A Stochastic Model to Analyze and Predict Transmission Dynamics of Tuberculosis in Ede Kingdom of Osun State

[1]Bashiru, K.A*.,[2]Fasoranbaku, O.A.,[1]Ojurongbe, T.A., [3]Lawal, M.,[4]Abiona, A.A. and [1]Oluwasanmi B.A.

[1]Department of Mathematical Sciences, Osun State University, Osogbo, P.M.B 4494 Osogbo.
[2]Department of Statistics, Federal University of Technology P.M.B 704 Akure, Nigeria
[3] Department of Mathematical and Computer Sciences, Fountain University Osogbo Nigeria
[4]ICT Unit, Federal Polytechnic, Ile Oluji, Ondo State.

Abstract

In this study the stochastic process model for estimating the incidence of tuberculosis (TB) infection in Ede kingdom (Ede North and Ede South Local Government Areas) of Osun State was carried out. The probability generating function approach was used to solve the associated birth process model to obtain the estimate of TB incidence. Also time series analysis was carried out using JMulti software to predict future incidence rate of the disease in the study area. Based on Autoregressive Integrated Moving Average (ARIMA) model, the autocorrelation and partial autocorrelation methods and a suitable model to forecast TB infection was obtained. The goodness of fit was measured using the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). Having satisfied all the model assumptions ARIMA (0,1,1) model with standard error, 6.37086 was found to be the best model for the forecast. It was observed that the forecasted series were close to the actual data series.

Keywords: Stochastic process, Tuberculosis, Incidence rate, Ede kingdom.

Introduction

Tuberculosis is an infectious disease caused by a bacterium called Mycobacterium tuberculosis. It is transmitted from human-to-human and is mainly spread by airborne route. The disease is spread by infection with pulmonary or laryngeal tuberculosis who expectorates bacilli. During coughing, speaking, or sneezing, the patient produces tiny infectious droplets. These particles, called droplet nuclei, are about 1 to 5 microns in diameter. Droplet nuclei can remain suspended in the air for several hours, depending on the environment (Erah and Ojieabu, 2009).Transmission may occur when these infectious droplets are inhaled. Sunlight, UV light and ventilation are effective in decreasing the ability of the droplets reaching the lung. The infectiousness of a patient is linked to the quantity of bacilli contained in his sputa. Patients with sputum smear-positive microscopy are by far the most contagious. Those with smear-negative results are less contagious. Patients whose sputum smears microscopy and culture are both negative are usually not contagious (Tachfouti et al, 2012).

A number of mathematical and statistical models have been developed and analyzed to explain the

*Corresponding author: Bashiru K.A
Email address: kehindebashiru@uniosun.edu.ng

dynamics of TB transmission. Many of these models are described by a system of ordinary differential equations formulated under reasonable assumptions and parameters (Enagi and Ibrahim, 2012; Nainggolan et al, 2013; Ntthiiri, et al., 2016). Some of them have included intervention to overcome the spread of TB in their models. Federal Ministry of Health (2010) recommended intervention is by giving BCG (BaccilusCalmette Guerin) vaccination to new – born babies. An individual vaccinated with BCG could be protected 70% - 80% from TB infection. Some have also developed mathematical model and made use of basic reproduction number, endemic and disease free equilibrium, local and global stabilities to model TB. Aparicio (2009) and Castilo-Chavez and Feng (2009) have also worked on Mathematical modelling of TB epidemics using cluster models and standard compartmental models. Castilo-Chavez and Feng (2009) modelled TB using bio-PEPA approach.

The method of probability generating function (p.g.f) has been applied widely in population studies, especially human reproduction process, birth and death process (Olowofeso and Waema, 2005, Bashiru and Fasoranbaku, 2009, Bashiru, 2014).

In this study, we develop a stochastic model for TB epidemic using birth process to analyze the effect of infection rate on the TB transmission in EDE kingdom and prediction of the future occurrence was carried out.

The Study Area

The study area (Ede kingdom) as presented in Figure 1 has two local governments – Ede North with an area of 111km^2 and Ede South with an area of 219km^2,Ede is one of the fastest growing cities in Osun State with populations of 83,831 and 76,035 for the North and South respectively according to population census 2006.

The two local governments lie between 7°44'N, 4°29'E and 7° 42'N, 4°27'E respectively.

Methodology

The data used in this study was collected from Primary Health Centre in Ede kingdom (Ede North
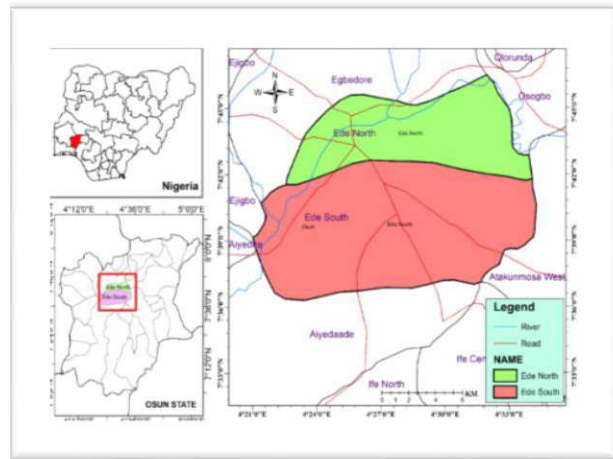

Figure 1: Map of Study Area – EDE Kingdom (Ede North and Ede South LGAs) in Osun State.

Local Government and Ede South Local Government), probability generating function approach was employed to solve the associated birth process model.

Formulation of Birth Process Model.

Consider a population that incorporates susceptible individuals $S(t)$, and infected individuals with Pulmonary TB $I(t)$. Assume that, at time $t = 0$ very new infection is regarded an addition to the Tuberculosis family. Assume further that contacts occur at random time point. Thus, at time $t = 0$, there is only one infected individual and $(n-1)$ susceptible population. So at time interval Δt the infected individual may transmit the disease to any susceptible person with the probability $\lambda_n(t)\Delta t + 0(\Delta t)$

The two different mutually exclusive events listed below will be considered to develop our model.

1. There are "n" units in the system at time t and no TB infection, hence there will be "n" units at time t.

2. There are $(n-1)$ units in the system at time t and 1 event take place to the TB family during time interval Δt.

From the above, there may be two mutually exclusive ways of having n units at time $t + \Delta t$, for $n > 0$. Hence, to determine the probability of " n " additions to the TB family in time t, we develop a governing differential equation for the above situations.

Let the rate of birth within time interval $(t, t + \Delta t) = \lambda_n(t)$

Let the probability of more than one birth in time interval $(t, t + \Delta t) = 0(\Delta t)$

If the probability of n events occurring at time $(t, t + \Delta t) = P_n(t + \Delta t)$

Then, Probability of 1 birth within time interval $(t, t + \Delta t) = \lambda_n(t)\Delta t + 0(\Delta t)$

Probability of 0 birth within time interval $(t, t + \Delta t) = 1 - \lambda_n(t)\Delta t + 0(\Delta t)$.

Therefore, the probability of " n " events at time $(t + \Delta t)$ is obtained as

$$P_n(t, t + \Delta t) = [1 - \lambda_n(t)\Delta t + 0(\Delta t)]P_n(t) + [\lambda_{n-1}(t)\Delta t + 0(\Delta t)]P_{n-1}(t)0(\Delta t) \quad (1)$$

Dividing eq. (1) by Δt and taking the limit $(\Delta t) \to 0$ Thus,

$$\lim \Delta t \to 0 \frac{P_n(t, t + \Delta t) - P_n(t)}{\Delta t} = [-\lambda_n(t)P_n(t) + \lambda_{n-1}(t)P_{n-1}(t)] \quad (2)$$

Therefore,

$$P_n'(t) = -\lambda_n(t)P_n(t) + \lambda_{n-1}(t)P_{n-1}(t) \quad (3)$$

Let $\lambda_n = n\lambda$, equation (3) becomes,

$$P_n'(t) = -n\lambda(t)P_n(t) + (n-1)\lambda(t)P_{n-1}(t) \quad (4)$$

Considering Probability generating function approach (p.g.f) to solve eq. (4).

$$G(z, t) = \sum_{n=0}^{\infty} P_n Z^n$$

Let $\qquad (5)$

$$= P_0 + PZ + P_2 Z^2 + P_3 Z^3 + \ldots$$

for n = 0, 1, 2,3,... $\qquad (6)$

Differentiating equation (6) with respect to z

$$\frac{\partial G(z,t)}{\partial z} = P + 2P_2 Z^2 + 3P_3 Z^3 + \ldots \quad (7)$$

Multiplying equation (4) by Z^n

$$\Rightarrow \frac{\partial G(z,t)}{\partial t} = -\lambda(nP_n Z^n) + \lambda(n - 1P_{n-1}Z^n) \quad (8)$$

For $nP_n Z^n : PZ + 2P_2 Z^2 + 3P_3 Z^3 + \ldots \quad (9)$

For $(n-1)P_{n-1}Z^n : PZ^2 + 2P_2 Z^3 + 3P_3 Z^4 + \ldots \quad (10)$

Substituting eq. (9) and (10) into equation (4) we have

$$\frac{\partial G(z,t)}{\partial t} = (-\lambda Z + \lambda Z^2)\frac{\partial G(z,t)}{\partial z} \quad (11)$$

Let $G(z, t) = \theta$, then eq. (11) becomes;

$$\frac{\partial \theta}{\partial t} + \lambda Z(1 - Z)\frac{\partial \theta}{\partial z} = 0 \quad (12)$$

Equation (12) is a linear probability differential equation with auxiliary

$$\frac{dt}{1} = \frac{dz}{\lambda Z(1 - Z)} = \frac{d\theta}{0} \quad (13)$$

From equation (13)

$$\frac{dt}{1} = \frac{d\theta}{0} \quad (14)$$

and

$$\frac{dt}{1} = \frac{dz}{\lambda Z(1 - Z)} \quad (15)$$

Integrating (14), we have;

$$\theta = k \quad (16)$$

Also from equation (15), we have;

$$\int \lambda dt = \int \frac{dz}{Z(1 - Z)} \quad (17)$$

Solving the above equation, we have;

$$\theta = \frac{e^{\lambda t}(Z - 1)}{Z} \quad (18)$$

Considering Olagunju et al (2007), Bashiru and Fasoranbaku (2009), we formulate the birth rate as

$$\lambda(t) = (1 - \psi)\beta + \alpha \quad (19)$$

where;
λ = Birth rate
ψ = Proportion of new birth that received BCG at birth
α = Rate of expiration of the efficacy of the vaccine
β = Incident rate

Therefore, equation (19) becomes

$$G(Z, t) = \theta = \frac{e^{((1-\psi)\beta + \alpha)t}(Z - 1)}{Z} \quad (20)$$

Under this study, we assume that the initial population $Z = 50000$, $\alpha = 0.6$ and $\beta = 0.3$ with varying value of ψ. The estimated TB incidences are obtained from equation (20) and the trends of TB incidence with varying value of ψ are presented in Figure2.
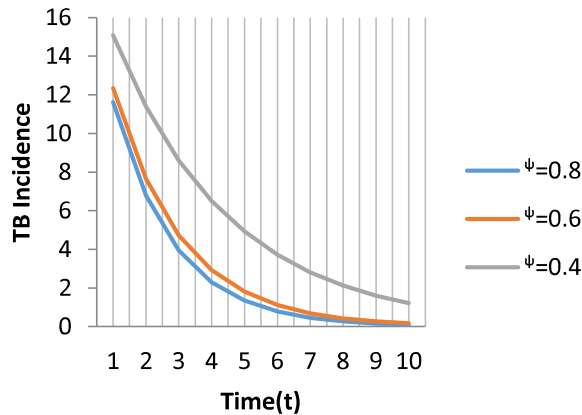

Figure 2: Trend of Tuberculosis (TB) incidences of EDE Kingdom in Osun State.

Time Series Analysis.

Time series analysis methods were employed to predict the future occurrence of the disease using J-Multi software. In time series analysis, the aims are to describe and summarize time series data, fit-low dimensional models and make forecasts.

Stationarity Test of Data.

By stationary test of data, we mean making the raw data time-independent. We carried out stationary test using the following methods:
* Time Plot.
* ACF/PACF (Auto Correlation Function / Partial Auto-Correlation Function.
* Unit Root Test.

Stabilizing the Data.

*After performing all the above tests and the data is non-stationary, we try to stabilize the data by differencing it.
Autoregressive

Where hi = observed value
hi^* = predicted value

*AR-The autoregressive process of order p is denoted by AR (p) and defined by

$$X_t = \sum_{r=1}^{p} \phi_r X_{t-r} + \varepsilon_t \qquad (21)$$

Where $\phi_1 \ldots \ldots \ldots \phi_r$ are fixed constants and ϵ_t is a sequence of independent (uncorrelated) random variables with mean 0 and variance σ^2
AR (1) process is defined by:
$$X_t = \phi_1 X_{t-1} + \varepsilon_t \qquad (22)$$
Generally, AR (p) process is defined by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t \qquad (23)$$

MA-The moving average process of order q is denoted by MA (q) and defined by

$$X_t = \sum_{s=0}^{p} \theta_s \varepsilon_{t-s} \qquad (24)$$

Where $\theta_1,\ldots,\theta_q$ are fixed constants, $\theta_0 = 1$ and ε_t is a sequence of independent random variables with mean 0 and variance σ^2.
ARMA-The autoregressive moving average ARMA (p, q) is defined by

$$X_t - \sum_{r=1}^{p} \phi_r X_{t-r} = \sum_{s=0}^{q} \theta_s \varepsilon_{t-s} \qquad (25)$$

ARIMA-If the original process (Y_t) is not stationary, we can look at the first order difference process:
$$X_t = \nabla Y_t = Y_t - Y_{t-1} \qquad (26)$$

The ARIMA process is employed to generate our tentative models.

Diagnosing Tentative Models to Select the Best Model.

The best model is selected by picking the model with the least value in AIC and SBC compartments respectively. If the least value does not fall on the same model in the AIC and SBC compartments, we deduce the standard error. Standard error is given by:

$$S.E = \sqrt{\frac{\sum (hi - hi^*)^2}{\mu - k}} \qquad (27)$$

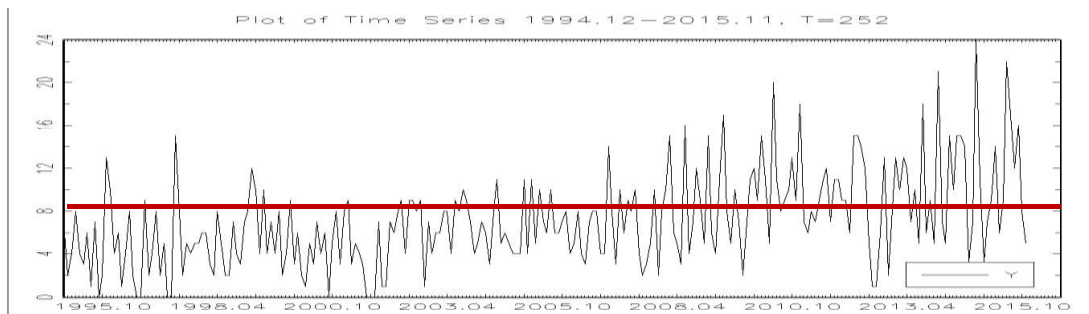μ = no of observations
k = no of order

Results and Discussion

It can be observed from Figure 2, that the TB prevalence is reducing as the proportion of newbirth that received BCG is increasing. At $t = 3$, when the proportion increases from 40% (0.4) to 60% (0.6) the prevalence decreases from 8.2 to 4.4 and also when the proportion is increased from 60% (0.6) to 80% (0.8) the prevalence decreases to 3.8.

Many factors could have contributed to the increases in number receiving the BCG at birth; among these could be the impact of the awareness campaign by the state and federal governments through the media houses ,also during the antenatal programme for pregnant women
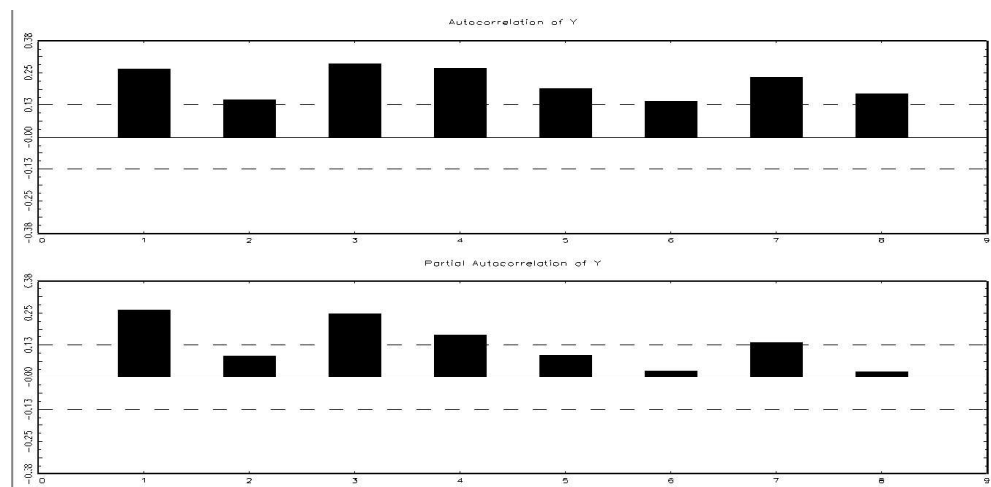
From Figure 3, it can be seen that the data is not stable. It needs to be stabilized by differencing. The red line drawn across the chart confirmed it as it does not pass through the centre. This shows that the data is not stationary or stable. Figure 4, shows that virtually all the spikes are outside the dotted line, which is the acceptable boundary for stability of the data. Hence, the data is not stable. From Figure5, it is clearly shown that the data is stable as the red line bisects the chart into almost equal halves. From Figure 6, the ACF & PACF of Differenced data, it shows that except for First and Second Spikes, the remaining six spikes of ACF are within the dotted acceptable boundary. This shows that the data is stable.
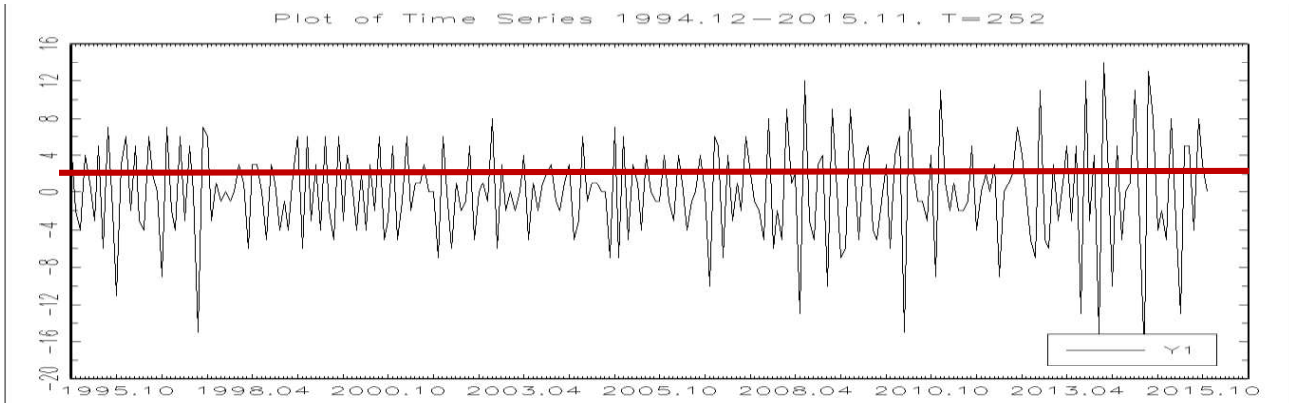

Source: Jmulti Analysis
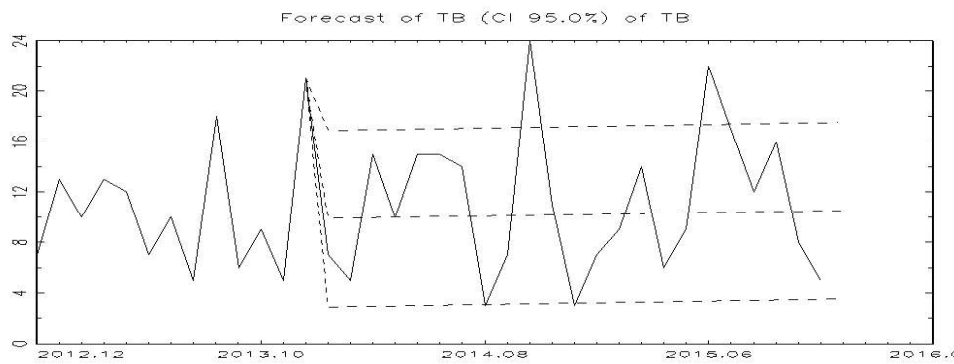Figure 3: Time Plot of Raw Data.

Source: Jmulti Analysis
Figure 4: Autocorrelation function and Partial Autocorrelation Function of Raw Data
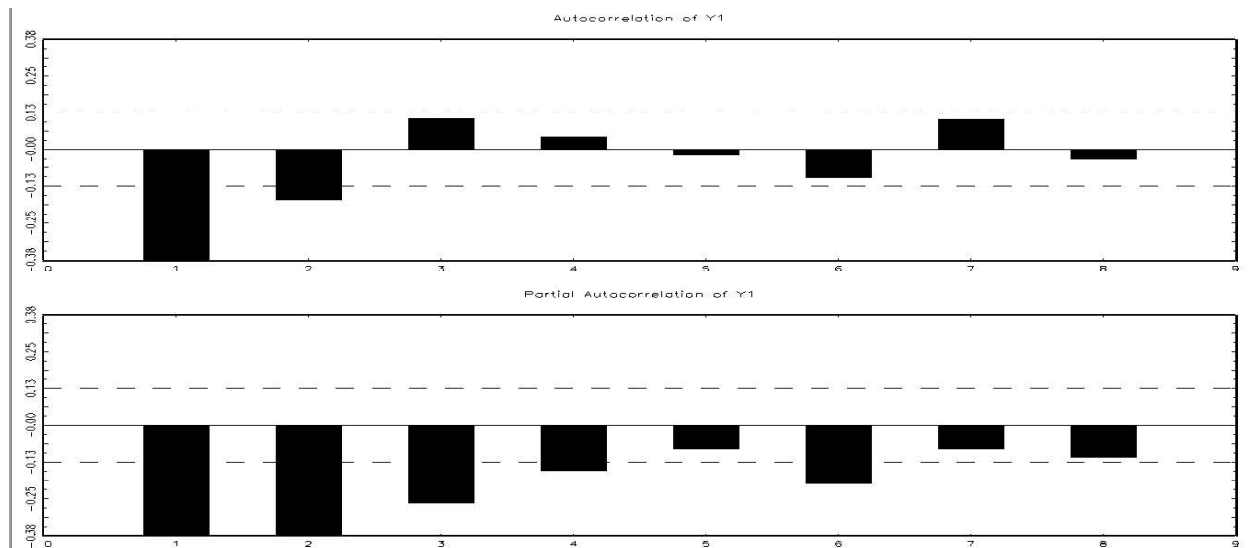
17

Source: Jmulti Analysis

Figure 5: First Difference Time Plot

Source: Jmulti Analysis

Figure 6: Autocorrelation function and Partial Autocorrelation Function of the first difference data.

Source: Jmulti Analysis

Figure 7: Forecast graph of Ede Local Government.

From Figure 7, the future occurrence of the disease was described. The upper course of the graph indicates that the disease is not properly managed and the condition deteriorates, thereby causing increase in the spread of tuberculosis. The middle course of the graph indicates that the disease is not properly managed and the condition remains constant. The lower course of the graph the spread of the disease will be reduced.

Table 1: AIC and SBC of ARIMA Models

ARIMA MODELS	AIC	SBC
0,1,0	1548.2139	1551.7393
0,1,1	1391.5980	1398.6489
0,1,2	1405.9691	1416.5455
0,1,3	1390.7609	1404.8628
1,1,0	1501.4579	1508.5088
1,1,1	1396.0359	1406.6123
1,1,2	1391.7055	1405.8073
1,1,3	1392.0638	1409.6910
2,1,0	1452.9784	1463.5547
2,1,1	1390.9044	1405.0063
2,1,2	1391.9212	1409.5484
2,1,3	1389.3659	1410.5186

Table 2: Standard Error of ARIMA Models

ARIMA MODELS	Standard Error
0,1,1	6.37086
2,1,3	8.58511

After considering all the models based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) also known as Schwartz Information Criterion (SBC), it can be seen from the tab.1 that the lowest value under AIC falls on model ARIMA (2,1, 3) and the lowest value under the SBC falls on model ARIMA (0, 1, 1). As a result, both tests did not choose the same model, and then we use Standard Error Formula as stated in equation 27. Then, the model with the lowest Standard Error is chosing as the best model. In the light of this model ARIMA (0, 1, 1) is the best model as stated in the tab.2 that the standard error for Model ARIMA (2, 1, 3) = 8.5851 and the standard error for model ARIMA (0,1,1) = 6.3709.

Conclusion and Recommendation.

In conclusion, the research study reported in this work has found that TB data in Ede kingdom in the Figure 1 could best be modelled with ARIMA (0,1,1). The study again found out that TB incidence rate in Ede kingdom can fade off with time by initiating comprehensive and strategic preventive measures.

References

Aparicio, P. J. & Carlos Castillo-Chavez (2009). Mathematical Modeling of Tuberculosis Epidemics:. Mathematical Biosciences and Engineering, 6(2), 209 – 237.

Bashiru , K. A. (2014). Stochastic and Dynamic Models Assessment of HIV/AIDS Epidemic in Nigeria. Unpublished PhD thesis. Federal University of Technology, Akure.

Bashiru, K. A., &Fasoranbaku , O. A. (2009). Statistical Modeling of Mother-to-child and Heterosexual Modes of Transmission of HIV/AIDS Epidemic. The Pacific Journal of Science and Technology, 10(2): 966 – 979.

Castilo-Chavez, C., & Feng, Z. (2009). Mathematical Models for the Disease Dynamics of Tuberculosis. New York: Biometrics Unit, Cornell University, Ithaca.

Enagi , A. I., & Ibrahim , M. O. (2012). Preventing Mother to Child Transmission of Tuberculosis Using Bacillus Calmette - Guerin vaccine: A Deterministic Modeling Approach. Research Journal of Mathematics and Statistics, 3(2), 67-71.

Erah, P. O., &Ojieabu, W. A. (2009). Success of the Control of Tuberculosis in Nigeria. International Journal of Health Research,2(1), 3-14.

Federal Ministry of Health. (2010). National HIV sero-prevalance Sentinel survey among pregnant women attending antenatal clinics in Nigeria. Abuja, Nigeria: Federal Ministry of Health.

Nainggolan , J., Sudradjat , S., Supriatna , A. K., &Anggriani, N. (2013). Mathematical Model of Tuberculosis Transmission with Recurrent Infection and Vaccination. Journal of Physics Conference Series. 423, 1-2.

Ntthiiri J.K , Lawi G.O, Akinyi C.O, Oganga D.O, Muriuki W.C, Musyoka M.J, Otieno P.O & Koechi .L (2016). Mathematical modeling of typhoid fever disease incorporating protection against infection. British journal of mathematics and computer Science, 14(1), 1-10.

Olagunju, S. O., Bashiru , K. A., & Olowofeso, O. E. (2007). Statistical Model for Estimating the Rate of Spread of Human Immune-Deficiency: A Case Study of Ondo Kingdom. Research Journal of Applied Sciences, 2(10), 1025-1030.

Olowofeso, O. E., & Waema, R. (2005). Mathematical modeling for human immunodeficiency virus (HIV) transmission using generating function approach. Kragujevac Journal of Science, 27, 115 – 130.

Tachfouti, N., Slama, K., Berraho, M., & Nejjari , C. (2012). The Impact of Knowledge and Attitude on Adherence to Tuberculosis Treatment. The Pan African Medical Journal. 2012; 12, 52 doi:10.11604/pamj.2012.12.52.137