



STATISTICAL MODELLING OF GENETIC DISORDER IN NIGERIA: A STUDY OF SICKLE CELL DISEASE

¹Nkpordee, L., & ²Wonu, N.

¹⁻²Department of Mathematics/Statistics, Ignatius Ajuru University of Education,
P.M.B 5047, Port Harcourt, Rivers State, NIGERIA

+Corresponding author(email): wonu.nduka@iaue.edu.ng

Abstract

This present work investigates trend analysis of genetic disorder in Nigeria a case study of Sickle Cell Disease (SCD). Four different trend models of time series were fitted such as; linear, quadratic, exponential and s-curve trend models on sickle cell disease cases in Nigeria (annual data set from 2010 to 2016). The best trend model was then identified using model accuracy measures [Mean Absolute Percentage (MAPE), Mean Absolute Deviation (MAD), & Mean Squared Deviation (MSD)]. The findings showed that of all three model accuracy methodologies tested, the quadratic and s-curve trend models compete favourably. The s-curve trend model, on the other hand, had the lowest MAPE and best fit the data. The s-curve trend model was used to produce a six-year projection of sickle cell disease cases in Nigeria and revealed that the number of sickle cell disease cases in the country (Nigeria) will decline during the following six years, from 2017 to 2022.

Keywords: Trend Analysis, Genetic, Sickle Cell, Disease, Disorder, Model Accuracy Measures

Introduction

Sickle Cell Disease(SCD) is a category of haemoglobin-related inherited diseases. Red blood cells are made up of haemoglobin, which is a molecule that delivers oxygen to the body's tissues. An issue with haemoglobin causes sickle cell disease, which causes red blood cells to be C-shaped (like a sickle) and sticky. As a result, they can stick in the cardiovascular system. They are also unable to deliver oxygen effectively. This can affect the body in various ways. In people with SCD, sickled red blood cells can clog small blood vessels and capillaries, preventing oxygen from reaching the tissues. As a result, many people with SCD have episodes of extreme discomfort. It can also cause organ damage, both acute and chronic. Children with SCD are at risk of contracting life-threatening infections. Since sickle cell disease impairs immune system function. Children with SCD are more likely to contract life-threatening infections, particularly those caused by Streptococcus pneumonia and Haemophilus influenza. (Gill, et al 1995). Sickle cell illness currently has no treatment. It can be fatal, there are however ways to manage the symptoms. Organizations such as the Sickle Cell Disease Association of America are attempting to raise awareness and encourage additional financing for sickle cell research (Cohen, et al., 2010).

Based on the National Institute of Health (2002) guidelines for the management of SCD, children with Hb SS or sickle S0-thalassemia should receive regular penicillin prophylaxis from the age of 2 months to 5 years, and parents should have the option to continue prophylaxis for patients older than 5 years; for patients with Hb SC, penicillin prophylaxis has been stated to be probably wise. In addition to antibiotic prophylaxis, children with SCD issues should be up to date on routinely recommended vaccines, such as pneumococcal and H. influenza type b vaccines (Hib). Children with SCD should get both the 13-valent pneumococcal-conjugated vaccine (PCV13) and the 23-valent polysaccharide vaccine starting at the age of 2–6 months. (NIH, 2002). Adamkiewicz, et al. (2008) established that people with SCD, as well as parents of children with SCD, need to know how to avoid complications and maintain their health. Dehydration prevention, avoiding exposure to excessive cold or heat, infection prevention, good nutrition, recognizing and seeking immediate medical attention for symptoms such as fever, and pain evaluation and management are just a few of them. (Wethers,

2000). As a result, information and awareness-related activities should be included in successful public health interventions for lowering morbidity and death among people with SCD.

Complication prevention knowledge and practice should be enhanced in areas such as follow-up of infants with favourable NBS results and provider contacts. Direct counselling, information material distribution, and reference to other sources are all possibilities for giving such information. Nearly all caregivers believed the written educational materials provided to children with SCD and their families were useful in a study (Mahat et al., 2007). Comprehensive SCD treatment clinics, for example, can provide disease-specific comprehensive care spanning a variety of medical skills and services (Grosse, et al., 2009). The number of specialized facilities, on the other hand, is restricted, and access to them is insufficient across the country. A medical home supervised by a primary care practitioner or a chronic care model may be more suited to provide primary and preventative care. (Grosse, et al., 2009). Day hospitals have also been reported to be an effective and helpful method of delivering health care, particularly for the treatment of sickle cell crisis and pain episodes (Raphael, et al, 2008).

Several variables contribute to the disease's return in Nigeria and Africa as a whole. In Africa, poverty has a role in the occurrence of sickle cell traits. Many of the world's poorest people, according to Pattanayak, et al. (2003), live in locations with a high incidence of sickle cell disease. Many people lack access to appropriate health care due to financial constraints. The economic status of a vulnerable country plays a role in defining preparedness and control measures in the event of an outbreak. (Sudhakar & Subramani, 2007). There appears to be limited information on the occurrence of SCD in Africa to enable community health experts and other health practitioners to give relevant counselling services to the general public regarding SCD. This study is designed to apply statistical modelling in forecasting the cases of SCD in Nigeria between 2016 and 2022.

As a statistical tool, time series analysis examines previous data and models it to produce forecasts for the future. Past observations of a phenomenon must be acquired to make any prognosis regarding the future. As a result, a time series can be defined as a collection of observations across time. The values Y_1, Y_2, \dots, Y_n of variable Y at times t_1, t_2, \dots, t_n can be written mathematically as; Secular trend, seasonal variation, cyclical variation, and irregular variation are all examples of motions in time series. As a result, time series analysis has shown to be a useful method for predicting macroeconomic variables. According to Chen and Hong (2012), econometric models are not superior to time-series techniques when fundamental changes occur in an economy. In this study, four trend models will be used to evaluate the movement in Nigeria; there have been epidemics of sickle cell disease. The linear trend model, quadratic trend model, exponential trend model, and S-curve model were used to determine the best model that matches the sickle cell disease outbreaks in Nigeria and, more importantly, to produce sickle cell disease forecast cases understudied.

Eke et al., (2015) compared three time-series models. This study fitted three time series trend models on Nigeria's Gross Domestic Product, viz: a linear trend model, a quadratic trend model, and an exponential trend model, using annual data from 1982 to 2012. The exponential trend model was found to have the lowest MAPE and best fit the data. The exponential trend model was used to construct a five-year forecast for Nigeria's Gross Domestic Product, revealing that the country's GDP will grow over the following five years. Lin et al., (2009) also used time series analysis to study the link between falciparum malaria in endemic provinces and imported malaria in non-endemic regions of China. An autoregressive integrated moving average model was used to fit the predictor variable at first. According to AIC and goodness-of-fit criteria, the seasonal ARIMA (1, 1, 1) and (0, 1, 1) models for malaria incidence matched the data the best of all the models examined. Briet, et al. (2008) created a short-term malaria prediction model for Sri Lanka. The ability of exponential moving average models, autoregressive integrated moving average models with seasonal components, and seasonal multiplicative autoregressive integrated moving average (ARIMA) models to predict the number of malaria cases one to four months ahead was compared using monthly time series of district malaria cases. The best model for forecasting and forecasting error differed significantly between districts. For example, in Ampara, the best model for predicting horizon was an ARIMA (2, 1, 1) with seasonality via a harmonic with one year and a harmonic with a six-month length for one month. The ARIMA (0, 1, 2) model with seasonality through a first-order seasonal autoregressive and a first-order seasonal moving average component proved optimal for the district of Ampara over longer forecasting horizons.

Contreras, *et al* (2003) used an ARIMA model to create a model for forecasting next-day power prices in mainland Spain and California markets. Their programme was able to predict the 24 market-clearing prices for the following day. The ARIMA model is a useful tool for forecasting time series. The time series analysis was used in the forecasting of malaria epidemic outbreaks in Nigeria by Nkpordee and Wonu (2018). The Nigerian National Bureau of Statistics, Social Statistics, provided secondary data for the study. The study used the Box-Jenkins (1976) approach to develop a suitable mathematical model by taking into account the ACF and PACF correlograms. With a 16-month lead, ARIMA (1, 0, 1) model was used to forecast monthly reported cases of malaria, resulting in a series with a progressive rise and decline. The study also looked at the monthly average of malaria cases recorded. The study suggested that the government should ensure that insecticide-treated nets, pesticides, anti-malaria drugs, and other products are available in rural communities in Nigeria.

Problem Specification

Poverty, unemployment, lack of education, lack of social support networks, and other disparities, for example, can all have an impact on the health and well-being of people. These factors may limit efforts to improve healthcare services. Despite the disease's severe impacts, some Nigerians do not fully comprehend the importance of a sickle cell disease-free environment in fostering economic development and poverty reduction. Sickle cell disease may have had a major impact that has not been quantified, which could persuade legislators, policymakers, programme managers, and development partners to commit the necessary attention and resources to combat this horrible disease. This study intends to fill the knowledge gaps about SCD trend behaviour in Nigeria to ascertain whether there is a decline or rise in the prevalence of the disease.

Purpose of the Study

The purpose of this study is to look at a trend analysis of genetic disorders in Nigeria, with a case study of sickle cell disease as a case study (SCD). The following are the study's particular objectives:

1. determine the parameter estimates (α_0 , α_1 and α_2) in the relevant models (i.e. linear, quadratic, exponential and s-curve trend models).
2. for each model, obtain the trend analysis plot for the sickle cell disease instances.
3. calculate the individual model's model accuracy measurements (MAPE, MAD, and MSD).
4. compare the results of the four trend models and choose the one with the lowest MAPE, MAD, and MSD as the best model for determining the sickle cell disease cases' trend.
5. using the best-fit trend model determined in step 4 above, forecast the trend of Nigerians with sickle cell illness

Materials and Methods

Research Design: The design for this study was a cross-sectional research design that focused on trend analysis of genetic disorder in Nigeria: A study of SCD.

Nature and Sources of Data: The collected and used data for this study was secondary statistical data extracted from survey information performed in Nigeria on sickle cell disease epidemic from records department of annual sickle cell disease cases in Nigeria matching the data length of 49 based on an aged group from 2010 to 2016. Long-term sickle cell disease cases in Nigeria data from the Nigeria SCDC Data - 2016, Average Number of Hospital Admissions were used. Appendix A contains the information.

Data Analysis: The parameters that make up the models are obtained using the MINITAB (16.0) application. The researcher used this software to estimate the parameters for Linear, Quadratic, Exponential, and S-cure models to make record evaluation easier. The MINITAB (version 16.0) programme was used to forecast the development of sickle cell illness in Nigeria after establishing the trend model with the greatest fit.

Model Specification: This research will use the linear trend model, quadratic trend model, exponential trend model, and S-cure model. Let Z_t be the time series, then

The Linear Trend Model

The linear trend model is used by default in trend analysis:

$$Z_t = \alpha_0 + \alpha_1 t + \mu_t \quad (1)$$

The average change from one period to the following time t is represented in this model.

Model for linear forecasting:

$$Z_t = \alpha_0 + \alpha_1 t \tag{2}$$

The Quadratic Trend Model:

$$Z_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \mu_t \tag{3}$$

Quadratic forecasting model:

$$Z_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 \tag{4}$$

The Exponential Trend Model:

$$Z_t = \alpha_0 (\alpha_1)^t + \mu_t \tag{5}$$

Exponential forecasting model:

$$Z_t = \alpha_0 (\alpha_1)^t \tag{6}$$

The S-curve Trend Model:

The S-curve model fits the Pearl-Reed logistic trend model. This is how the case when the series follows an S-shaped curve is described. The following is the model:

$$Z_t = \frac{10^a}{(\alpha_0 + \alpha_1 \alpha_2^t)} \tag{7}$$

where

α_0 = Estimated Z intercept

α_1 = Estimated linear effect on Z

α_2 = Estimated quadratic effect on Z

Models Accuracy Measures

To find the best fit, the model with the fewest accuracy measurements will be used (MAPE, MAD and MSD). The best model is one that reduces the criterion to the bare minimum. Several criteria for selecting various models, such as Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Squared Deviation (MSD), have been proposed in recent years (MSD). A couple of these standards are as follows:

Mean Absolute Percentage Error (MAPE): Because this is a percentage, it measures the correctness of fitted time series values and expresses accuracy as a percentage of the error.

$$MAPE = \frac{\sum |(Z_t - \hat{Z}_t) / Z_t|}{n} \times 100 \quad (Z_t \neq 0) \tag{8}$$

where \hat{Z}_t is the number of observations, Z_t is the number of fitted values, and is the actual values at time t.

Mean Absolute Deviation (MAD): The precision of time series values that have been fitted is measured with this metric. It expresses precision in the same way as data does. This also aids in theorizing the error.

$$MAD = \frac{\sum_{t=1}^n |(Z_t - \hat{Z}_t)|}{n} \tag{9}$$

where \hat{Z}_t , Z_t and n as defined in Equation (8)

Mean Squared Deviation (MSD): Regardless of the model, the denominator, n, is used to determine this. MSD values can now be compared across models. As a result, MSD is more vulnerable to forecast inaccuracy than MAD.

$$MSD = \frac{\sum_{t=1}^n |(Z_t - \hat{Z}_t)|^2}{n} \tag{10}$$

where \hat{Z}_t , Z_t and n as defined in Equation (8)

RESULTS

The Actual Data's Time Series Plot:

The sickle cell disease cases in Nigeria (2010–2016) were displayed in a time series plot in fig. 1 below. The graph depicts seasonal variation in the early period (seasonality of order 7) and upward trend movement. This implies we'll model the data by fitting a general time series trend model to figure out which model best fit the data set for predictions.

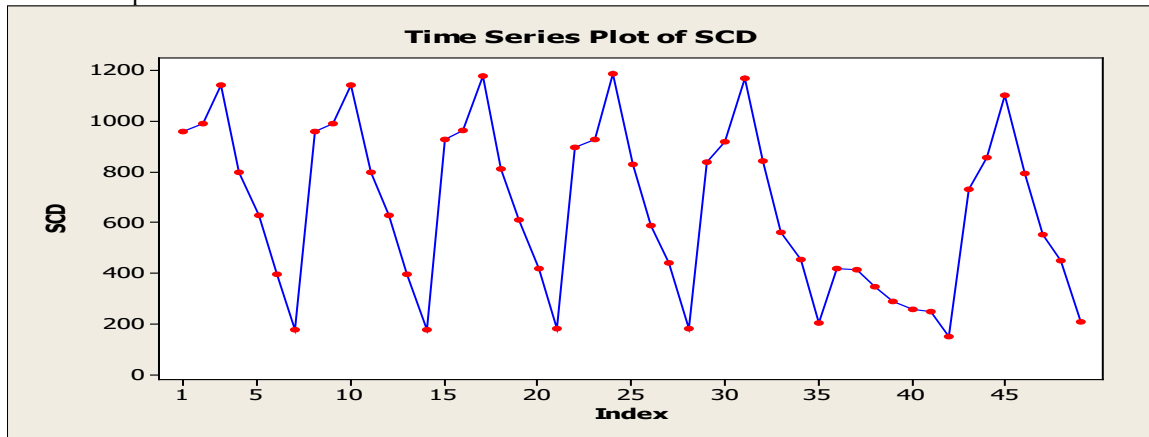


Fig.1 Time series plot of the actual data

The Linear Trend Model:

Equation (2)'s predicted linear trend model is written as $Z_t = 843.8 - 7.55510 * t$.

The linear trend model does not fit the data, as shown by the graphic in fig.2.

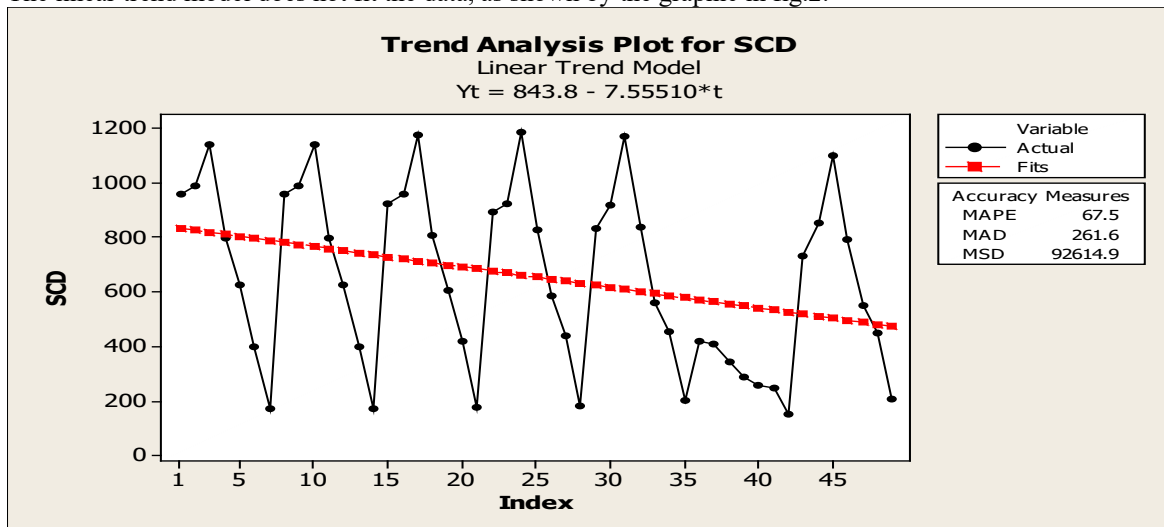


Fig.2 Trend analysis plot of the linear trend model

Table 1: Accuracy Measures of Linear Trend Model

Accuracy Measures	
MAPE	67.5
MAD	261.6
MSD	92614.9

Data in Table 1 shows that MAPE has a value of 67.5 in the linear trend model's accuracy metrics, while other values are relatively large.

The Quadratic Trend Model:

Equation (4)'s predicted quadratic trend model is written as $Z_t = 855 - 8.8t + 0.026t^2$. The quadratic trend model, as depicted in fig.3, does not suit the data.

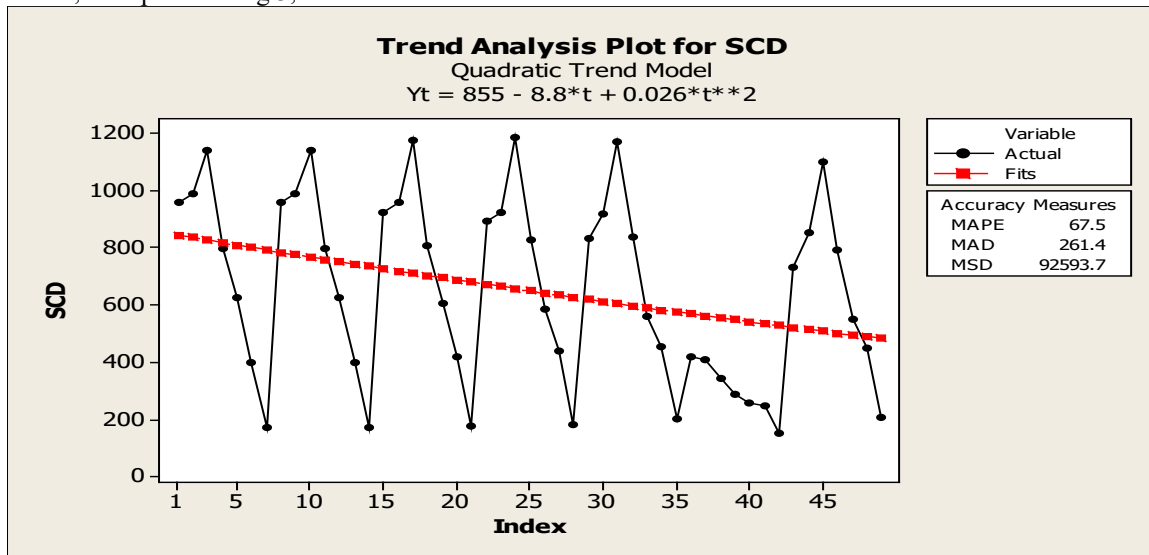


Fig.3 Trend analysis plot of the quadratic trend model

Table 2: Accuracy Measures of Quadratic Trend Model

Accuracy Measures	
MAPE	67.5
MAD	261.4
MSD	92593.7

Data in Table 2 shows that MAPE has a value of 67.5 in the quadratic trend model's accuracy metrics, while other measures are relatively large.

The Exponential Trend Model:

Equation (6)'s projected exponential trend model is written as $Z_t = 773.472 * (0.98698)^t$. The exponential trend model does not suit the data, as shown by the graphic in fig. 4.

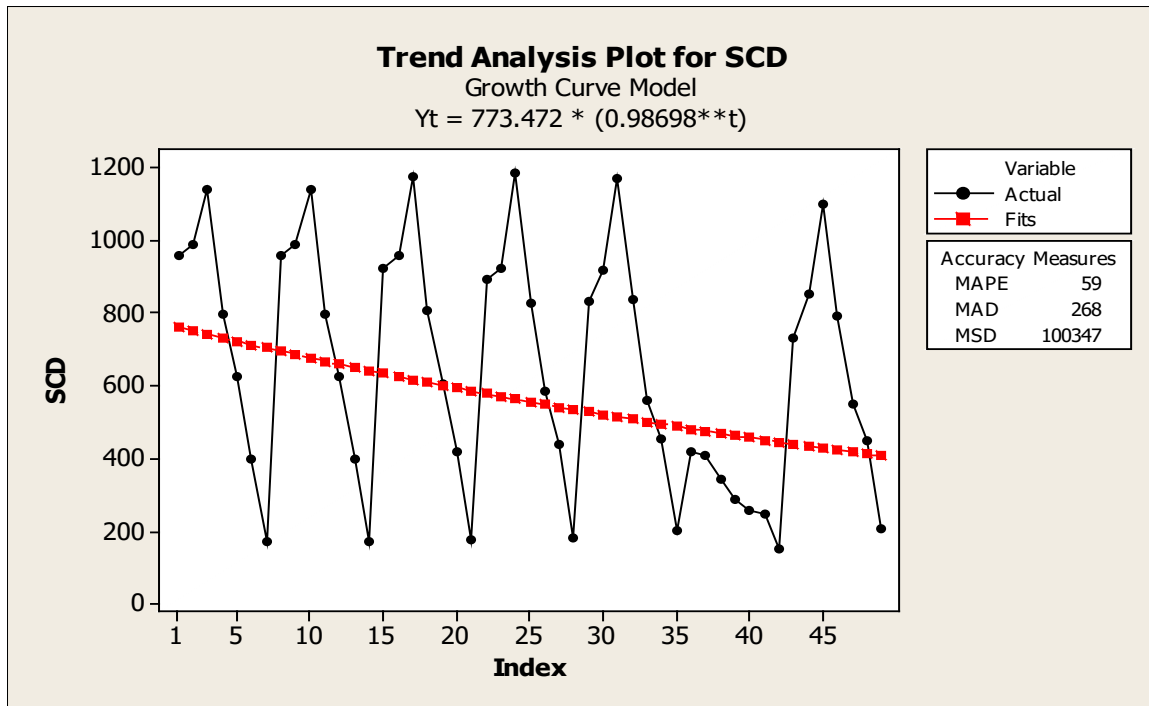


Fig.4 Trend analysis plot of the exponential trend model

Table 3: Accuracy Measures of Exponential Trend Model

Accuracy Measures	
MAPE	59
MAD	268
MSD	100347

The exponential trend model's accuracy measurements in Table 3 demonstrate that MAPE has a value of 59, while other values are relatively substantial.

The S-curve Trend Model:

The estimated s-curve trend model of equation (7) is given as $Z_t = \frac{10^4}{(18.7369 + 0.00193279 * (1.21864)^t)}$

. The s-curve trend model matches the data with slight variance, as seen in Figure 5.

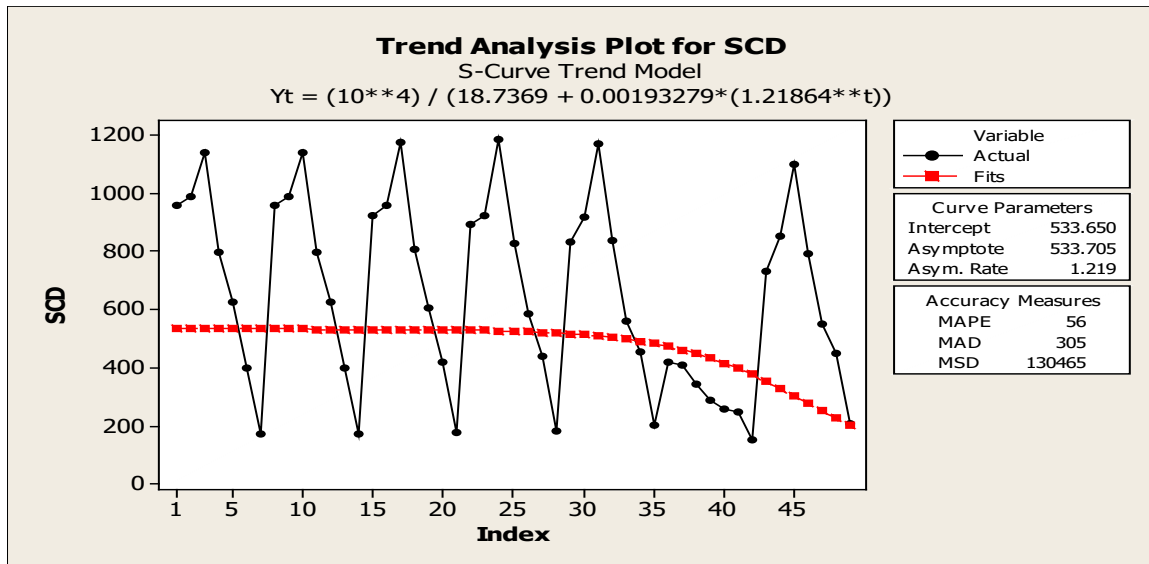


Fig.5 Trend analysis plot of the s-curve trend model

Table 4: S-curve Trend Model Accuracy Measures

Accuracy Measures	
MAPE	56
MAD	305
MSD	130465

Table 4 shows that MAPE has a value of 56 in the s-curve trend model's accuracy metrics, while other values are relatively substantial.

Comparison of the Accuracy Measures:

Table 5: Comparing the Accuracy Measures

Model	MAPE	MAD	MSD
Linear	67.5	261.6	92,614.9
Quadratic	67.5	261.4	92,593.7
Exponential	59.0	268.0	100,347.0
S-curve	56.0	305.0	130,465.0

When comparing the accuracy of the measurements of these four models in Table 5, the MAD and MSD for all of them are quite large. However, when comparing the four accuracy metrics, the s-curve model has the least MAPE measure of 56, indicating that sickle cell disease cases in Nigeria do not follow a linear trend model; rather, they follow an s-curve trend model. The s-curve trend model will be used to calculate a six-year projection of sickle cell disease cases.

Six Years Forecast of the Sickle Cell Disease Cases in Nigeria

The s-curve trend model forecast of sickle cell disease cases in Nigeria shows that the number of sickle cell disease cases will decline, as shown in Table 6 below.

Period	2017	2018	2019	2020	2021	2022
Forecast	176.212	153.702	132.998	114.244	97.491	82.711

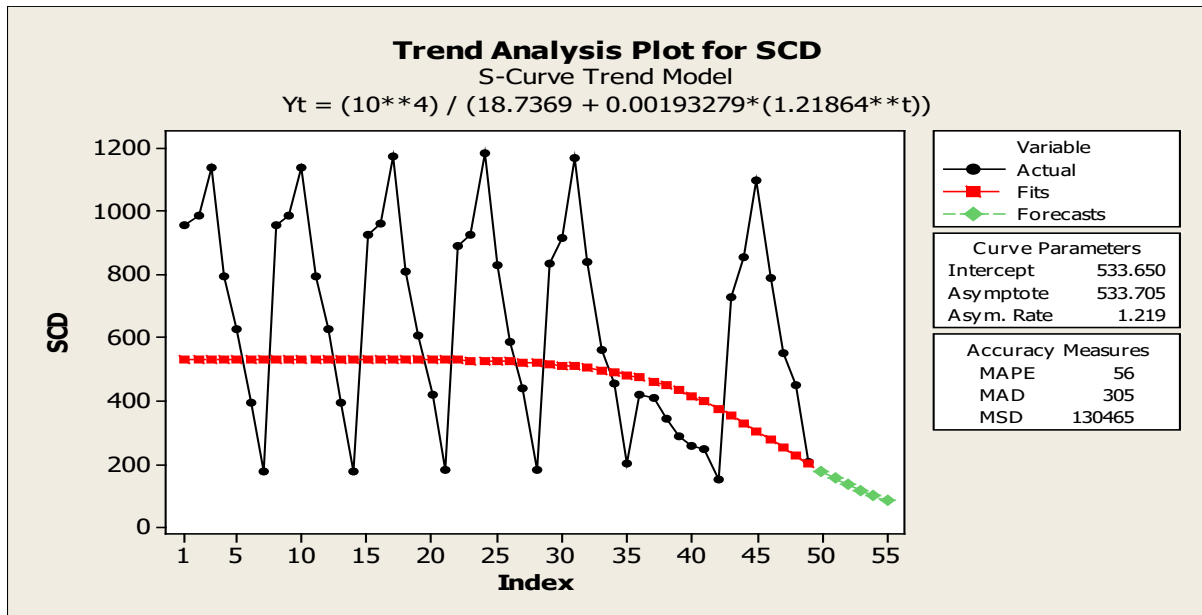


Fig.6 Trend analysis plot of the s-curve trend model with Forecasted values

Discussion of Finding

The growth behaviour of sickle cell disease cases in Nigeria was analyzed using four time-series models, which were compared in this study. The four models encompassing Linear, Quadratic, Exponential, and S-curve were not all supposed to be able to reflect the long term trend of sickle cell disease cases in Nigeria by comparing and analyzing the model accuracy metrics of MAPE, MAD, and MSD values of each model. The S-curve trend model, on the other hand, was in the best agreement with the actual long-term sickle cell disease cases in Nigeria, as measured by MAPE, MAD, and MSD values. The real sickle cell disease patients were detected using an s-curve trend model, which was established by the MAPE value (Mean Absolute Percentage Error). In a previous study, Nkpordee and Wonu (2018) applied time series analysis was used in the forecasting of malaria epidemic outbreaks in Nigeria. The Box-Jenkins (1976) approach was used to develop a suitable mathematical model by taking into account the ACF and PACF correlograms. The ARIMA (1, 0, 1) model was used to forecast monthly reported cases of malaria, resulting in a series with a progressive rise and decline. The study also looked at the monthly average of malaria cases recorded.

CONCLUSION

It was established that of all three model accuracy methodologies tested, the quadratic and s-curve trend models compete favourably. The s-curve trend model, on the other hand, had the lowest Mean Absolute Percentage Error (MAPE) and best fit the data (2010-2016). The s-curve trend model was used to produce a six-year projection of sickle cell disease cases in Nigeria and revealed that the number of sickle cell disease cases in the country (Nigeria) will decline during the following six years, from 2017 to 2022. The implication of the findings of this study is positive because the decline in the number of cases of SCD is an indication of a reduction in the problems associated with SCD. The pregnancy and childbirth psychological effects, lower educational achievement and difficulty in finding a job as well as other vital considerations related to genetic counselling can be optimally done with public health efforts with the available findings. These are significant factors that should be included in public health initiatives aimed at preventing SCD.

Recommendations

The following recommendations were made based on the findings of the study:

1. A national public health agenda with the requisite public health services as its base is required in addressing the issues and concerns raised in this study to increase public awareness of SCD.
2. Poverty, unemployment, lack of education, insufficient social support networks, and related disparities should be given serious consideration as variables to improve the health and well-being of people.

3. Important health-related programmes and policies that are focused on or coordinated with socioeconomic and environmental aspects should be more effective.

References

- Adamkiewicz, T. V., Silk, B. J., Howgate, J., Baughman, W., Strayhorn, G., Sullivan, K., & Farley, M. M. (2008). Effectiveness of the 7-valent pneumococcal conjugate vaccine in children with sickle cell disease in the first decade of life. *Pediatrics*, *121*(3), 562-569.
- Box, G.E., & Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*. Rev. ed. Holden-Day.
- Briët, O. J., Vounatsou, P., & Amerasinghe, P. H. (2008). Malaria seasonality and rainfall seasonality in Sri Lanka are correlated in space. *Geospatial Health*, *2*(2), 183-190.
- Chen, B., & Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, *80*(3), 1157-1183.
- Cohen, R. T., DeBaun, M. R., Blinder, M. A., Strunk, R. C., & Field, J. J. (2010). Smoking is associated with an increased risk of acute chest syndrome and pain among adults with sickle cell disease. *Blood, The Journal of the American Society of Hematology*, *115*(18), 3852-3854.
- Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE transactions on power systems*, *18*(3), 1014-1020.
- Eke, C. N., Egwim, K. C. Igbo, C. A. & Onuoha, D. O. (2015). Comparative analysis of three time series trend models on gross domestic product in Nigeria. *European Journal of Business and Management*, *7*(3), 11-15
- Gill, F.M., Sleeper, L.A., Weiner, S.J., Brown, A.K., Bellevue, R., Grover, R., Pegelow, C.H. and Vichinsky, E., (1995). Clinical events in the first decade in a cohort of infants with sickle cell disease. Cooperative Study of Sickle Cell Disease. *Blood* 1995;86: 776-83.
- Grosse, S. D., Schechter, M. S., Kulkarni, R., Lloyd-Puryear, M. A., Strickland, B., & Trevathan, E. (2009). Models of comprehensive multidisciplinary care for individuals in the United States with genetic disorders. *Pediatrics*, *123*(1), 407-412.
- Lin, H., Lu, L., Tian, L., Zhou, S., Wu, H., Bi, Y., Ho, S.C. and Liu, Q., 2009. Spatial and temporal distribution of falciparum malaria in China. *Malaria Journal*, *8*(1), 1-9. <https://doi.org/10.1186/1475-2875-8-130>
- Mahat, G., Scoloveno, M. A., & Donnelly, C. B. (2007). Written educational materials for families of chronically ill children. *Journal of the American Academy of Nurse Practitioners*, *19*(9), 471-476.
- National Institutes of Health (2002). *The management of sickle cell disease*. 4th ed. NIH, MD2002.
- Nkpordee, L. & Wonu, N.(2018). Application of time series analysis on the forecasting of the outbreak of malaria epidemic in Nigeria. *International Journal of Statistics and Applied Mathematics* 2018; *3*(6): 98-105.
- Pattanayak, S. K., Mercer, D. E., Sills, E., & Yang, J. C. (2003). Taking stock of agroforestry adoption studies. *Agroforestry systems*, *57*(3), 173-186.
- Raphael, J. L., Kamdar, A., Wang, T., Liu, H., Mahoney, D. H., & Mueller, B. U. (2008). Day hospital versus inpatient management of uncomplicated vaso-occlusive crises in children with sickle cell disease. *Pediatric blood & cancer*, *51*(3), 398-401.
- Sudhakar, P., & Subramani, P. (2007). Literature review: insights into formulating a protective malarial medicine. *Journal of Young Investigators*. 16(1)
- Wethers, D. L. (2000). Sickle cell disease in childhood: Part I. Laboratory diagnosis, pathophysiology and health maintenance. *American Family Physician*, *62*(5), 1013-1020.