A Dynamic Regression Modeling of the Prevalence of Stroke in South-East of Nigeria

[1]Christogonus Ifeanyichukwu Ugoh, [2]Chukwunonso Nelly Amolo, [3]Akindele Emmanuel Oni, [4]Mercy Chinyere Kalu, [5]Tosin Babatunde Adewale

[1]Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria
[2,4,5]Department of Microbiology, Faculty of Life Science, University of Benin, Benin, Nigeria
[3]Department of Pharmacology, Faculty of Basic Medical Sciences, University of Port Harcourt, Nigeria

Citation: Ugoh, C.I., Amolo C.N., Oni A.E., Kalu M.C, Adewale T.B (2022) A Dynamic Regression Modeling of the Prevalence of Stroke in South-East of Nigeria, *International Journal of Public Health, Pharmacy and Pharmacology*, Vol. 7, No.4, pp.31-40

ABSTRACT: *Stroke is increasingly becoming one of the major public health issues in Nigeria. This study aimed to predicting future prevalence of stroke in south-east Nigeria in 2022 and 2023, using previous prevalence. The number of patients recorded with stroke (transient ischemic attack) and risk factors hypertension (HT), diabetes mellitus (DM), dyslipidemia (DY), and alcohol (AL) on monthly basis from January 2017 to December 2021 in Enugu State University Teaching Hospital were extracted. The dynamic regression model was applied to the data, best model is selected using Akaike information criterion corrected (AICC), and the model generated is used to predict the number of patients with stroke in 2022 and 2023. A total of 1216 patient records were included in this study. The proportion of hypertension was 51.49%, diabetes mellitus was 7.65%, alcohol was 6.4%, and dyslipidemia was 24.81%. Regression with ARIMA(0,0,1) was the best model. The prediction showed that by December 2022 the number of patients will increase by 29.63% and by December 2023 it will rise to 36.67%. The findings of this study suggest the prevalence of stroke in south-east Nigeria is high and will still rise in the future. There is still need of further research on stroke and other risk factors towards establishing appropriate policy, preventive and management measures.*

KEYWORDS: Stroke, transient ischemic attack, dynamic regression model, risk factors, Nigeria

INTRODUCTION

Stroke is a non-communication disease and it is one of the leading causes of mortality and disability in the world. Stroke's burden, as well as its increasing incidence and economic impact is very high in developing countries (Mukherjee & Patil, 2011; Emmanuel et al, 2015), It is the second leading cause of death worldwide. The prevalence and incidence of stroke vary from community to community and from time to time worldwide (Kelechi et al, 2014). According to

World Health Organization (WHO) stroke accounts for 10.8% mortality and 3.1% of disease burden worldwide. There is an epidemiologic and demographic transition of diseases in most developing countries with increased risk for cardiovascular diseases (Mensah, 2008). According (Towfighi & Saver, 2011), by the year 2030, low and middle income countries of the world will experience prevalence of about 80% of all stroke cases. A review from Feigin et al., (2003), on global stroke cases showed that as the developed countries are experiencing a decline in stroke incidence most of the developing countries are experiencing a rise in stroke incidence of about 100 percent. Africa countries are always face with limitations on information or data on risk factors and epidemiology of stroke, however, availability of such information are mostly based on western populations (Osuntokun et al, 1979; Ogun et al, 2015). Managing diseases such as stroke in Nigeria is largely conservative, because there is little or no funding for high-quality research. Primary prevention is optimum means to reducing the burden of the disease especially in developing countries (Kolawole, 2008). Studies conducted in both developed and developing nations revealed that crude prevalence of stroke ranges from 0 to 10.2 per 1,000 populations, and that the burden of stroke has been projected to have an increasing trend (Feigin et al., 2003; Danesi et al., 2007). Although many studies and research about prevalence of stroke have been done extensively in the literature, no studies about stroke have been done using dynamic regression model. The essence of this study is to use the already existing data on the number of patients with transient ischemic attacks recorded in Enugu State University Teaching Hospital for the period under review, to predict for the future prevalence using the techniques of dynamic regression model.

MATERIALS AND METHODS

In this study, the dataset consists of the number of patients with stroke with emphasis on transient ischemic attack (TIA), i.e. the prevalence of stroke and four modifiable risk factors hypertension (HT), diabetes mellitus (DM), dyslipidemia (DY), and alcohol (AL) reported and recorded in a monthly basis from January 2017 to December 2021 in Enugu State University Teaching Hospital.

The data are analyzed using the techniques of linear regression model, autoregressive integrated moving average (ARIMA) model, and the combination of linear regression and ARIMA (known as the dynamic regression), to capture the effects of the risk factors on the prevalence of stroke. The descriptive statistics (mean, median, and standard deviation) are used to represent the average and variations of the occurrence of the risk factors; the frequency and proportion (in %) are used to represent the rate of occurrence of the risk factors. All the analysis is done using the R programming package.

Linear regression model

Regression analysis is a technique that deals with the relationship between a response variable and one or more predictor variables. The relationship between a response variable and one predictor variable is known as simple regression, while the relationship between a response variable and more predictor variables is known as multiple regression.

Using the multiple linear regression analysis to develop a model connecting linearly the Transient ischemic attack (TIA) and the predictor variables hypertension (HY), diabetes mellitus (DM), dyslipidemia (DY), and alcohol (AL). The linear regression model which is the function of TIA and predictor variables is expressed as

$$TIA = \beta_0 + \beta_1 HY + \beta_2 DM + \beta_3 DY + \beta_4 AL + \varepsilon_t \quad (1)$$

where $\beta_0, \beta_1, \beta_2, \beta_3,$ and β_4 are parameters of the risk factors (the predictor variables)

2.2 Autoregressive integrated moving average (ARIMA) model

Autoregressive integrated moving average (ARIMA) is a statistical a statistical analysis model that make use of time series data to predict future trends. ARIMA is an acronym where AR stands for autoregressive (AR) process that is used to predict future values based on past values, integrated (I) represents the differencing of the raw observations allowing the time series to become stationary, and moving average (MA) represents the connection of observations and the residual error. If the time series data are stationary, then autoregressive moving average (ARMA) is modeled instead and this written as

$$TIA_t = c + \phi_1 TIA_{t-1} + \cdots + \phi_p TIA_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (2)$$

If the observations of the time series, in this regard the prevalence of stroke are differenced, then autoregressive integrated moving average (ARIMA) is modeled to the data. It is written as

$$\nabla^d TIA_t = c + \phi_1 TIA'_{t-1} + \cdots + \phi_p TIA'_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (3)$$

Stationary AR(p) process: This is the regression of current observations y_t on past observations y_{t-i}, and it is defined as

$$TIA_t = \delta + \phi_1 TIA_{t-1} + \cdots + \phi_p TIA_{t-p} + \varepsilon_t, \qquad \delta = \mu\left(1 - \phi_1 - \cdots - \phi_p\right) \quad (4)$$

Stationary MA(q) process: This is the regression of current disturbance on past disturbances, and it is defined as

$$TIA_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (5)$$

where p is AR order, q is MA order, d is degree of differencing, $\phi_1, \cdots, \phi_p$ are the autoregressive process parameters, $\theta_1, \cdots \theta_q$ are the moving average process parameters, $TIA'_{t-1}, \cdots, TIA'_{t-p}$ are the differenced stroke series, δ is the drift, μ is the mean of MA process, ε_t is white noise and $\varepsilon_{t-1}, \cdots, \varepsilon_{t-q}$ are the residual errors.

The best ARIMA model is selected form the list of competing models using Akaike information criterion corrected defined as

$$AICC = -2lnL + 2p + \frac{2p(p+1)}{n-p-1} \tag{6}$$

where L is the likelihood of the data, p is the number of fitted model parameters, and n is the sample size.

2.3 Dynamic regression model

This is a linear combination of regression and ARIMA model, where the ARIMA is modeled to the residual error obtained as a result of the regression model of the TIA and the four risk factors HY, DM, DY and AL. Dynamic regression model is written as

$$TIA = \beta_0 + \beta_1 HY + \beta_2 DM + \beta_3 DY + \beta_4 AL + a_t \tag{7}$$

where a_t can be AR process, MA process or both.

Dynamic regression model with AR process is written as

$$TIA = \beta_0 + \beta_1 HY + \beta_2 DM + \beta_3 DY + \beta_4 AL + \delta + \phi_1 TIA_{t-1} + \cdots + \phi_p TIA_{t-p} + \varepsilon_t \tag{8}$$

Dynamic regression model with MA process is written as

$$TIA = \beta_0 + \beta_1 HY + \beta_2 DM + \beta_3 DY + \beta_4 AL + \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{9}$$

RESULTS

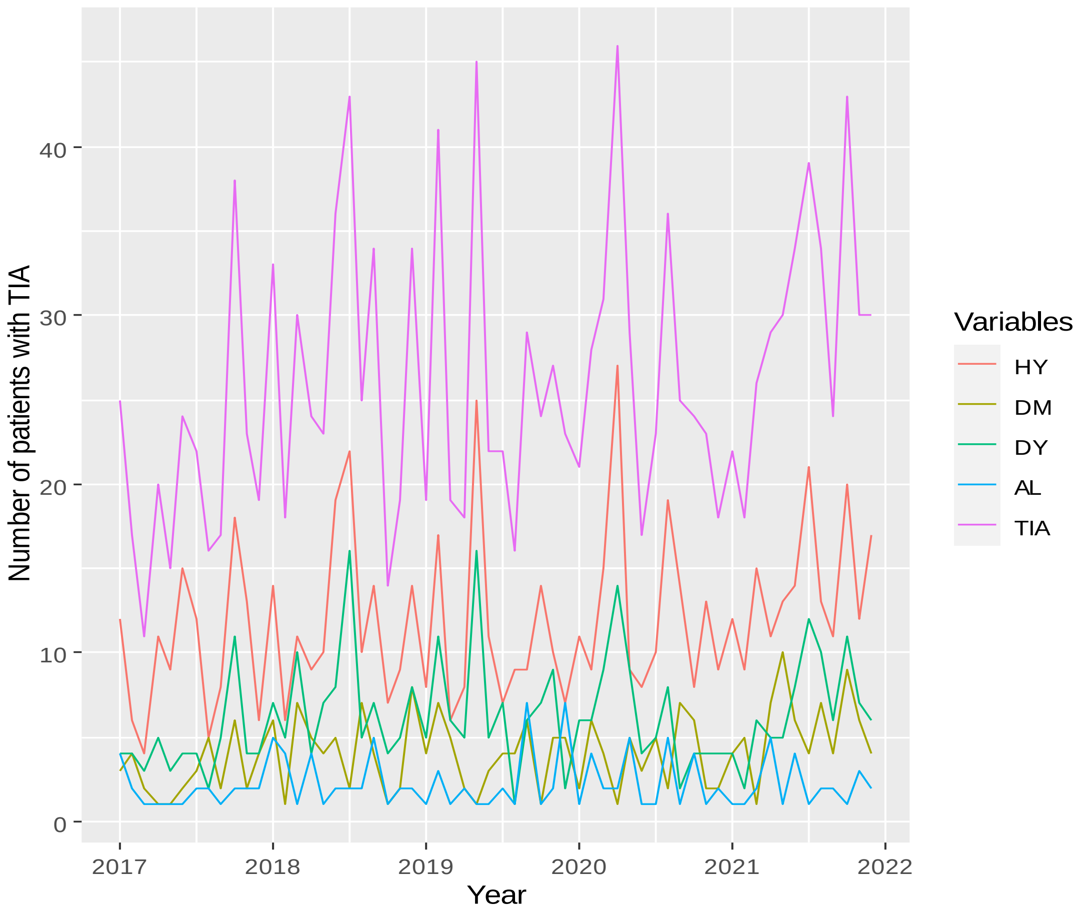

Figure 1. Time plot of the number of patients with transient ischemic attack (TIA) and the risk factors associating to TIA

Figure 1 is the time plot of the prevalence of transient ischemic attack (TIA) and the four risk factors: hypertension, diabetes mellitus, dyslipidemia, and alcohol. The plot shows no evidence of trend, that is the prevalence of TIA is not affected with change in time. Thus, this implies that TIA is stationary, and do not require any differencing.

In Figure 2, the boxplot for Transient ischemic attack (TIA) show high variation in their mean prevalence within the time review, hypertension, diabetes mellitus, and dyslipidemia show slight variations, while alcohol show very low variation in their respective mean prevalence over the time under review. There is significant difference between the four risk factors based on their distribution.

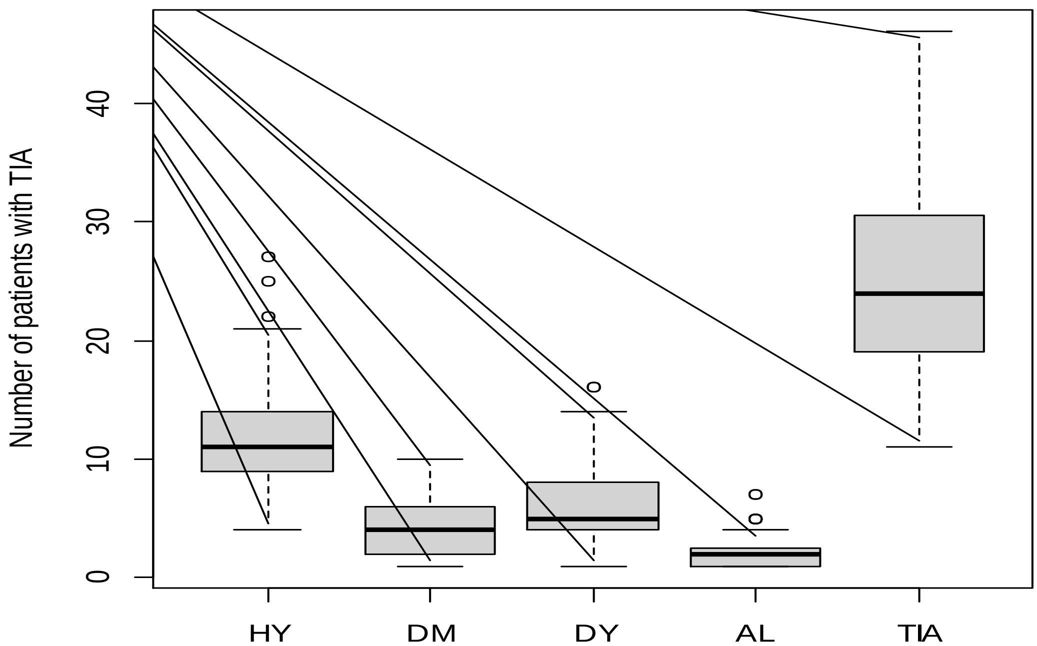

Figure 2. Box plot for transient ischemic attack (TIA) with the risk factors

Table 1. Descriptive statistics of the risk factors

	Min.	Max.	Mean	Median	Standard deviation	Total	Proportion
Hypertension	4	27	12	11	4.73	693	51.49%
Diabetes mellitus	1	5	2	1	1.08	103	7.65%
Dyslipidemia	1	16	6	5	3.36	334	24.81%
Alcohol	1	4	2	1	0.67	86	6.40%

In Table 1, over the period January 2017 to December 2021, hypertension is the highest prevalent risk factor 693 (51.49%), with average prevalence of 12 patients per month and a standard deviation of approximately 5, hypertension is followed by dyslipidemia with 334 (25.81%) and a mean prevalence of 6 patients monthly. This result shows that hypertension is the major cause of transient ischemic attack (TIA), followed by dyslipidemia.

Table 2. Estimated coefficient for regression with ARIMA(0,0,1) errors

	MA1	Intercept	HT	DM	DY	AL
Coefficients	-0.6075	2.4927	1.0082	1.0300	1.0097	1.1726
Standard error	0.1311	0.3449	0.0363	0.0506	0.0468	0.0870

From Table 2, the dynamic regression model is written as

$$TIA = 2.4927 + 1.0082HY + 1.03DM + 1.0097DY + 1.1726AL + 0.6075\varepsilon_{t-1} \quad (10)$$

This equation is used to make predictions in Table 3. Figure 3 presents the model adequacy check for the dynamic regression model obtained. The time plot for the residuals from regression with ARIMA(0,0,1) shows evidence of constant variance (no trend is found); there is no significant lag in the autocorrelation function (ACF) plot for the forecast errors, and this implies that there is non-zero autocorrelations at lag 1-20 for the in-sample forecast errors, and the histogram shows that the distributions of the forecast errors are normally distributed.This implies that the dynamic regression model obtained is a good model for predicting the prevalence of patients with stroke (transient ischemic attack) in Enugu state university teaching hospital. In addition, the dynamic regression model obtained forecasts the TIA by utilizing the effect of the risk factors, in order to show how the prevalence of stroke behaves.

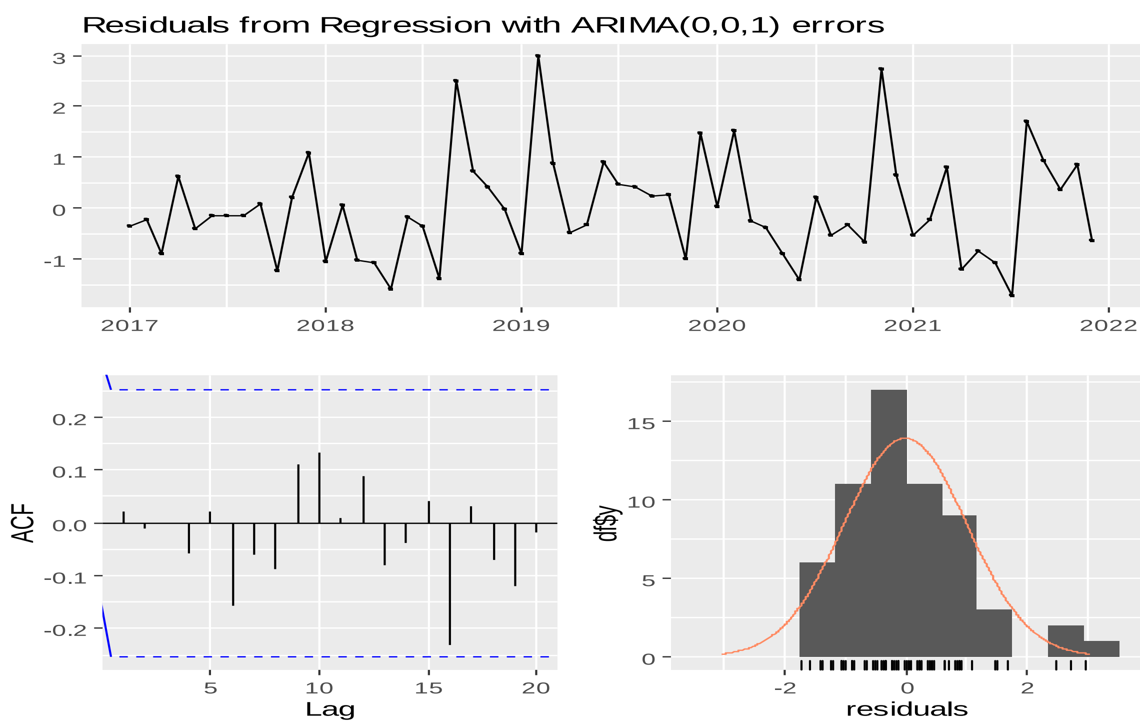

Figure 3. Model adequacy check for regression with ARIMA(0,0,1) errors

Table 3. Predicted number of patients with TIA from Jan 2022 to Dec 2023

Year/Month	Predicted number of patients with TIA	
	2022	**2023**
January	27	36
February	19	19
March	13	32
April	21	25
May	17	25
June	25	37
July	25	45
August	17	27
September	19	34
October	40	16
November	24	21
December	19	35

Table 3 shows the forecasted number of patients with stroke (transient ischemic attack) in Enugu state university teaching hospital. The result shows that in January 2022, Enugu state university teaching hospital will experience a prevalence of 27 patients with stroke, and that by December 2022, the prevalence of patients with stroke will decrease to 19 patients showing a percentage increase of 29.63%, and in January 2023, the prevalence of patients with stroke will rise again to 36 patients showing a percentage increase of 47.22%, and by December 2023, the prevalence of patients with stroke will reduce by one, therefore showing a percentage decrease of 2.78%. Thus there a decrease in the prevalence of patients with stroke from December 2021 (30 patients) to December 2022 by 36.67%, and by December 2023, the prevalence will rise by 14.29%. However, by the end of 2023, Enugu state university teaching hospital will be expecting a high prevalence of stroke patients.

Figure 4 shows the time plot for the actual prevalence of stroke patients and the predicted prevalence of stroke patients for a period of four years.

International Journal of Public Health, Pharmacy and Pharmacology
Vol. 7, No.4, pp.31-40, 2022
Print ISSN: (Print) ISSN 2516-0400)
Online ISSN: (Online) ISSN 2516-0419)

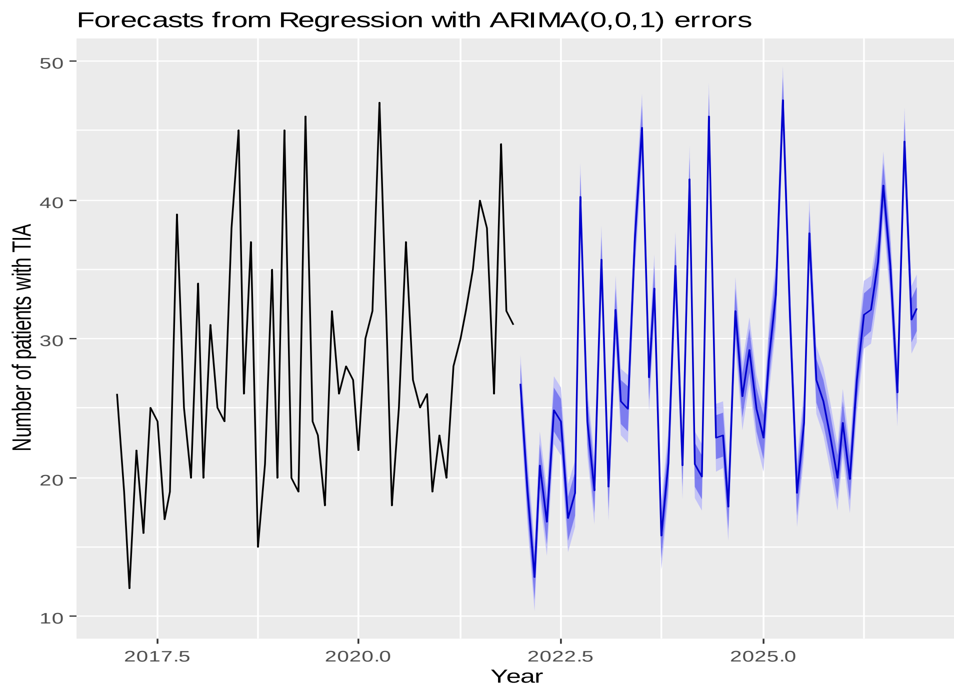

Figure 4. Plot showing the predicted number of patients with TIA from Jan 2022 to Dec 2025

DISCUSSION

Many studies have been done and published on the prevalence of stroke in south-east Nigeria but without any consideration on the prediction of patients with transient ischemic attack (TIA) using time series models. The findings of this study revealed that the number of patients with transient ischemic attack in the south east Nigeria is constant over the period under review. The future estimates of patients with TIA were based on the statistical application of multiple regression model and autoregressive integrated moving average model, and descriptive statistics.

Based on this study, hypertension was the leading modifiable risk factor, similar to findings from other studies in Nigeria, Africa, and the rest of the world (Desalu et al., 2011; Ekeh et al., 2015). This study showed that the number of transient ischemic attack is very high in the south eastern side of Nigeria and that in the future the prevalence of stroke will rise the more. In addition, the TIA will rise by 29.63 % in December 2022, and by December 2023, it will further rise to 36.67%.

CONCLUSION

This study found that the number of transient ischemic attack in south east Nigeria is rising more and more, and the prevention of stroke has been affected mainly by wealth systems and poor government response (Bonita and Truelsen, 2003). This study suggests an increasing trend of of number of patients with transient ischemic attack in south-east Nigeria. However, this study had some limitations. The current low availability of data, due to the fact that some case files were missing and could not be included in the data analysis. Thus, there is need for more research capacity, basically to conduct and fund higher research that could raise awareness on stroke, and registry-based study will provide more liable information on stroke and risk factors.

REFERENCES

Arabambi, B., Oshinaike, O., Akilo, O.O., Yusuf, Y., & Ogun, S.A. (2021). Pattern, risk factors, and outcome of acute stroke in a Nigerian university teaching hospital: A 1-year review. *Niger J Med*, 30: 252-258.

Bonita R. and Truelsen T. (2003). Stroke in sub-sahara Africa: a neglected chronic disease. *Lancet neurol*, 2: 592.

Emmanuel, O.S., Olufemi, O.D., Feyiyemi, A., Sunday, A.A., Akeem, S., & Olabode, O. (2015). Prevalence of stroke in the three semi-urban communities in middle-belt region of Nigeria: a door to door survey. *Pan African Medical Journal*, 20: 33.

Ekeh B, Ogunniyi A, Isamade E, Ekrikpo U. (2015). Stroke mortality and its predictors in a Nigerian teaching hospital. *Afr Health Sci*, 15:74-81.

Danesi M, Okubadejo N, Ojini F. (2007). Prevalence of stroke in an urban, mixed-income community in Lagos, Nigeria. *Neuroepidemiology*. 28:216–223.

Desalu OO, Wahab KW, Fawale B, Olarenwaju TO, Busari OA, Adekoya AO, et al. (2011). A review of stroke admissions at a tertiary hospital in rural Southwestern Nigeria. *Ann Afr Med*, 10:80-5.

Feigin, V.L., Lawes, C.M.M., Bennet, D.A., Anderson, C.S. (2003). Stroke epidemiology; a review of population based studies of incidence, prevalence and case-fatality in the late 20th century. *Lancet Neurol*, 2 (1):43-53.

Kelechi, O.E., Maduaburochukwu, C.N.,Adesola, O., Paul, O.N., Azuoma, L.A, Ezinna, E.E. (2014). Epidemiology of stroke in a rural community in southeastern Nigeria. *Vascular Health and Risk Management*, 10: 375-388.

Kolawole, W.W. (2008). The burden of stroke in Nigeria. *International Journal of Stroke*, 3(4): 290-2.

Mensah, G.A. (2008). Epidemiology of stroke and high blood pressure in Africa. *Heart*, 94 (6):697-705

Mukherjee, D. and Patil, C.G. (2011). Epidemiology and the Global Burden of Stroke. *World Neurosurgery*, 76: 85-90.

Ogun SA, Ojini FI, Ogungbo B, Kolapo KO, Danesi MA. (2005). Stroke in South West Nigeria: a 10-year review. Stroke. 36 (6): 1120-1122.

Osuntokun BO, Bademosi O, Akinkugbe OO, Oyediran AB, Carlisle R. (1979). Incidence of stroke in an African city: results from the stroke registry at Ibadan, Nigeria, 1973-197 Stroke. 10 (2): 205-207.

Towfighi A, Saver J. (2011). Stroke declines from third to fourth leading cause of death in the United States: historical perspective and challenges ahead. Stroke, 42 (8):2351-55