Analysis of Reported Cases of Breast Cancer in Abuja, Nigeria: A Time Series Approach

SANI MARY[1], OLANREWAJU S.O[2], OGUNTADE E.S.[3]

[1]National Bureau of Statistics Headquarter, Abuja. FCT. Nigeria

[2, 3] Department of Statistics, University of Abuja, Abuja. FCT. Nigeria

Abstract- Breast cancer is the commonest cancer in women and characterized by regional variations which is prevalently common in low and middle income countries including Nigeria. It is now the most common disease and the second leading cause of death among women. . This study assessed the reported cases of breast cancer for both rural and urban women in University Teaching Hospital Gwagwalada Abuja. The main objectives are to determine the trend movement of the disease and to forecast the trend of the disease. The statistical techniques used for the analysis are logistic regression and autoregressive moving average (ARMA). The result of the analysis revealed that age of cancer patients does not affect chances of survival or death and we used the best model to forecast next eleven cancer victim ages, which ranges between 36 to 56 years.

Indexed Terms- Abnormal growth, ARMA, Cancer, Time series,

I. INTRODUCTION

Cancer is one of the fatal diseases that involve abnormal growth of cells that increases compoundedly. It is a malignant tumor which snicks the surrounding tissues without any wall or boarder through the roots and spread to other parts of human body. If the cancer spread, small tads of cancer cell chuck off the original tumor and move to the other parts of the body. The spread can either be direct, through blood or through lymphatic system. Cancers can touch various organs, and each type of cancer has its unique characteristics. Cancer of various types are known in relation to the location of cancers namely cervical cancer, lung cancer, gynecological cancer, skin cancer, brain cancer, breast cancer etc. the specific type of cancer that is more common to an area or community than the other cancer is breast cancer.

There was a stable rise in frequency of cancer across the period where a total of 1990 cancer cases were recorded consisting of 1001 (50.3%) males and 989 (49.7%) females. Cervical cancer (22.9%), Breast cancer (18.9%), Ovary cancer (8.2%), non-melanoma cancer of skin (6.3%), and Uterus cancer (6.2%) were the most common female cancers. In males, prostate cancer (16.5%), bladder cancer (10.2%), non-melanoma skin cancer (9.9%), colorectal cancer (9.3%) and cancer of connective tissue (6.3%) were most popular. Burkitt's lymphoma cancer (31.4%), other lymph reticular cancers (23.8%) then retinoblastoma cancer (20%) predominated in children.

II. LITERATURE REVIEW

According to [1], he said "Breast cancer usually begins in the cell of the lobules which are the milk producing gland or the duct. The passage that drains the milk from the lobules to the nipples". He said that women whose mother or sister has breast cancer have a high risk of developing the disease themselves. It has been discovered that breast cancer susceptibility genes are from one of the parents. The most common of the genes is the BRCA (gene mutated in breast and ovarian cancer) i.e breast cancer gene.

The gene account for about 10% of all breast cancer cases in families that have these genes, the risk of breast cancer can be very high. However, it is important to realize that 85-90% of breast cancers are not from their families.

A few related works of the use of SARIMA methodology to model economic and financial data include the following [2] works on fitting the best univariate ARIMA model (Box-Jenkin Methodology) to forecast the Pediatrics patient's incoming at Outpatients Medical Laboratory (OPML), Outpatients

Department (OPD). The empirical analysis of their results indicated that SARIMA (1, 1, 1)*(1, 0, 1)4 is best fitted model for patients data for short run forecasting. The estimated results of model showed that Peads incoming is influenced by seasonal variation of data. [3] works on Energy Consumption Forecasting Using Seasonal ARIMA with Artificial Neural Networks Models. The quarterly energy consumption of the United States from January 1973 to June 2015 is used. It aimed to forecast the residential energy consumption in U.S. using the Box-Jenkins methodology and Artificial Neural Network approach and compared their results in order to know the best model for predicting energy consumption in U.S. from their results they concluded that the forecasting accuracy is not quite significant. But, the performance of ANN model is better than SARIMA model in terms of forecasting accuracy from the test data using MAE and MAPE, the opposite result is happened for MSE. While the SARIMA model fits better the historical data (training data) than ANN models using all performance parameters.

[4] Epidemiology of Breast Cancer among Male In University Of Abuja Teaching Hospital Gwagwalada. examined the trends in the prevalence rate of breast cancer in men among tissues submitted to histopathology laboratory university of Abuja Teaching Hospital. A total of 544 data collected consisting of men between the age 17-86years with the mean aged group of 56years and was analysed using Epi-Info version 6.1. it was found that the prevalence of breast cancer among men was 4(2.6%), fibroadenoma197(36.8%), fibrocytic disease 120(22.4%), granulomatouse mastitis 14(2.6%) lactating adenoma 11(2.1%), sclerosingadenosis 8(1.5%), the highest prevalence rate was found between the aged group of 39-48years(50%) followed by 39-48years(25%) and 79-88years (25%) respectively.

[5], fitted a SARIMA (1, 0, 1) x (1, 1, 1)12 model to cucumber market prices in China.

For the forecasting evaluation, and MAPE was used to measure forecast accuracy, it is shown that the ARIMA model build based on past three months data is the best model in term of forecasting two to seven days ahead and ARIMA model based on past six months data is the best model to forecast one day ahead.

[6] observed that the number of reported cases of dengue in Campinas, State of Sao Paulo, Brazil tended to show a maximum in the rainy season and a minimum in the dry season. Such seasonal series may be modeled using a seasonal Box-Jenkins approach. Moreover a seasonal autoregressive integrated moving average (SARIMA) model is proposed and fitted to the call rates. This is with a view to providing basis for possible forecasting of the series.

[7] reviewed the trend of cancer incidence in Lagos University Teaching Hospital (LUTH) from January 2004 to December 2013.Their study is a retrospective study of histologically confirmed malignancies seen at Radiotherapy Department, LUTH from January 2004 to December 2013. Case files were retrieved through the record department, and the information required was extracted with the aid of a data extraction form. They found out that a total of 3,314 new cases of cancer were recorded in LUTH during the study period. The mean age of cancer presentation is 48.52 (±16.44). The median age is 49.00 years with an age range of 1–100 years. The peak age incidence for males was 50–54 years accounting for 10% of all male presentation while females had a peak age incidence of 40–44 years accounting for 14% of female cases. The male-to-female ratio was 1:3. Breast (38.1%), cervical (17.0%), and colorectal cancers (3.3%) are the common ones recorded. In males, the most common cancer was prostate cancer (12.8%) followed by colorectal cancer (4.5%). They concluded that in general, cancer incidence in Nigeria appears low compared to developed countries which may not truly reflect the burden of the disease. This could be due to poor population-based statistics and poor health patronage of orthodox medical care.

[8], worked on Trend Analysis of Cancer Mortality and Incidence in Panama, Using Joinpoint Regression Analysis. Their aim is to utilize Joinpoint regression analysis to study the trends of the incidence and mortality of cancer in Panama in the last decade. Result shows that the trend of age-adjusted cancer mortality in Panama has declined over the last 10 years (−1.12% per year). The cancers for which there was a significant increase in the trend of mortality were

female breast cancer and ovarian cancer; while the highest increases in incidence were shown for breast cancer, liver cancer, and prostate cancer. Significant decrease in the trend of mortality was evidenced for the following: prostate cancer, lung and bronchus cancer, and cervical cancer; with respect to incidence, only oral and pharynx cancer in both sexes had a significant decrease.

[9]Their research was undertaken to evaluate trends in breast cancer incidence in Egypt from 1999 to 2008 and to make projections for breast cancer occurrence for the years 2009–2015. They utilized joinpoint regression and average annual percent change (AAPC) measures with 95% confidence intervals (CI) to describe the trends in breast cancer incidence rates from the Gharbiah Cancer Registry by age and stage at diagnosis and to estimate expected breast cancer caseloads for 2009–2015. Results show thatFrom 1999 to 2008, the AAPC in breast cancer incidence rates in Gharbiah significantly increased among women 50 years and older and among localized tumors (AAPC %, 95% CI, 3.1% to 8.0%). Our results predict a significant increase in breast cancer caseloads from 2009 to 2015 among women aged 30–39 (AAPC %, 95% CI, 0.9% to 1.1%) and among women aged 40–49 years (AAPC %, 95% CI, 1.0% to 2.6%).

III.   RESEARCH DESIGN

The research design to be adopted for this study is a descriptive and Box-Jenkins research design. Descriptive survey design is a research design in which data is collected consistently to explain and predict the given situation. For this purpose non-seasonal Box Jenkins approach is used to find the best fitted, the best forecasting model and the accuracy of the forecasting values are checked by comparing residuals. The steps of the suggested model and its forecasting can be explained in the following steps. Determining whether the time series is stationary or not is a very important concept before making any inference in time series analysis. Therefore Augmented Dickey Fuller (ADF) and Phillips-Person (PP) tests will be used to check the stationarity of the data series. There are several methods that can be used to fit a time series model, among them, ARMA,

ARIMA, and SARIMA model which will be used on the stationary data of this study.

3.1   Population of the Study and Research Sample
The population for this research work is the entire women with breast cancer in Abuja, University of Abuja teaching hospital, Gwagwalada Abuja is the sample of the study.

3.2   Method of Data Collection
Documentary method of data was used for this research work. Documentary method means consulting past record to obtain information in Hospital at Abuja.

3.3   Technique of Data Analysis and Model Specification
The advances in Time series enable researchers to use those techniques in their analysis to reanalyze the traditional rotation analysis applied in earlier studies ([10]). The software that was used for the test is Eviews 4.0 version.

3.4   Seasonal Autoregressive Integrated Moving Average (SARIMA) Models
The ARIMA model (3.5) is for non-seasonal non-stationary data. Box and Jenkins have generalized this model to deal with seasonality. Their proposed model is known as the Seasonal ARIMA (SARIMA) model. In this model seasonal differencing of appropriate order is used to remove non-stationarity from the series. A first order seasonal difference is the difference between an observation and the corresponding observation from the previous year and is calculated as $X_t = Y_t - Y_{t-s}$ s. For monthly time series S=12 and for quarterly time series S=4 This model is generally termed as the SARIMA (p, d, q)x (P,D,Q)$_S$

For a seasonal time series of order s, [11] proposed that $\{X_t\}$ be modelled by:

$$A(L)\Phi(L^s)\nabla^d\nabla_s^D X_t = B(L)\Theta(L^s)\varepsilon_t \qquad (1)$$

where the series must have been subjected to seasonal differencing D times and non-seasonal differencing d times, $\nabla s = 1 - L_s$, being the seasonal differencing operator. Moreover $\Phi(L)$ and $\Theta(L)$ are the seasonal autoregressive and moving average operators

respectively. These seasonal operators are polynomials in L.

Suppose that $\Phi(L) = 1 + \varphi_1 L + \varphi_2 L^2 + ... + \varphi_P L^P$ and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_Q L^Q$, then the time series $\{X_t\}$ is said to follow a multiplicative seasonal autoregressive integrated moving average model of orders p, d, q, P, D, Q and s, designated (p, d, q)x(P, D, Q)$_s$ SARIMA model.

Suhartono (2011), using moving average (MA) symbolism, defines a subset SARIMA model as
$$\nabla^d \nabla^d_s X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + ... + \beta_s \varepsilon_{t-S} + \beta_{s+1} \varepsilon_{t-S-1} \quad (2)$$

where $\beta_{s+1} \neq \beta_1 \beta_s$. Otherwise, it is a multiplicative SARIMA model. If $\beta_{s+1} = 0$, the model is said to be an additive SARIMA model. He goes on to propose the following set of steps for SARIMA fitting:

Fit a subset SARIMA model.

Find out if $\beta s = 0$. If so, the model is additive but If not, find out if the model is multiplicative. If not, the model is subset.

3.5 Forecasting Evaluation

The forecasting ability is subsequently assessed using the symmetric loss functions which are the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The equations are:

$$\text{MAE} = \frac{1}{n} \sum | r_t^2 - \sigma_t^2 | \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(r_t^2 - \sigma_t^2 \right)^2} \quad (4)$$

where r_t^2 is used as a substitute for the realized or actual variance and σ_t^2 is the forecasted variance.

3.6 Model Selection Criteria

Akaike Information Criteria (AIC,) and Schwarz Criteria (SIC) are the most commonly used model selection criteria

$$\text{AIC} = 2K - 2\ln(L) = 2K + \ln\left(\frac{RSS}{n}\right) \quad (5)$$

The Schwarz Information Criterion is given as

$$SIC = K \log n - 2\ln(LL) = K \log n + \ln\left(\frac{RSS}{n}\right) \quad (6)$$

where k is the number of parameters in the model and L is the maximized value of the likelihood function for the model and $RSS = \sum e^2$ is the residual sum of squares.

3.7 Lagrange Multiplier test

The test for conditional heteroscedasticity is the Lagrange multiplier test of Engle. The test is equivalent to the common F-statistic for testing $\alpha_i = 0$ (i=1,..., m) in the linear regression:
$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_m a_t^2 + \varepsilon_t, \quad t = m+1,... ,T(7)$$

Where ε_t denote the error term, m is a prespecified positive integer, and T is the sample size. The null hypothesis is
$$H_0 : \alpha_1 = ... = \alpha_m = 0 \quad \text{Versus}$$
$$H_a : \alpha_i \neq 0 \quad \text{for some } i \in \{1,..,m\}$$

The test statistic
$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1(T - 2m - 1)} \quad (8)$$

$$SSR_1 = \sum_{t=m+1}^{T} e_t^2$$
Where , $\hat{e}_t$ is the least square residual of the linear regression.

$$SSR_0 = \sum_{t=m+1}^{T} (a_t^2 - \varpi)^2, \quad \varpi = \frac{1}{T} \sum_{t=1}^{T} a_t^2 \text{ is the}$$
sample mean of a_t^2.

The test statistic is asymptotically distributed as chi-squared distribution with m degrees of freedom under the null hypothesis. The decision is to reject the null hypothesis if $F > \chi_m^2(\alpha)$, where $\chi_m^2(\alpha)$ is the

upper 100(1-α)th of the χ^2_m or the p- value of F less than α.

IV. DATA PRESENTATION, ANALYSIS AND DISCUSSION OF THE RESULTS

4.1 DATA ANALYSIS
4.2.1 Unit Root Test (Test for Stationarity)

H_0: the data is non-stationary
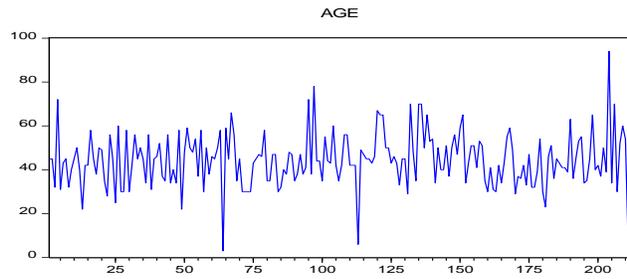H_1: The data is stationary

Figure1. Ages of Cancer patients at level

Table 4.1 Result of Augmented Dickey Fuller Unit Root Test (Test for Stationarity)

Variable	I(0): (at Level)	I(1): (at First difference)
Number of patient	Stationary	

Source: Eview output

Figure 1 above is the time plot of cancer patients that was diagnosed at National Hospital Abuja from 2011 to 2018.

Figure 1 above is the time plot of the number of patient at first difference. From here we can observe that the series is stationary i.e they revolve around the same mean. Augmented Dickey-Fuller test statistic also show that the data is stationary at level with probability of 0.0000, which is less than 0.05 level of significant at 5% confidence interval of the level of accepting the null hypothesis. Therefore, we conclude that the data is stationary or has no unit root.

4.2.2 Identification of the AR and the MA process (Correlogram)

Table 4.2 ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) Plot

Source: Eviews output

Table 4.2 shows that the series is stationary and it is only at lag 12, 15 and lag 23 that it is above the error bound or 95% confidence interval. There for the only combinations we can have are AR(15), AR(23), MA(12), MA(15) and MA(23), giving ARMA(15, 12), ARMA(15, 15), ARMA(15, 23), ARMA(23, 12), ARMA(23, 15) and ARMA(23, 23).

Criteria for the best model are that, the model must:
- have the best significant co-efficient
- have the least Volatility (SIGMASQ)
- Have the lowest AIC (Akaike Info Criterion) and SIC (Schwarz Criterion)
- Have the highest Adjusted R^2.

4.2.3 Test for Best Fit Model

Table 4.2: Result of the all the ARIMA models

	ARMA (15, 12)	ARMA (15, 15)	ARMA (15, 23)	ARMA (23, 12)	ARMA (23, 15)	ARMA (23, 23)
significant co-efficient	1	0	2	2	1	0
Volatility (Sigma^2)	141.907	143.983	141.262	140.387	141.500	143.859
AIC	7.834	7.847	7.830	7.826	7.832	7.848
SIC	7.897	7.911	7.893	7.890	7.895	7.911
Adjusted R^2	0.0438	0.0298	0.0481	0.0540	0.0465	0.0306

Source: Eview

Observing these models, ARMA(23,12) have high significant co-efficient, have lowest volatility, have the lowest AIC and SIC value and have the highest Adjusted R^2. Therefore ARMA(23,12) should be the best model for forecasting. To ascertain that the model is the best fit, the correlogram plot of that ARIMA model will be the determinant.

4.2.4 MODEL DIAGNOSTIC

Table 4.3: Correlogram for ARMA(23, 12)

Included observations: 212
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.099	-0.099	2.0927	
		2	0.082	0.073	3.5520	
		3	0.024	0.040	3.6799	0.055
		4	0.037	0.038	3.9836	0.136
		5	-0.067	-0.066	4.9640	0.174
		6	0.043	0.024	5.3652	0.252
		7	-0.050	-0.036	5.9129	0.315
		8	-0.023	-0.033	6.0318	0.420
		9	-0.117	-0.117	9.1037	0.245
		10	0.032	0.012	9.3308	0.315
		11	0.073	0.108	10.547	0.308
		12	-0.017	0.002	10.610	0.389
		13	-0.015	-0.029	10.661	0.472
		14	0.037	0.013	10.974	0.531
		15	0.113	0.133	13.938	0.378
		16	0.062	0.087	14.834	0.390
		17	0.011	-0.013	14.863	0.461
		18	0.016	-0.018	14.921	0.530
		19	0.070	0.083	16.074	0.519
		20	-0.045	-0.000	16.560	0.554
		21	0.021	-0.010	16.662	0.613
		22	-0.024	-0.039	16.801	0.666
		23	-0.015	0.005	16.858	0.720
		24	-0.085	-0.035	18.604	0.670
		25	0.030	0.013	18.817	0.712

Source: Eviews output

The above table shows that some of the lag structure did not fall between the 95% confidence interval or standard error bounds (i.e it is flat). This also means that all the information has not been captured in the model. Therefore ARMA(23,12) is not the best fit/most ideal for predicting future occurrences. This calls for the inclusion of some of the lags (15 and 23)

Table 4.2: Result of the all the ARIMA models

	ARMA (23, 12)	AR(23) AR(15) MA(12)	AR(23) MA(12) MA(15)	AR(23) A(12) MA(15)MA(23)
significant co-efficient	2	1	0	4
Volatility (Sigma^2)	140.387	137.375	137.818	134.922
AIC	7.826	7.834	7.847	7.817
SIC	7.890	7.897	7.911	7.912
Adjusted R^2	0.0540	0.0438	0.0298	0.091

Source: Eview

4.2.5 Test for Autocorrelation

Table 4.3: Test for Autocorrelation (Ljung-Box Test for Squared residual)

Sample: 1 223
Included observations: 212

		AC	PAC	Q-Stat	Prob
1		-0.008	-0.008	0.0131	0.909
2		0.092	0.092	1.8465	0.397
3		0.005	0.006	1.8517	0.604
4		-0.077	-0.086	3.1484	0.533
5		-0.003	-0.005	3.1502	0.677
6		-0.016	-0.001	3.2078	0.782
7		0.186	0.190	10.825	0.146
8		-0.007	-0.009	10.835	0.211
9		0.048	0.011	11.359	0.252
10		-0.025	-0.032	11.504	0.320
11		-0.035	-0.011	11.787	0.380
12		-0.039	-0.033	12.135	0.435
13		-0.001	0.013	12.136	0.517
14		-0.041	-0.076	12.525	0.564
15		-0.007	-0.009	12.536	0.638
16		0.095	0.093	14.640	0.551
17		-0.049	-0.036	15.201	0.581
18		0.055	0.038	15.916	0.598
19		-0.017	-0.001	15.981	0.659
20		-0.076	-0.078	17.363	0.629
21		-0.056	-0.042	18.113	0.642
22		-0.014	0.009	18.159	0.697
23		0.047	0.028	18.683	0.719
24		0.033	0.042	18.940	0.755
25		0.053	0.012	19.629	0.766

Hypothesis

H$_0$: There exist no autocorrelation among the variables.
H$_1$: There exist autocorrelation among the variables.

Table 4.4 above shows that the model is free from autocorrelation since the probability values are greater than 0.05 at 5% level of significant.
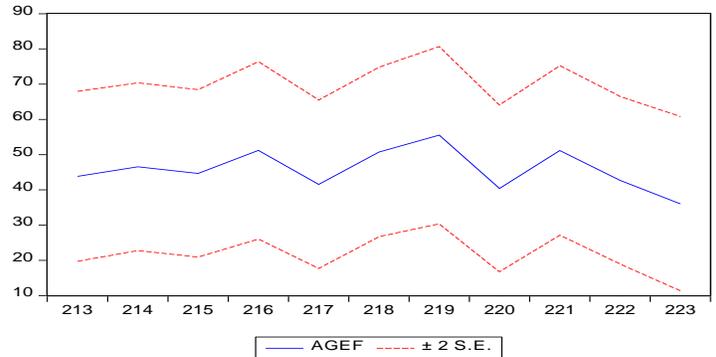
4.2.6 Forecasting

Figure 3. Ages that might possibly have cancer
Source: Eview output.

The chart above shows the ages that might possibly have cancer. The figures are given below.

Table 4.5 Forecast Table

S/N	Age
213	44
214	47
215	45
216	51
217	42
218	51
219	56
220	40
221	51
222	43
223	36

Source eviews

The above result shows that the minimum age of the next eleven (11) patients that can have cancer is 36year and the maximum age is 56year.

4.2 LOGISTIC REGRESSION RESULT

Below is the result of the relationship between age of cancer patients and chances of survival.

4.3.1 Model Summary

Table 4.6a Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	176.400[a]	.000	.000

Source: SPSS Output

The model summary above shows that the -2 log likelihood statistic is 176.400[a]. The statistic measures how the model predicts the mortality rate of cancer patients, the smaller the statistic the better the model. SPSS give us the statistic for the model that had only the intercept, to be 807.388. Adding the variable reduced the -2 log likelihood statistic by 807.388 – 176.400 = 468.393, the chi-square statistic we interpret Table 4.16. And the p-value for the test is p $[X^2(1) < 630.988]$ which is highly significant at 0.05 levels. The null hypothesis is rejected and we conclude that at least one of the beta's coefficients is different from zero.

The Nagelkerke R Square show that the model doesn't fit as the independent variable age of the cancer patients doesn't account for their survival or death (with result value 0.000).

4.3.2 MODEL FITTNESS

The Hosmer- Lemeshow statistic tests the null hypothesis that there is a linear relationship between the predictor variables and the log odds of the criterion variable. Cases are arranged in order by their predicted probability on the criterion variable. A chi-square statistic is computed comparing the observed frequencies with those expected under the linear model. A non-significant chi-square indicates that the data fit the model well.

Table 4.6b   Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	7.418	8	.492

Source: SPSS Output

Table 4.6c Contingency Table for Hosmer and Lemeshow Test

	Dead/Alive = Dead		Dead/Alive = Alive		Total
	Observed	Expected	Observed	Expected	
Step 1 1	4	3.516	19	19.484	23
2	2	2.703	16	15.297	18
3	1	2.984	19	17.016	20
4	3	3.105	18	17.895	21
5	4	2.937	16	17.063	20
6	2	3.358	21	19.642	23
7	5	3.050	16	17.950	21
8	5	3.168	17	18.832	22
9	1	2.987	20	18.013	21
10	4	3.192	19	19.808	23

Source: SPSS Output

The Table 4.6b computed above for the model was C = 7.418 and the corresponding p-value computed from the chi-square distribution with 8 degree of freedom is 0.492 this indicates that the model seems to fit quite well (that the predictions made on table 4.6c are accurate).

Table 4.6d Variables in the Equation

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 0 Constant	1.765	.194	82.405	1	.000	0.171

Table 4.6e   Variables in the Equation

	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Step 1[a] Age	.003	.016	.029	1	.864	1.003	.972	1.035
Constant	1.643	.732	5.040	1	.025	5.173		

a.   Variable(s) entered on step 1: Age
Source: SPSS Output

Table 4.6d above, which is the variables in the equation show that the intercept only model is in (odds) = 1.765. if the exponent are on both sides of this expression we find that our predicted odds [Exp(B)] = 0.171. That is, the predicted odds of death of cancer patients are 0.171. Since 31 of the total number of patients with cancer died and 181 of the total number of patients with cancer were still alive, which gives us an observed odd to be 31/181 =0.171.

Hypothesis
H_0: B=0 where B is regression coefficient.
H_1: not H_0

Decision rule
If 1 falls between the lower bound and the upper bound, that is the null hypothesis is been supported else do not accept the null hypothesis.

From table 4.6e result above, 1 fall between the upper bound and the lower bound for age signifying that age doesn't contribute significantly to the determinant of the patient survival or death.

Table 4.6f Omnibus Tests of Model Coefficients

	Chi-square	Df	Sig.
Step 1 Step	.029	1	.864
Block	.029	1	.864
Model	.029	1	.864

Source: SPSS Output

Also considering age as predictor variable. We see that the Omnibus Tests of Model Coefficients table above gives us a Chi-square of 0.029 on 1 df, significant beyond .864. This is a test of the null hypothesis that age variable in the model does not have significant effect on cancer patient death or survival. The above result shows that age doesn't statistical significantly determine whether a cancer patient will live or die.

Table 4.6g Classification Table[a]

	Observed		Predicted		
			Dead/Alive		Percentage
			Dead	Alive	Correct
Step 1	Dead/Alive	Dead	0	31	.0
		Alive	0	181	100.0
	Overall Percentage				85.4

Source: SPSS Output

The classification table above allows us to correctly classify that 0/31 = 0% of subjects where the predicted event (out of total number of patients diagnosed of cancer and died) was observed. This is known as the sensitivity of prediction, the P(correct | event did occur), that is, the percentage of occurrences correctly predicted. We also see that this rule allows us to correctly classify 181/ 181 = 100% of the subjects where the predicted event was not observed. This is known as the specificity of prediction, the P (correct | event did not occur), that is, the percentage of nonoccurrence's correctly predicted. Overall our predictions were correct 181 out of 213 times, for an overall success rate of 85.4%. from this table, 31 patients was observed to be dead and all the 0 patients were predicted to be dead which in turn is not in line with the observed value having a 0%. While 181 was observed to be alive and 213 was predict to be alive.

V.   SUMMARY

This study examined the Statistical Analysis of Reported Cases (Incidence) Of Breast Cancer at University of Abuja Teaching Hospital, Gwagwalada Abuja. ARMA forecasting models was used on reported cancer cases between "2011", to "2018" .The preliminary analysis of the data obtained shows that the ages of recorded cancer cases are stationary at level. The Parameter of the ARMA models and Models selection, ARMA were estimated with most of the parameter significant at 5%. AIC was used to select the best model that will be used for ARMA model because is the combination of AR and MA model. From the AIC, ARMA (23, 12) was selected to be the best model since it has the smallest AIC. The diagnostic test shows that ARMA (23, 12) shows evidence that the residual are dependent with also the

Q-Q plot result confirm that the model is not normally distributed. This calls for the inclusion of some of the lags that was left behind (i.e lag 15 and 23) which after testing, AR(23) MA(12) MA(15) MA(23) Happens to be the best model for forecasting: Having the smallest AIC and the diagnostic shows that all the data are captured in the model.

CONCLUSION

This research had come out with some finding in putting forth the forecasting model for ages of cancer victim at University of Abuja Teaching Hospital. From the results of the forecasting models, The ARMA (15, 12), ARMA (15, 15), ARMA (15, 23), ARMA (15, 15), ARMA (23, 12), ARMA (23, 15) and ARMA (23, 23) are the adequate forecasting model in estimating ages of cancer victims. Furthermore, in terms of forecasting accuracy, the forecasting models were evaluated using some critarion and from the results, ARMA (23, 12) is most suitable for forecasting. Moreover, using the forecasting models shows the next eleven (11) cancer victim ages, this range between 36 to 56 years. This research work had examined the best forecasting models for the age of cancer victim. The best model was computed using information criterion, AIC and SIC, and diagnostic tests was run on each of the models. The forecasting performance of the models was evaluated using model evaluation performance measures such as the Root Mean Square Error and Mean Absolute Error. The post estimation evaluation carried out revealed various estimating models to capture information for forecasting the ages of cancer victim. The ARMA (15, 12), ARMA (15, 15), ARMA (15, 23), ARMA (15, 15), ARMA (23, 12), ARMA (23, 15) and ARMA (23, 23) were adequate in forecasting the ages of cancer victim over time. But from the results of evaluation obtained, it shows that ARMA (23, 12) is the best forecasting model for ages of cancer patients over time. After carrying out a diagnostic test on ARMA (23, 12), is we discover that all information for forecasting was not captured by ARMA (23, 12). After the inclusion of other lags, AR(23) MA(12) MA(15) MA(23) was discovered to be the best forecasting model and have captured all the information for predicting the ages of future cancer victims.

The result from the logistic regression also indicates that age of cancer patients doesn't affect chances of survival or death. Whether a victim died or lives is not determined by age.

REFERENCES

[1] Henry, A.A. (1993) A to Z Community Health and Social Medicine. Published by 3:00am communication Nigeria.

[2] Mohammed MA, Moles RJ, Chen TF. (2015) Medication-related burden and patients' lived experience with medicine: a systematic review and metasynthesis of qualitative studies. BMJ Open 2016;6:e010035.doi:10.1136/bmjopen-2015-010035

[3] Sissoko D, Laouenan C, Folkesson E, M'Lebing A-B, Beavogui A-H, Baize S, et al. (2016) Experimental Treatment with Favipiravir for Ebola Virus Disease (the JIKI Trial): A Historically Controlled, Single-Arm Proof-of-Concept Trial in Guinea. PLoS Med 13(3): e1001967. doi:10.1371/journal.pmed.1001967

[4] Dangana A, et.al (2015). Epidemiology of Breast Cancer among Male In University Of Abuja Teaching Hospital Gwagwalada. National Biomedical Science Journal.

[5] Luo Xiangzhong , Xiaoqiu Chen , Lin Xu , Ranga Myneni and Zaichun Zhu (2013): Assessing Performance of NDVI and NDVI3g in Monitoring Leaf Unfolding Dates of the Deciduous Broadleaf Forest in Northern China. Remote Sens. 2013, 5, 845-861; doi:10.3390/rs5020845

[6] Martinez J, et al. (2011) Genomic-wide methods to evaluate transcription rates in yeast. Methods Mol Biol 734:25-44

[7] Sowunmi Anthonia, Adewunmi Alabi, et.al (2018). "Trend of cancer incidence in an oncology center in Nigeria".West Africa Journal of Radiology. Volume :25|Issue : 1|Page : 52-56

[8] Politis, Michael, et al (2015). "Trend Analysis of Cancer Mortality and Incidence in Panama, Using Joinpoint Regression Analysis". Medicine: June 2015 - Volume 94 - Issue 24 - p e970.

[9] Kelly A. Hirko et at.(2013). "Trends in Breast Cancer Incidence Rates by Age and Stage at Diagnosis in Gharbiah, Egypt, over 10 Years

(1999–2008)". Journal of Breast cancer journal (www.webmd.com).

[10] CHIA-Chi Chang, Ming-Hua Yang, Hwei-Mei Wen and Jiing-Chuan Chern (2002) Estimation of Total Flavonoid Content in Propolis by Two Complementary Colorimetric Methods. Journal of Food and Drug Analysis, Vol. 10, No. 3, Pages 178-182

[11] Box, G. E. P and Jenkins, G.M., (1976). "Time series analysis: „Forecasting and control," Holden-Day, San Francisco.