MODELLING THE TREND OF HEART DISEASES IN NIGERIA: APPROACH OF REGRESSION MODEL

[1*]Nkpordee, Lekia; [2]Ibinabo, Magnus Ogolo; [3]Lawal, Oluwafemi Oluwasegun

[13]*Department 0f Mathematics/Statistics,*
Ignatius Ajuru University of Education,
Rumuolumeni, Port Harcourt, Rivers State, Nigeria

[2]*School of Foundation Studies,*
Rivers State College of Health Science and Management Technology,
Rumueme, Port Harcourt, Nigeria

[*]*Corresponding Author Email: lekiafnkpordee1@gmail.com*

Abstract

This study examined modelling the trend of heart diseases in Nigeria using regression model a case study of incidence of Tuberculosis (TB). The accumulated data used for this study was a secondary statistical data extracted from survey information performed in Nigeria on incidence of Tuberculosis (TB) from the World Development Indicators- World Bank from 2000 to 2021. It was found out that based on p-value and Pearson (rho), there was a strong and positive relationship. The logistic regression model was best fit and was used to make forecast of the trend of the incidence of tuberculosis in Nigeria. It clearly showed that incidence of tuberculosis cases will be on the decreased within the next 12 months starting from January, 2022 to December, 2022.

Keywords: *Trend Analysis, Regression Model, Tuberculosis, Heart Diseases, Modelling*

1. Introduction

The WHO has designated Nigeria as one of the countries with a high burden of tuberculosis (TB) for the past two decades in order to encourage focused interventions and advocacy for funding and policies to promote TB control. This initiative has led to focused and practical actions for TB control worldwide. The Nigeria National Tuberculosis Control Programme and its donor partners have recently begun to scale up the availability and accessibility of improved TB diagnosis and treatment regimens. While these efforts are necessary and well-deserved in Nigeria, there are few practical policy options for addressing growing TB risk factors in the population, such as diabetes, alcohol consumption, and cigarette smoking. Country-specific epidemiologic studies which investigate trends in TB disease burden and the attributable risk factors for TB would be useful for public health experts and policy-makers to strengthen TB control and preventive efforts.

Tuberculosis (TB) is still a major public health concern in low- and middle-income countries, where it is the largest cause of mortality from a single infectious disease, ranking ahead of HIV/AIDS and the human immunodeficiency virus. In 2016, 6.3 million new cases of tuberculosis were reported among HIV-negative adults, according to the World Health Organization's (WHO) Global Tuberculosis Report (2017), up from 6.1 million in 2015. Similarly, the Global Burden of Diseases, Injuries, and Risk Factors (GBD) Study 2016 anticipated 9.0 million new and relapse TB-HIV-negative event cases, up from 8.8 million in 2015. These reports highlighted the considerable burden of TB globally. For example, the WHO African region accounted for 25% of the total number of incident cases (i.e., TB-HIV-negative and TB-HIV infection) globally, where Nigeria accounted for8% or 407 cases per 100,000 populations in 2016, up from 322 cases per 100,000 populations in 2015. These estimates may be lower than the actual number of TB cases in Nigeria because only less than a quarter of TB cases (15%) were notified in 2015.

The estimated number of new and relapse tuberculosis cases developing in a particular year, stated as a rate per 100,000 people, is known as tuberculosis incidence. All types of tuberculosis are mentioned, including instances in HIV-positive persons. As new information becomes available and methodologies are developed, estimates for all years are revised, thus they may differ from those previously published.

Regression analysis is a collection of statistical techniques used to estimate the connection between a dependent variable and one or more independent variables in statistical modeling. Several researchers have contributed in developing predictive models. There are several models, like Weibul, Asymptotic Regression/Growth Model, Richards, Gompertz, Hill, Logistic, and S-shaped Curves etc. These models are Sigmoidal Growth Models (Sigmoidal curves) that can see in numerous applications including the following fields; engineering, bioassay, tree diameter, height distribution in forestry, signal detection theory, agriculture, fire size, high-cycle fatigue strength predication, seismological data analysis for earthquakes and economics.

In a statistical model, the unknown variable can be forecasted by using maximum likelihood estimation and least squares method. The method of least squares is a standard method to approximate the solution of over determined systems. In least squares method, the overall solution minimizes the sum squares of error (SSE) made in the results of all the single equation. The first concise presentation of the method of least square was published by Legendre in1805.

Evidence shows that TB mortality among HIV-negative people has declined in many developing countries (including Nigeria); but that TB incidence has remained unchanged in many communities. To guarantee that the TB disease burden in Nigeria continues to decline, it is critical to understand not only the trends in the disease burden, but also the amount to which risk factors contribute to the disease burden in order to influence targeted and high-priority TB programs. Because this was not possible in the GBD capstone publications due to the large size and scope of the study, we have provided a detailed exposition of TB disease burden in Nigeria from the GBD findings, which has also led to further characterisation of the results for other health focus areas and locations. We also hope that by condensing the data for TB burden in Nigeria, we will raise awareness and comprehension of TB estimates in Nigeria for tuberculosis, doctors, national, and worldwide health professionals are needed to prevention and control programmes, especially that Nigeria is the largest recipient of developmental assistance for health in Sub-Saharan Africa. Using data from the World Health Organization, the current study aims to highlight the incidence, prevalence, mortality, disability-adjusted life years, and risk factors for tuberculosis in Nigeria from 2000 to 2021 via Development Indicators- World Bank Study 2021.

Purpose of the Study

The aim of this present study is to examine modelling the trend of heart diseases a case study of the incidence of Tuberculosis (TB) in Nigeria: approach of regression model. The study's specific goals are as follows:

1. To obtain the series plot of the incidence of Tuberculosis (TB) in Nigeria
2. To estimate the descriptive statistics of the data under study
3. Obtain the estimates of the parameters ($\beta_{0,}\, \beta_{1,} \beta_2$)of the Logistic growth model.
4. To obtain the (SSR, R^2-value, R^2adjusted-value, F-value, rho-value, p-value and Log-likelihood) for the model.
5. To estimate the (AIC, BIC, HQC and Durbin-Watson) values for the model
6. To forecast the trend of Tuberculosis (TB) cases in Nigeria using the best fitted regression model.

Scope of the study

The study is centred on modelling the trend of heart diseases a case study of Tuberculosis (TB) incidence in Nigeria: approach of regression model to build a best fit regression model that will be used to forecast the trend of the Tuberculosis (TB) data under this study, specifically, annual Tuberculosis (TB) cases in Nigeria matching the data length of 22 based on number of persons admitted from 2000 to 2021.

However, due to inadequate time, unavailability of information and lack of related literature‘s constraint, the researcher focuses his attention on the yearly cases of Tuberculosis (TB) incidence in Nigeria using secondary records from the World Development Indicators- World Bank Data - 2021, Average Number of Hospital Admissions. The growth model to be used in this study is the Logistic Regression Model among other nonlinear regression models.

2. Literature Review

2.1 Concept of Tuberculosis

TB is a disease that has been around for millennia. It was first used in 1839 and is derived from the word "tubercular," which refers to little scars in tissues (i.e. small lumps). An outbreak of the disease swept Europe and the United States in the nineteenth century. This sparked a flurry of research into the causes and treatments for tuberculosis (WHO, 2008).

Causes of Tuberculosis

Multiple factors contribute to the global increase in TB infection. The human immunodeficiency virus (HIV) that causes acquired immune deficiency syndrome (AIDS) poses the greatest risk to activate and progress TB. People with HIV have a weakened immune system that increases their susceptibility to TB, and in these people, TB often progresses rapidly from the primary to secondary stage. This is observed in statistics showing that the increase in TB incidence is highest in Africa and Asia, which consists of the highest number of people infected with HIV (WHO, 2009a; 2009b).

TB resurgence is also contributed by the failure of patients to complete the full six months of antibiotics therapy required to cure the disease. When people start to feel better, they often stop taking antibiotics. Successful TB treatment, on the other hand, necessitates treatment beyond the onset of symptoms. Patients who do not adhere to prescribed treatments risk becoming actively infectious and spreading the sickness to others. Furthermore, failure to complete full therapy can cause emergence of TB bacterial strains with acquired drug resistance, thereby further complicating treatment by increasing the length and cost of therapy.

Other factors that contribute to the spread of TB include migration, international air travel, and tourism. The extreme difficulty of screening immigrants and travellers for TB allows the disease to cross international borders easily. The substantial increase in homelessness and the related circumstances of poverty, overcrowding, and malnutrition also contributed to the increased incidence of TB in the United States and other industrialized countries during the early 1990s. The current TB resurgence has been controlled in industrialized countries with good public health systems. However, curbing the spread of TB on a global scale will require on-going international efforts. In the future, combating TB throughout the world will require advances in molecular biology, in-depth research on the genetics of TB in order to understand drug resistance, and the on-going research of new medications, as well as the possibility of synthesizing more vaccines (Atun et al., 2005; Rios et al., 2000, Bacar et al., 2008).

Cures for Tuberculosis

In the literature, several TB cures have been proposed. The sickness struck American physician Edward Trudeau twice, in 1873 and 1876. He went to Saranac Lake in the Adirondack Mountains of New York to spend his final days when he felt he was dying. He ascribed his recuperation to the fresh air of the highlands when his symptoms gradually disappeared. Trudeau established the first American sanatorium in 1885, which later served as a model for numerous others. That became the main stay of TB treatment in the late 19th century and 20th century. By 1930, there were 600 sanatoriums in the United States, with a total of 84,000 beds. Trudeau also established the Trudeau laboratory, which was responsible for training most physicians versed in the treatment of TB.

Early in the 19th century, TB was considered a refined disease – that is, one that affected artistic, morally superior individuals. However, as the epidemic continued and claimed a larger circle of people, often the poor and the disadvantaged. In the absence of scientific understanding, TB was often blamed on the victims themselves, and TB was linked to a person's lifestyle. The search for the true origin of tuberculosis continued, and in 1882, German physician Robert Koch found the bacterium that caused the disease and utilized basic but exact observations and tests to verify the presence of the bacteria and how it was transmitted (WHO, 2008, 2009a, 2009b).

In Paris, French bacteriologists Albert Calmette and Camille Guerin worked with a virulent strain of bovine (low) tubercle bacillus at the Pasteur institute. In 1924, they prepared the BCG vaccine in hopes of protecting the world against TB. The vaccination was given to a newborn baby who was at high risk of contracting tuberculosis. The vaccine was successful and the child never contacted the disease. Selman Waksman, an American microbiologist, discovered Streptomycin from the fungus Streptomycin Lavendula in 1944, marking the beginning of modern antibiotic therapy for tuberculosis. Sources of drug therapy and declining rates of disease incidence and mortality over the next 30 years instilled a sense of confidence in public health officials that TB could be conquered. As antibiotic therapy became the primary treatment, mortality rates from TB decreased significantly. Deaths from TB in the United States dropped from 188 per 100,000 in 1904 to about 1 per 100,000 in 1980. Between 1953 and 1984, the average annual decrease in cases was roughly 3%. As a result, funding for public health programs in the United States, including those for the prevention and treatment of TB, was drastically curtailed in 1980s (Bacar et al., 2008; Atun et al., 2005; Rios et al., 2000; Atun et al., 2005).

2.2 Empirical Review

Purnachandra and Ayele (2013) researched on the solutions of rate-state equation explaining biological growths. In the study, the rate-state ordinary differential equation $f'(t)= r_t f(t)$ which describes biological growths was considered. They constructed growth function $f(t)$ and relative growth function r_t for the models: Generalized logistic, Case of generalized logistic, Richards, Von Bertalanffy, Brody, Classical logistic, Gompertz, Weibull, Generalized Weibull, Monomolecular and Mitscherlich. Detailed derivations were presented bearing in mind the non-mathematicians studying in the fields of Biological sciences and the paucity of information relating to these derivations in literatures. They accepted that there is a constraint on the acceptability of the values of B in each of the models. They further stated that the rate-state equation can generate more useful solutions. Further suggestions were made on the possibility of exploring the potentials in developing such models, assessing the relationships between the models and their inflection points.

Dagogo, Nduka, and Ogoke (2020) carried out a comparative analysis of Richards, Gompertz and Weibull Models. They stated for nonlinear models that there is a need for starting point (initial variables or guess values) to begin the optimization process. In the study, they suggested that nonlinear model must be expressed with the parameter names declared and initial variable values specified, then the variables are predicted through an iterative approach. A computer program can be used for the estimation of three growth models (Richards, Gompertz and Weibull model) employing a modified version of the Levenberg-Marquardt method for solving non-linear regression model. The growth models were decomposed by additive and multiplicative error terms which help in revealing the most adequate model for growth studies. The challenge of the initial variables was addressed by second-order regression equation before an iterative approach employed. The result shows the final estimate of the variables, standard errors, p-values, and model adequacy criteria, employed to determine the most appropriate growth model. The study was able to reveal that the Weibull Growth Model with Additive Error Terms as the best growth model. The researchers recommended and suggest the Weibull Growth Model for further growth studies.

Yaya*et al.,* (2019) carried out a study on the comparison of nine growth curve models to describe growth of partridges (Alectorischukar). The study assessed nine non-linear growth curve models. The body weight measurements of a total of 178 partridges (Alectorischukar), 93 females, and 85 males, were utilized to calculate and test the models' goodness of fit. The R^2 values for total partridges, females, and males in Brody, Gompertz, Logistic, von Bertalanffy, asymptote regression, exponential, Monomolecular, Richards, and Weibull-type regression were 0.985, 0.980, and 0.985, respectively. and 0.984, 0.997, 0.998 and 0.998, 0.996, 0.999 and 0.999, 0.995, 0.995 and 0.996, 0.985, 0.980 and 0.984, 0.891, 0.871 and 0.892, 0.985, 0.980 and 0.984, 0.997, 0.999 and 0.999, 0.997, 0.999 and 0.999, respectively. The R2 values for Gompertz, Logistic, von Bertalanffy, Richards and Weibull-type were >0.99, while the exponential (<0.90) had the lowest. What's more, the Gompertz, Logistic, Richards and Weibull-type models best explained the data because of lower mean square error (MSE) Akaike's information criteria (AIC) and Schwarz Bayesian information criterion (BIC), higher adj. R^2(Adjusted coefficient of determination) and r(the correlation coefficient between measured body weight and predicted body weight) and there was not an autocorrelation between the residual values. As a result, based on goodness of fit criteria; R^2, adj.R^2, MSE, r, AIC, BIC values, the Weibull-type model best described live weight data of the Partridges (Alectorischukar).

Eke *et al.,* (2015) carried out a comparative analysis of three time series trend models on gross domestic product in Nigeria. This work fitted three time series trend models namely, linear trend model, quadratic trend model and exponential trend model on Gross Domestic Product of Nigeria using an annual data from 1982 to 2012. It was found out that the exponential trend model had the least MAPE and fitted the data appropriately. The exponential trend model was used to make a five year for forecast of Nigeria's Gross Domestic Product, which showed that the country's Gross Domestic Product will be on the rise within the next five years.

Adejumo, Akinrefon, Odetunmibi and Ademola (2013) investigated tuberculosis: a study of patients in Nigeria using binary logit models. This study aims to examine the influence of age, state, and year on gender of Tuberculosis patients using binary logitmodeling. Binary logit models were computed using data sets of registered Tuberculosis patients from 2006 to 2009. Results suggest that males face higher risks for Tuberculosis as compared to females in all age groups, states, and years. In addition, risk variation was observed in age groups, states, and years. This study contributes to a better understanding of Tuberculosis patients in Nigeria in terms of age, state, year, and gender.

3. Methods and Materials

3.1 Research Design: The design for this study was cross sectional research design which focused on modelling the trend of heart diseases in Nigeria using regression model a case study of incidence of Tuberculosis (TB).

3.2 Nature and Source of Study Data: The accumulated data used for this study was a secondary statistical data extracted from survey information performed in Nigeria on incidence of Tuberculosis (TB) from the World Development Indicators- World Bank from 2000 to 2021. This data can be provided on demand.

3.3 Data Analysis Tools: The Gretl (version 32) programme was used to acquire the parameters which constitute the models. To facilitate parameter estimation, the researcher made use of this software in estimating the parameters for Logistic Regression Model. This software was also used to estimate the values of AIC, BIC, Hannan-Quinn, R^2, R^2-Adj., SSE and MSE.

3.4 Model Specification: The growth model to be used in this study is the Logistic Regression Model.

Logistic Growth Model

The Logistic model with four parameters is expressed as

$$y_i = \beta_0 + \left[\frac{(\beta_1 - \beta_0)}{1 + \ell^{\left(\frac{(x_i - \beta_2)}{\beta_3} \right)}} \right] \quad (3.1)$$

Where

ℓ represents Euler number (e = 2.71828)

x_i represent time

β_0 represent upper asymptote when time approaches $+\infty$

β_1 represent positive number of the shape parameter related to initial time (or displacement along the x axis)

β_2 represent positive number of the growth range

y_i is the i^{th} observation at time t_i

Link functions: Minitab provides three link functions-logit (the default), normit (also called probit), and gompit (also called complementary log-log)-allowing you to fit a broad class of binary response models. These are the inverse of the standard cumulative logistic distribution function (logit), the inverse of the standard cumulative normal distribution function (normit), and the inverse of the Gompertz distribution function (gompit). This class of models is defined by:

$$g(p) = b_0 + \mathbf{x'b} \tag{3.2}$$

where:

p = the probability of a success

g(p) = the link function (described below)

x = a vector of predictor variables

b = a vector of unknown coefficients associated with the predictors

The inverse of a distribution function is the link function. The following are the link functions and their corresponding distributions:

Name	Link Function	Distribution
logit	$g(p) = \log_e (p/ (1 - p))$	logistic
normit (probit)	$g(p) = F^{-1}(p)$	normal
gompit (complementary log-log)	$g(p) = \log_e (-\log_e(1 - p))$	Gompertz

You want to choose a link function that result in a good fit to your data. To compare fits using different link functions, goodness-of-fit statistics can be employed. For historical reasons or because they have a special connotation in a discipline, certain link functions may be used. For example, an advantage of the logit link function is that it provides an estimate of the odds ratios.

Characteristics of Estimated Equation

Log likelihood: Derived from the individual probability density functions, the expression is maximized to yield optimal values of b. The log likelihood cannot be used alone as a measure of fit because it depends on sample size but can be used to compare two models.

The formula for the log-likelihood used here is:

$$L(b) = S_j (y_j \log p_j + (m_j - y_j) \log(1 - p_j)) \tag{3.3}$$

where p_j = probability, y_j = response, and m_j = number of trials or subjects associated with the j^{th} factor/covariate pattern. If your data contain one trial per factor/covariate pattern, m_j is 1. If not, m_j is the number of trials per factor/covariate pattern.

Coefficients: With a binary response, the estimated coefficient for each predictor represents the change in the log of P(success)/P(failure) for every unit changed in the corresponding predictor while the other predictors are held constant.

To find the value of b that maximizes L(b), L(b) is differentiated with respect to b_0 and b_i and the resulting expressions are set to zero:

$$S_j (y_j - m_j p_j) = 0 \text{ and } S_j x_{ji} (y_j - m_j p_j) = 0 \quad (3.4)$$

These expressions are nonlinear in $b_0, b_1,\ldots,b_p$ and Minitab uses an iterative reweighted least squares method to obtain the estimates of the coefficients, which is equivalent to maximum likelihood estimation.

Z: Used to determine whether the predictor is significantly related to the response. A significant association is indicated by larger absolute values of Z. The formula is:

$Z = b_i$ / standard error

For small samples, the likelihood-ratio test may be a more reliable test of significance.

p-value (P): Used in hypothesis tests to decide whether to reject or fail to reject a null hypothesis. The p-value is the probability of obtaining a test statistic that is at least as extreme as the actual calculated value, if the null hypothesis is true. The 0.05 cut-off number for the p-value is often employed. We reject the null hypothesis if the estimated p-value of a test statistic is less than 0.05.

Odds Ratio

Useful in interpreting the relationship between a predictor and response. The odds ratio is provided only if you select the logit link function, which is the default. Any nonnegative value can be used as the odds ratio (q). The odds ratio of 1 is used as a comparison point. If $q = 1$ indicates there is no association between the response and predictor. If $q > 1$, the odds of success are higher for the reference level of the factor (or for higher levels of a continuous predictor). If $q < 1$, the odds of success are less for the reference level of the factor (or for higher levels of a continuous predictor). Stronger degrees of correlation are represented by values that are further away from 1.

Note For the binary logistic regression model with one covariate or factor, the odds of success are:

$$\frac{p_j}{1 - p_j} = \exp(b_0 + b_1 x) \quad (3.5)$$

The exponential relationship provides an interpretation for b: The odds increase multiplicatively by e^{b1} for every one-unit increase in x. The odds ratio is equivalent to $\exp(b_1)$. For example, if b is .75, the odd ratio is exp(.75), which is 2.11. This indicates that there is a 111% increase in the odds of success for every one unit increase in x.

Diagnostic Measures

Pearson residuals: Elements of the Pearson chi-square that can be used to detect ill-fitted factor/covariate patterns. Minitab stores the Pearson residual for the jth factor/covariate pattern. The formula is:

$$r_j = \frac{(y_i - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \tag{3.6}$$

where:

y_j = the number of successes for the j^{th} factor/covariate pattern

m_j = the number of trials for the j^{th} factor/covariate pattern

$_j$ = estimated probability for the jth factor/covariate pattern

Goodness-of-fit Statistics

1. Pearson: A summary statistic based on the Pearson residuals that indicates how well the model fits your data. Pearson chi-square (x^2) isn't useful when the number of distinct values of the covariate is approximately equal to the number of observations, but is useful when you have repeated observations at the same covariate level. Higher x^2 and lower p-values values indicate that the model may not fit the data well.

The formula is:

$$x^2 = \Sigma_j r_j^{\,2} \tag{3.7}$$

where r_j^2 = Pearson residual for the j_{th} factor/covariate pattern.

2. Akaike Information Criteria (AIC): Akaike (1974) developed a procedure which is known as Akaike Information criteria. The form of this statistics is given below:

$$AIC = n \ln\left[\frac{SSE}{n}\right] + 2(k) \tag{3.8}$$

Where;

n=sample size

k=number of parameter and

SSE=sum of square error.

3. **SCHWARZ Criterion:** Craven and Wahba (1978) developed a procedure which is known as SCHWARZ (BIC) criteria. The form of this procedure is given below as:

$$BIC = n\ln\left[\frac{SSE}{n}\right] + k\ln(n) \tag{3.9}$$

The value of SCHWARZ (BIC) will also decrease provided there are at least 8 observations (Ramanathan, 1995).

4. Hannan-Quinn **Criteria (HQC)**

The HQC is a model selection criterion that involves selections among a finite set of models. The HQC is given by

$$HQC = \mathrm{T}\ln\left[RSS/\mathrm{T}\right] + p\ln\left(\ln(\mathrm{T})\right) \tag{3.10}$$

5. Coefficient of Determination (R^2): R^2 Is one theimportant statistical parameters used in decision making and for statistical inferences. It is a procedure used to determine the percentage of effects of one or more variables over the others. The form of this procedure is represented as:

$$R^2 = \frac{SSR}{SST} \tag{3.11}$$

Where;

SSR= Sum of square Residual

SST= Sum of Square Total

RESULTS

The Actual Data Time Series Plot

The time series plot of the data in fig. 1 below, displayed the incidence of tuberculosis cases in Nigeria (2000 – 2021). The plot shows trend movement with seasonal variation with decreasing behaviours.

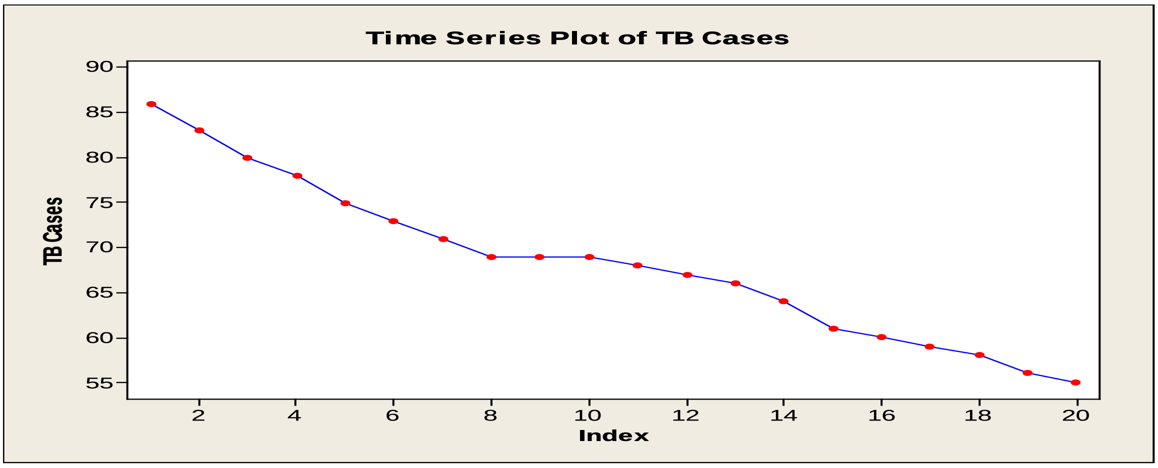

Fig.1 Time series plot of the actual data

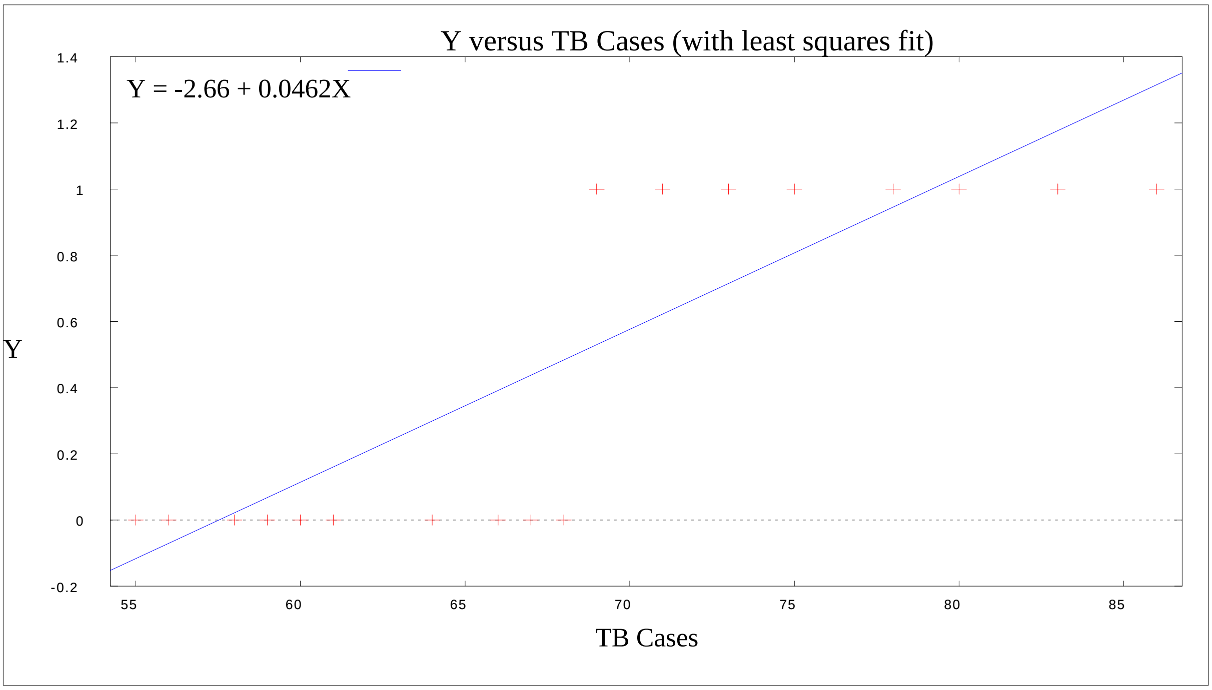

Fig.2 The Scatter Plot of the Response(Y) variable Versus TB Cases

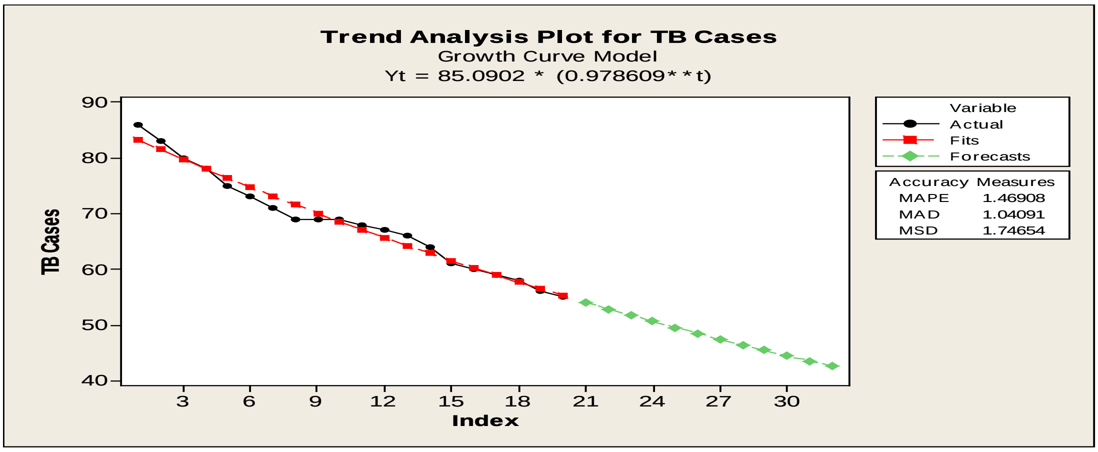

Fig.3: The Trend Plot of the TB Cases

Table 1: Summary of Descriptive Statistics, using the Observations (2000 – 2021)

Variable	Mean	Median	Minimum	Maximum
TB Cases	68.3500	68.5000	55.0000	86.0000
Y	0.500000	0.500000	0.000000	1.00000
Variable	**Std. Dev.**	**C.V.**	**Skewness**	**Ex. kurtosis**
TB Cases	8.89870	0.130193	0.327243	-0.747107
Y	0.512989	1.02598	0.000000	-2.00000

Data in table 1 above shows the descriptive results of the incidence of tuberculosis in Nigeria with mean of 68.35(SD=8.899). The minimum cases of the tuberculosis cases were 55 cases (in 2019) and the maximum recorded cases were 86 (in the year 2000). This shows a significant decrease in the recorded cases. The result showed based on Skewness that the distribution of tuberculosis cases in Nigeria is normally skewed. Again, the result also revealed based on Kurtosis that the distribution of tuberculosis cases in Nigeria is normally kurtik.

Table 2: Logistic Model Summary, using Observations 2000-2021 (T = 22)

Dependent variable: TB Cases

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	0.468196	0.0915397	5.1147	0.00007	***
Y	0.675989	0.129457	5.2217	0.00006	***

Footnote: Significant at $p< 0.05$

The regression model is

y_t= 100 / (1 + exp(-X*b))

$$y_t = \frac{100}{(1+\exp^{(-0.468196x)})}$$

The table 2 summarised parameter estimates, t-test, p-value and standard error of the actual values for the TB cases in Nigeria. The results of parameters obtained are significant at 5%.

Table 3A: Statistics based on the Transformed Data

Sum squared resid	1.508313	S.E. of regression	0.289474
R-squared	0.602355	Adjusted R-squared	0.580264
F(1, 18)	27.26656	P-value(F)	0.000058
Log-likelihood	-2.531366	Akaike criterion	9.062732
Schwarz criterion	11.05420	Hannan-Quinn	9.451487
rho	0.658385	Durbin-Watson	0.399662

Table 3B: Statistics based on the Actual Data

Mean dependent var	68.35000	S.D. dependent var	8.898699

Sum squared resid	541.5588	S.E. of regression	5.485125

Data in table 3a shows that the logistic regression models has log-likelihood estimate of (-2.531366) and equal p-value of (0.000058). This implies that the model best describe the data. This result also revealed that at 100 iterations, logistic regression model converges. The Pearson (rho) value of 0.658385 shows a strong and positive relationship. The R^2 value of (0.602355) shows that 60.2 percent of the variation in the TB cases can be attributed to the model whereas the other 39.8 percent is due to other elements not included in the model.

Table 4: Forecasts of the TB Cases in Nigeria

Period	Forecast
January, 2022	54.0353
February, 2022	52.8794
March, 2022	51.7483
April, 2023	50.6414
May, 2024	49.5581
June, 2025	48.4981
July, 2026	47.4607
August, 2027	46.4455
September, 2028	45.4520
October, 2029	44.4797
November, 2030	43.5283
December, 2031	42.5972

Discussion of Finding

In this study, logistic regression model was used to analyze the trend behaviour of the incidence of tuberculosis in Nigeria. Via estimating and analysing the log-likelihood and model selection criteria of R^2, Adj.R^2, MSE, SSE, AIC, BIC and Hannan-Quinn values of the model, the logistic regression model was assumed to be able to describe the incidence of tuberculosis in Nigeria. However, based on p-value and Pearson (rho), there was a strong and positive relationship. The forecast using the logistic regression model shows that there will be a significant decreased in the incidence of tuberculosis in Nigeria from January 2022 to December 2022. This finding is in agreement with that of Adejumo, Akinrefon, Odetunmibi and Ademola (2013) that investigated tuberculosis: a study of patients in Nigeria using binary logit models. Their finding revealed that males face higher risks for Tuberculosis as compared to females in all age groups, states, and years.

Conclusion

The fitting of appropriate trend model to the tuberculosis cases in Nigeria (2000-2022) showed that logistic regression model fit the data and with the smallest AIC value. Since the logistic regression model was best fit, it was used to make forecast of the trend of the incidence of tuberculosis in Nigeria. It clearly showed that the incidence of tuberculosis cases will be on the decreased within the next 12 months starting from January 2022 to December 2022.

Recommendation

Based on the findings of this study, the scholar recommended the following:

1. A national public health agenda for TB is required, including the necessary public health services as its foundation. Addressing the challenges and concerns raised in this study, as well as others, can contribute to a stronger public health focus on TB.
2. Serious attention should be given to the rate of poverty, unemployment, lack of education, weak social support structures, and related disparities as factors that may affect health and well-being. These concepts may be beneficial to efforts to enhance health that is capable of reducing the rate of TB in Nigeria.
3. Important programs and strategies focused on or coordinated with socioeconomic and environmental factors that affect health (TB) should often be implemented and should as well be more effective.
4. Government should also implement policies that will regulate programmes geared towards TB spread rate and as well make provisions for facilities to public health sectors.

References

1. Adejumo, A. O. (2005). *Modeling for raters' agreement measurements using a generalized linear(log linear) model.* Frankfurt am Main: Peter Lang.

2. Adejumo, A. O. and Arabi, F. J. (2013). A study of HIV sero-prevalence surveillance survey data using log linear model. *Global Journal of Pure and Applied Mathematics*. Vol. 9 No. 1, pp. 73-81.

3. Agresti, A. (2002). *Categorical data analysis*. New York: John Wiley and Sons.

4. Atun, R. A, Samyshkin, Y. A, and Drobniewski, F. (2005). Seasonal variations in tuberculosis hospitalization in Russia: hospitals as social care institutions. *European Journal of Public Health*, Vol. 15, pp. 350-354.

5.Bacaër, N., Ouifki, R., and Pretorius, C. (2008). Modeling the joint epidemics of TB and HIV in a South African township. *Journal of Mathematical Biology*, Vol. 57, pp. 557-593.

6.Bayaga, A. (2010). The use and use of multinomial logistic regression in risk analysis. *Journal of Applied Quantitative Methods*, Vol. 5, No. 2, pp. 288-297.

7.Christensen, R. (1997). *Log linear models and logistics regression*. New York: Springer-Verlag.

8.Hosmer, D. W., and Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley.

9.Peng, C. J., Lee, K. L., and Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Education Research*, Vol. 96 No. 1, pp.3-14.

10.Rios, M., Garcia, J. M, Sanchez, J. A, et al. (2000). A statistical examination of pulmonary TB seasonality. *European Journal of Epidemiology*, Vol. 16 No. 5, pp. 483-488.

11.World Health Organization. (2008). Global tuberculosis control: surveillance, planning, financing. *WHO Report 2008*, Geneva.

12.World Health Organization. (2009a). Tuberculosis control in the South East Asia region. *WHO Report 2009*, Geneva.