O-E)^2/V
Sex=1 40       22     26.8      0.85      2.64
Sex=2 26       18     13.2      1.72      2.64

 Chisq= 2.6  on 1 degrees of freedom, p= 0.1
>
```

Figure 6: Log rank test output for sexes (male and female)

```
Call:
survdiff(formula = s1 ~ Hypertn.stage, data = hypertension)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
Hypertn.stage=1 20       11     3.42    16.791    21.050
Hypertn.stage=2 27       16    14.07     0.264     0.478
Hypertn.stage=3 19       13    22.51     4.015    12.225

 Chisq= 26.7  on 2 degrees of freedom, p= 2e-06
```

Figure 7: Log rank test output for hypertension stage

```
Call:
survdiff(formula = s1 ~ Diabetes, data = hypertension)

            N Observed Expected (O-E)^2/E (O-E)^2/V
Diabetes=1 39       24     29.8      1.14      4.77
Diabetes=2 27       16     10.2      3.35      4.77

 Chisq= 4.8  on 1 degrees of freedom, p= 0.03
```

Figure 8: Log - rank test output for diabetics' patients

```
Call:
survdiff(formula = s1 ~ Age, data = hypertension)

           N Observed Expected (O-E)^2/E (O-E)^2/V
Age=old   37       21       21  6.39e-05  0.000141
Age=young 29       19       19  7.04e-05  0.000141

 Chisq= 0  on 1 degrees of freedom, p= 1
```

Figure 9: Log rank output for age

3.4 Cox Proportional Hazard Model for Time to Control Hypertension

3.4.1 Univariate Cox Regression Analysis

In this case, the aim was to check the significance of the single factor to control time. These factors included age, sex, hypertension stage, and diabetes status of the patients. Two of the factor (hypertension stage and diabetes status) were statistically significant while the rest were not. Table 2 showed the calculation of the factors that played an important role in duration to control BP through Cox-regression. The results revealed that hypertension stage and the diabetes status of the patient played a significant role in the time taken to control hypertension.

Table 2: Univariate cox regression analysis

Factor	B	Sig.	Hazard Ratio	95.0% CI for Exp(B)	
				Lower	Upper
Hypertension Stage	-1.3603	4.82E-06	0.2566	0.1432	0.4597
Diabetes status	0.7255	0.0326	2.0667	1.602	4.019
Age of the patient	-0.0047	0.519	0.9954	0.9825	1.009
Sex of the patient	0.5117	0.1149	1.6681	0.8903	3.125

The univariate models were as follows

$$h(t, \text{hypertension stages}) = h_0(t)exp^{\Sigma\,\beta(-1.3603)}$$

$$h(t, \text{age}) = h_0(t)exp^{\Sigma\,\beta(-0.0047)}$$

$$h(t, \text{sex}) = h_0(t)exp^{\Sigma\,\beta(0.5117)}$$

$$h(t, \text{diabetes}) = h_0(t)exp^{\Sigma\,\beta(0.7255)}$$

3.4.2 Multivariate Cox Regression analysis

Table 3 shows the calculation for all combined together factors which were found the significant in the individual level in control time. This combine Cox regression analysis found that only hypertension stage plays a significant role in time to control hypertension. Diabetes was found to be insignificant after the multivariate analysis despite being significant for univariate cox regression analysis.

Table 3: Multivariate Cox Regression Analysis

	B	Sig.	Hazard Ratio	95.0% CI for Exp(B)	
				Lower	Upper
Hypertension Stage	-1.3039	1.77E-05	0.2715	0.1497	0.4924
Diabetes status	0.4063	0.253	1.5012	0.7486	3.0106

Finally, the final multivariate Cox Proportional Regression model is given by

$$h(t, hypertension\ stage, \text{diabetes}) = h_0(t)exp^{\Sigma\ \beta_1(-1.3039)+\beta_4(0.4063)}$$

IV. CONCLUSION

In conclusion, adherence and related disease are significantly effect on the control time the hypertension patients. The findings indicated that female hypertension patients had better control probability than male hypertension patients. The hypertension patients with age less than 25 years had the lowest control probabilities when compared to other age groups. Hypertension patient with diabetes had higher control probability as compared to that with non- diabetic patients. From the Cox proportional hazard model (univariate) two factors that were found statistically significant were hypertension stage and diabetes with (P vaule=4.82E-06, P value=0.0326) at 5% level of significance respectively. The two factor were eligible for the multivariate analysis after which only one of them (hypertension stage) was found to statistically significant.

REFERENCES

[1] Awino, B. O., Ogonda, L. A., Barno, G. C., & Magak, N. W. G. (2016). Awareness status and associated risk factors of hypertension among adult patients attending Yala sub-county hospital, Siaya county, Kenya (Doctoral dissertation, Maseno University).

[2] Barraclough, H., Simms, L., & Govindan, R. (2011). Biostatistics primer: what a clinician ought to know: hazard ratios. Journal of Thoracic Oncology, 6(6), 978-982.

[3] Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), 187-202.

[4] Erango, M. A., Gergiso, K. T., & Hebo, S. H. (2019). Survival time analysis of hypertension patients using parametric models. Adv. Res, 1-10.

[5] Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), 457-481.

[6] Mohamed, S. F., Mutua, M. K., Wamai, R., Wekesah, F., Haregu, T., Juma, P., ... & Ogola, E. (2018). Prevalence, awareness, treatment and control of hypertension and their determinants: results from a national survey in Kenya. BMC public health, 18(3), 1-10.

[7] Nshimirimana, A. D., Gikonyo, A. K., Gikonyo, D. K., Ponoth, P. M., & Gikonyo, B. M. (2019). Prevalence of Systemic Hypertension in Kenya: A Multicentre Analysis of the Karen Hospital Database. Journal of Kenya Association of Physicians, 2(1).

[8] Ogola, E. N., Barasa, F., Barasa, A. L., Gitura, B. M., Njunguna, B., Beaney, T., ... & Poulter, N. R. (2019). May Measurement Month 2017: the results of blood pressure screening of 14 845 individuals in Kenya—sub-Saharan Africa. European Heart Journal Supplements, 21(Supplement_D), D71-D73.

[9] World Health Organization. (2014). Hypertension in Kenya and worldview. Nairobi, Kenya