TIME SERIES ANALYSIS OF MALARIA IN ABA, ABIA STATE: USING AUTO REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS TO FORECAST FUTURE INCIDENCE

Igwe, N. O[1], Okonta, C. A[2], Okeahialam, A. H[3],
Bassey Uwabunkonye[4] , Igwe, Favour O.[5]

[1,2,3,4.]Department of Maths/ Statistics,
Akanu Ibiam Federal Polytechnic, Unwana, Ebonyi State Nigeria.
[5]Department of Biochemistry,
Abia State University Uturu, Nigeria.
ndyigwe@gmail.com, charlesokonta1973@gmail.com,ohossanna@yahoo.com,
okorieiwo@gmail.com, okoriefavour16@gmail.com
08064090157[1],08063463468[2], 08030893317[3],08037381137[4], 08037299216[5]

Abstract

The study sought to assess the trends of Malaria incidence in Aba Metropolis and forecast future incidence. A retrospective comparative study design was employed using secondary data from Abia State University Teaching Hospital Aba from January 2014 to December 2020. Trend of malaria prevalence was analysed and compared by years and months. The Quadratic model was used for the forecasting of the half year incidence of Malaria while Auto regressive integrated moving average (ARIMA) (1, 1, 2) was used for forecasting monthly malaria incidence for the years 2021 and 2022 in Aba Metropolis. For the general pattern, July recorded the highest number of cases whereas January recorded the lowest cases in each year. Also, 2014 was the best performing year since it recorded the lowest number of malaria cases (10,336). The projected malaria case for the first half year of 2022 is expected to be 61,371.8, while the second half year is expected to be 77,842.0. This model is recommended to the metropolitan health directorate and researchers who want to monitor the malaria reported cases in the metropolis and other parts of the world. It is suggested that measures should be put in place to curb malaria incidence during the period of the year when high incidence were recorded.

Keywords

Malaria; ARIMA; Forecasting; Future Incidence

Introduction

This paper deals with the peculiarity that human health is extremely important in every society. A society filled with unhealthy people cannot excel both in society development and otherwise. Malaria is caused by the *Plasmodium* genus that is transmitted between humans by *Anopheles* mosquitoes (Reindolf, 2018). *P. falciparum* and *P. vivax* are the most common species that cause malaria in humans. *P. falciparum* is the most dangerous because of the multi-drug resistance on this strain of the disease (Medical Research Council, 2001). A severe episode of cerebral malaria can result in epilepsy, cerebral palsy, or intellectual or physical disabilities (Davies & Eaton, 2018). These malaria victims are in the poorest, and sometimes most remote parts of the world, increasing the difficulty in finding support to cope with the disease (Davies & Eaton, 2018).

According to the World Health Organization, in 2012, there were approximately 207 million cases of malaria resulting in 627,000 deaths (World Health Organization, 2014). The overwhelming majority that is 90% of these cases occur in Africa (Medical Research Council, 2001). Most of the deaths occur in children. However, the rate of deaths in children has been reduced by 54% since 2000 (World Health Organization, 2014). The Countries with the most confirmed cases are in sub-Saharan Africa and India. Moreover, malaria contributed to 2.05% to the total global death in 2000 and was responsible for 9% of all death in Africa (WHO, 2003). WHO also estimated that the total cost of malaria in Africa was US$ 1.08 billion in 1995 and US$2 billion in 1997 (WHO, 1997). Malaria is therefore a massive problem which plagues all segments of the society. Malaria remains a major health challenge to mankind all over the world (World Health Organization, 2013). This is tied to the report that over three billion people in the world stand the risk of having malaria (World Health Organization, 2013). Despite local and international efforts towards the prevention of the disease, the rate at which people become sick and eventually die as a result of malaria is outrageous (Adebayo, *et al.*, 2015).

The future values of these variables are often predicted from their history. A time series analysis often does this and therefore this study was conducted to identify the trends of malaria cases in Aba Metropolis over the period 2014 to 2020 using time series analysis and to forecast future incidence for 2021 and 2022. A good way of describing the incidence of malaria is important and it will go a long way to ensure proper planning and evaluation in the implementation of programs to monitor and control the disease, especially in endemic zones.

Materials and Methods

A retrospective comparative study design was employed using Malaria cases as reported by Abia State University Teaching Hospital Aba from January 2014 to December 2020. The data included monthly number of malaria cases among the various age groups. Trend of malaria incidence was analysed and compared by years and months.

Auto Regressive Integrated Moving Average Model

An Auto Regressive Integrated Moving Average (ARIMA) model was used and applied to Time Series Data of Malaria incidence in Kumasi. The model looks for temporal dependence between successive observations (Helfenstein, 1991). Due to the transmissibility and seasonality of malaria, models with an ARIMA structure have more predictive power compared to other methods (Nobre, *et al.*, 2001); such models have been applied to predict numerous infectious and noninfectious diseases with similar periodic patterns over the past decades (Ture & Kurt, 2006). Another advantage of the ARIMA approach is the relative simplicity and stability of the model in predicting malaria cases in a context where poor resources have led to lack of detailed data, which makes it difficult to calculate parameters needed for construction of more complex models of malaria (Pascual, *et al*., 2008). ARIMA models are in theory the best models for forecasting a time series. The procedure involves fitting an appropriate model, estimating the parameters and verifying the model. This model was used to forecast malaria incidence for the year 2022.

Results and Discussions

The monthly reported cases from January 2014 to December 2020 were fed into Minitab Spread sheet to generate the best trend fit, which was given by the quadratic model, $Y_t = 3665.84 - 2.1049t + 0.5051t^2$. The seasonal indices (S.I) were obtained by dividing the actual values at time t by its respective trend forecast. The quadratic trend model for the half year was given as $Y_t = 13341 - 215t + 114t^2$ In order to choose the best model to forecast for the year 2018, an output of measure of accuracy to determine whether the use of ARIMA (1, 1, 1), ARIMA (1, 1, 2), ARIMA (0, 1, 1) or ARIMA (0, 1, 2) was appropriate for the monthly forecast of malaria in Aba for the years 2022 was performed and the results indicated that the ARIMA (1, 1, 2) was best fit for forecasting monthly malaria incidence for the year 2022. This was as a result of low values for the MAPE, MAD and MSE as shown in Tables 1 and 2 below.

Table 1. Tentative ARIMA models and their corresponding normalized BIC and MAPE

Models	Normalized BIC	MAPE	MSE
ARIMA (1, 1, 1)	7.820	13.539	801,626
ARIMA (1, 1, 2)	7.585	13.070	734,201
ARIMA (0, 1, 1)	7.854	13.624	862,245
ARIMA (0, 1, 2)	7.918	13.686	894,013

Table 2. Measure of model accuracy

Models	MAPE	MAD	MSE
ARIMA (1, 1, 2)	13	851	734201
Moving Average(MA)	15	2005	796442
Single Exponential Method	13	961	512204
Double Exponential Method	13	979	581861
Linear Trend Model	23	1450	1094211
Quadratic Trend Model	21	1474	611052

Therefore, the model selected for forecasting monthly incidence of Malaria in 2022 was ARIMA (1,1, 2) because this model had the minimum normalized Bayesian Information Criterian (BIC) of 7.585. The ARIMA (1, 1, 2) model also gives the best fit for forecasting monthly malaria incidence since it has the least MAD (851) and MSE(734201) among all the models.

Considering the results, ARIMA (1, 1, 2) appeared fit for forecasting monthly malaria incidence for the year 2022 in Aba metropolis. Hence the model for predicting future malaria incidence is $Y_t = 3665.84 - 2.1049t + 0.5051t^2$ and $Y_t = 13341 – 215t + 114t^2$.

$$Z_t = X_t - X_{t-1}$$

and the model obtained was;

ARIMA(1,1,2): $X_t = \mu + \alpha_1 X_{t-1} + \beta_1 Z_{t-1} + \beta_1 Z_{t-2} + Z_t$

where, μ is a constant, α, β is a parameter and is the $Z_t$residual term.

Substituting the estimates of the parameters, we obtained

$X_t = 7.5131 + 0.9126X_{t-1} + 1.1312Z_{t-1} - 0.2196Z_{t-2} + Z_t$

Table 3. Parameter estimates of the model

Type	Coefficient	SE Coefficient	T	P
AR(1)	**0.9126**	**0.0114**	**6.75**	**0.000**
MA(1)	**1.1312**	**0.0053**	**29.14**	**0.000**
MA(2)	**-0.2196**	**0.0491**	**-3.19**	**0.031**
Constant	**7.5131**	**1.2991**	**11.14**	**0.000**

The final estimate of parameters for the model (Table 3) shows that the AR (1), MA (1) and MA (2) parameters have *p*-values of which indicate a significant model parameter. The model converges after 25th iterations. The test statistic value of the constant is 11.14 and *p*-value is 0.000, indicating that the constant is significant at 5% level of significance. Thus the independent assumption is not violated. Once the assumptions hold the model can be seen as valid for prediction.

Quadratic model:

Also, in order to choose the best model for forecasting for the half year malaria incidence for 2021 and 2022, an output of measure of accuracy showing whether the use of Linear or Quadratic trend analysis was performed and the results have been summarized in Table 4 below.

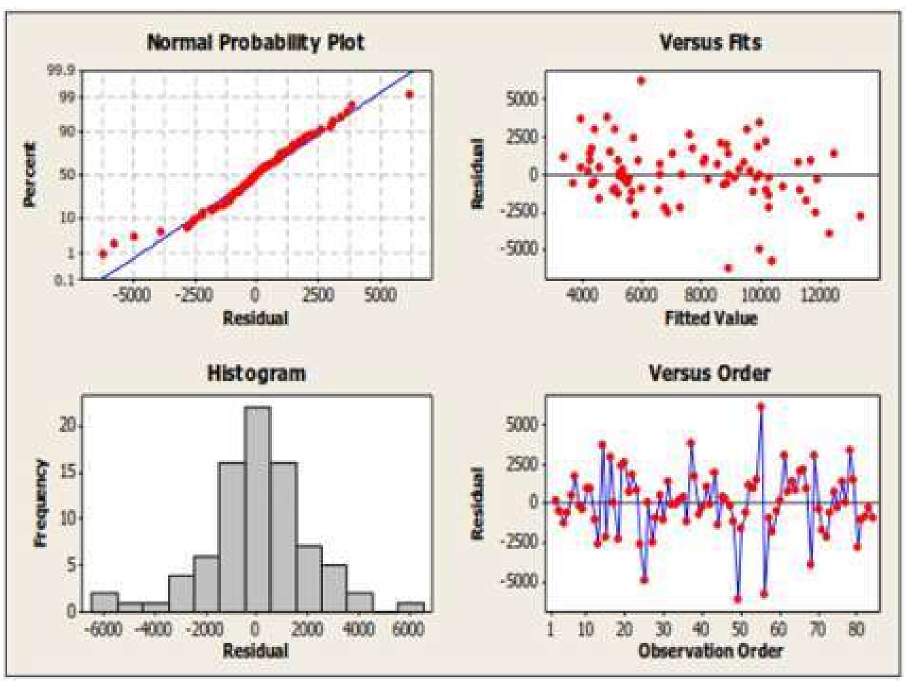
Table 4. Measure of model accuracy

Models	MAPE	MAD	MSE
Linear trend analysis	**14**	**6414**	**9844389**
Quadratic trend analysis	**13**	**5937**	**6110410**
Exponential trend	**14**	**6955**	**9847139**

The Quadratic model appeared to be fit for the forecasting of half year malaria incidence. This was as a result of low values for the MAPE, MAD and MSE. The quadratic trend model used for the forecast of half year malaria incidence was $Y_t = 13341 - 215t + 114t^2$

The selection of the appropriate model also depended on the values of Normalized BIC and the ACF together with the PACF. The graphs of the ACF and PACF are shown in Figures 1 and 2 respectively. Four tentative models were examined and the model with the minimum Normalized BIC was chosen.

Figure 1



From the normal plot of residual, it was observed that the residuals did not deviate much from the straight line. This indicated that the errors were quite close to normal with no clear outliers. Thus, the normality assumption holds. The histogram of residuals confirmed this assumption. The plot of residuals vs. the fitted values exhibited no trend in dispersion. This indicated that the model satisfies the constant variance assumption. The plot of residuals vs. the order of the data suggested that the residuals were uncorrelated. This fact is supported by the Ljung Box statistics (Table 5) which gives non-significant *p*-values.

Table 5.Modified box–pierce (Ljung-Box) χ^2 Statistic

Lag	χ^2	DF	P- value
12	**7.9**	**8**	**0.455**
24	**30.5**	**20**	**0.063**
36	**43.8**	**32**	**0.080**
48	**61.7**	**44**	**0.127**

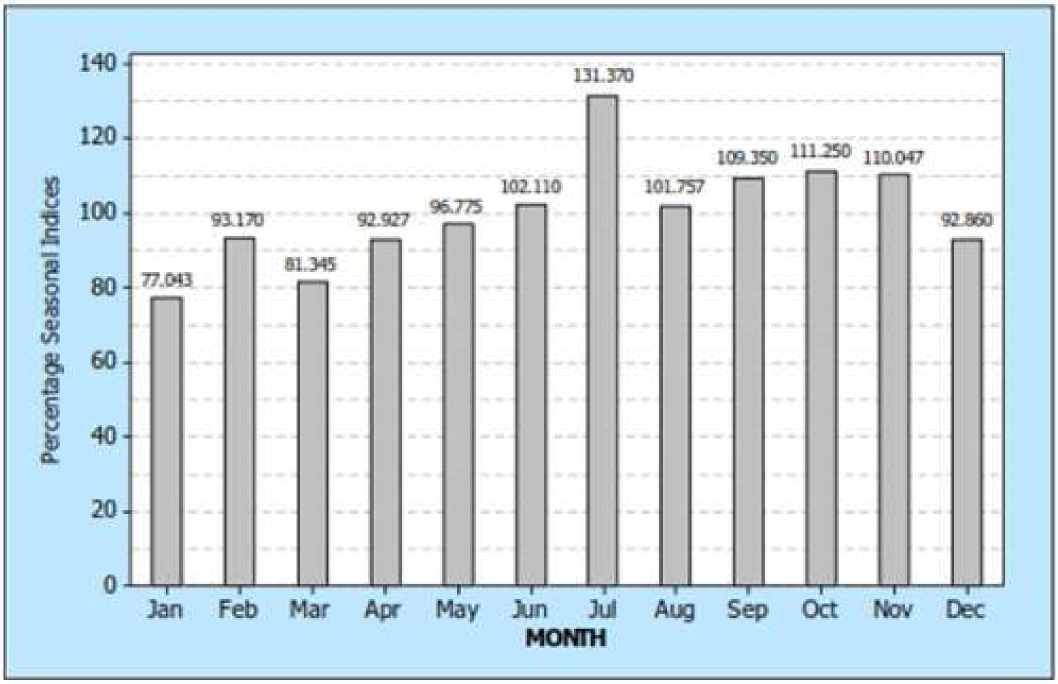
Table 5 shows the modified Box-Pierce (Ljung-Box) $\chi 2$ statistic. It can be seen that all the lags have a *p*-value greater than the level of significant (0.05). This indicates non-significance implying that this model was appropriate. The statistical software used for the analysis was Minitab Version 16, SPSS Version 16. The data for the analysis were grouped as Monthly malaria cases and Half year malaria cases. The monthly reported cases from January 2014 to December 2020 were fed into a Minitab spread sheet and analyzed to obtain the descriptive statistics. That was followed by graphical and tabular exploration of yearly, monthly, and half year incidence of malaria and forecast of future incidence.

Table 6: Descriptive statistics of monthly malaria cases

Year	Number of months	Mean	Standard deviation	Minimum	Maximum	Sum
2014	12	15,185	910	10,621	27,720	262,223
2015	12	18,104	2,807	13,101	22,072	297,242
2016	12	15,050	757	13,918	26,619	260,597
2017	12	19,100	801	17,928	20,769	290,197
2018	12	14,866	2,517	12,686	22,128	258,386
2019	12	19,822	2,013	17,284	22,618	217,861
2020	12	20,294	1,693	18,207	23,870	223,532

Table 6 displays the Malaria Cases from 2014 to 2020 and from the table, the lowest average annual Malaria cases were recorded in 2018 with 14,866 cases while the highest were recorded in 2020 with 20,294 cases. However, the highest Malaria cases for the period were recorded in 2015 with 297,242 cases while the lowest were recorded in 2019 with 217,861 cases. The highest standard deviation is seen in the year 2015, which implies that, the reported cases in that year had the highest monthly variations whereas the lowest standard deviation was recorded in 2016 implying that the monthly reported cases did not vary significantly.

Figure 2: Malaria incidence in months

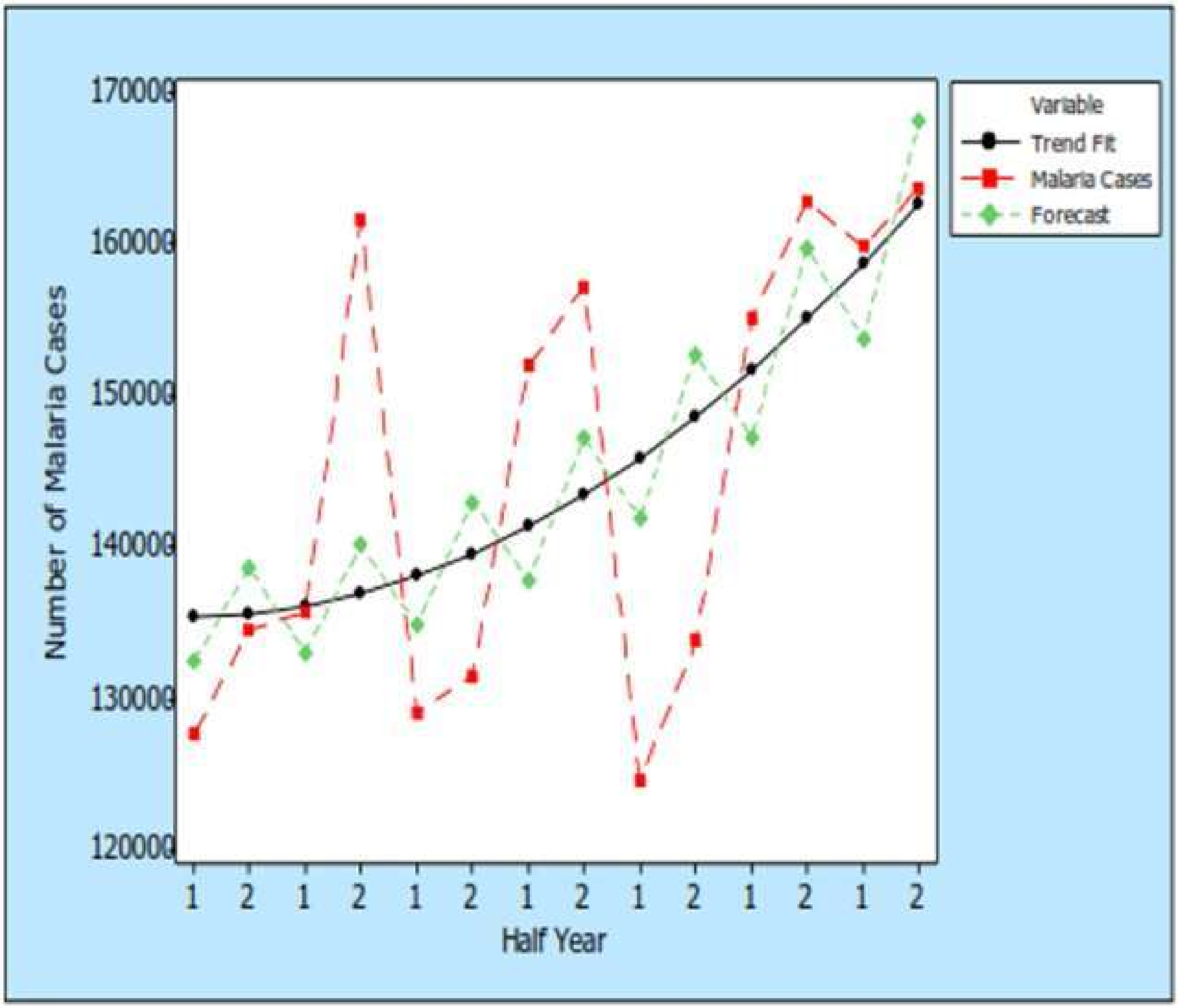


The monthly analysis of malaria cases consisted of 84 months. That is, 12 month each for the 7 year period. The seasonal indices are obtained and shown in Figure 2. The seasonal indexes plot indicates that, malaria cases in June, July, August, September, October and November are above the average mark of 100%, Whiles those of January, February, March, April May, and December are below the average. The averages were adjusted so that those below 100% were below the monthly average and those above 100% were above the monthly average. Hence, from Figure 2, malaria incidence for July was 31.4% (131.37–100%) above the rate of incidence. January incidence on the other hand was 22.96% (100–77.043%) below the average. The same calculation can be done for all the other months. Thus July had the highest incidence of malaria whilst January had the lowest. In general, the second half of the year records higher incidence of malaria than the first half.

Table 7: Time series analysis showing half year malaria incidence

Year	Half year	Time (t)	Malaria cases (yt)	Forecast ($\hat{y}t$)	*S.It*	Adjusted (S.I)
2014	1	1	127,628	135,445	0.77946	0.91541
	2	2	134,595	135,635	0.97082	1.08589
2015	1	3	135,692	136,143	0.98752	0.91541
	2	4	161,550	136,969	1.66491	1.08589
2016	1	5	129,088	138,113	0.76320	0.91541
	2	6	131,509	139,575	0.79618	1.08589
2017	1	7	151,961	141,355	1.25646	0.91541
	2	8	157,236	143,453	1.31719	1.08589
2018	1	9	124,579	145,869	0.53585	0.91541
	2	10	133,807	148,603	0.69557	1.08589
2019	1	11	155,069	151,655	1.06609	0.91541
	2	12	162,792	155,025	1.14115	1.08589
2020	1	13	159,845	158,713	1.01928	0.91541
	2	14	163,687	162,719	1.01543	1.08589

Figure 3: Time series plot showing half year malaria incidence



The time series plot of the half year reported cases from January 2014 to December 2020 shows that the time series was non-stationary. The time series exhibit an increasing trend and seasonal variations implying that there has been an increment and seasonal fluctuation in malaria incidence for the half year.

The plot also showed that, there has been a consistent rise in malaria cases for the second half year over the first half year. The half year analysis of malaria cases consisted of fourteen half years. That is, two half years for each of the years from 2014 to 2020. The seasonal indices were obtained and shown in Figure 3. The seasonal indices plot indicates that, malaria cases for first half year were below the average mark of 1 by 8.46% (1–0.91541), Whiles the second half year is above the average mark of 1 by 8.56% (1.08589–1). These indices suggest that Malaria cases fall below the average number in the first half of the year, whilst it rises above the average number in the second half of the year.

Table 8: Monthly forecast for the year 2021

Period (2021)	Forecast	Lower limit	Upper limit
January	8,924.00	5,174.12	12,673.9
February	8,764.79	4,500.01	13,029.6
March	8,671.77	4,218.86	13,124.7
April	8,624.55	4,107.08	13,142.0
May	8,609.03	4,073.77	13,144.3
June	8,615.44	4,077.62	13,153.3
July	8,637.04	4,099.18	13,174.9
August	8,669.15	4,129.16	13,209.1
September	8,708.53	4,163.13	13,253.9
October	8,752.94	4,199.01	13,306.9
November	8,800.84	4,235.81	13,365.9
December	8,851.15	4,273.06	13,429.2

Table 8 gives the expected monthly forecast of malaria incidence using the model ARIMA (1, 1, 2). The forecast shows consistent but gradual increase in incidence from June to December, but a decrease from January to May.

Table 9: Trend analysis for the half year malaria incidence for 2021 and 2022

Year	Period	Forecast values
2021	First half year	61,371.8
	Second half year	77,842.0
2022	First half year	70,161.6
	Second half year	88,959.4

A forecast for each half year for the next two years using the ARIMA (1, 1, 2) model is shown inTable 9. The projected malaria cases for the first half of 2021 are expected to be 61,371.8 cases, while the second half of the year is expected to record 77,842.0 cases of Malaria. In the first half of 2022, it is expected that 70,161.6 cases of malaria will be recorded which will rise to 88,959.4 cases in the second half of the year. This suggests a general reduction of malaria cases from previous years.

Time series analysis was adopted for this study to identify the trends in malaria cases in the Aba Metropolis over the period 2010 to 2016 and to forecast future incidence for 2020 and 2021. Time series methods were used for analysing the series data in order to extract meaningful statistics. Preliminary and further analysis conducted was based on Half Year and Monthly Malaria cases. In all these cases, malaria cases have shown a continuously increasing trend. The seasonal indices analysis for malaria cases for first half year was below the average mark of 1 by 8.46% whiles the second half year was above the average mark of 1 by 8.56%. This implies that Malaria cases consistently rise in the second half of each year.

The time series model developed for predicting the number of monthly cases of malaria was ARIMA (1, 1, 2) while the Quadratic model was used for the forecasting of the half year incidence of Malaria. This implies that ARIMA (1, 1, 2) can be used as a forecasting model to project the future values of a series based entirely on its own inertia. ARIMA works best when data exhibits a stable or consistent pattern over time with a minimum amount of outliers (Labys, 2006).

Conclusion

Based on the findings of the study, we make a number of conclusions and recommendations.The general trend of both monthly and half year malaria cases follows an increasing quadratic trend and there is seasonality in both cases. For the general pattern, July recorded the highest number of cases whereas January recorded the lowest cases in each year and the second half of each year records higher number of malaria cases. It was projected that Malaria cases for 2020 and 2021 will reduce when compared to previous years and therefore a projection for the first half of 2021 was 61,371.8 cases, while the second half cases will rise to 77,842.0. For the first half of 2019, it was projected that 70,161.6 cases of malaria will be recorded which is expected to rise to 88,959.4 cases in the second half of the year.

The ARIMA (1, 1, 2) model was used for forecasting the number of expected monthly cases of Malaria. The model was used to predict a twelve-month lead period of malaria cases for 2021. The Quadratic model was also used for the forecasting of expected half year incidence of Malaria cases for 2018 and 2019. This model is recommended to the metropolitan health directorate and researchers who would want to monitor the malaria reported cases in the metropolis and other parts of the world.

Recommendations

The ARIMA (1, 1, 2) model is recommended to the metropolitan health directorate and researchers who want to monitor the malaria reported cases in the metropolis and other parts of the world. It is suggested that measures should be put in place to curb malaria incidence during the period of the year when high incidence were recorded. It is recommended that in some instances, indoor residual spraying should be applied to minimize environmental contamination. Indoor residual spraying is suggested for application prior to the rainy season to prevent and control epidemic outbreaks. The use of Insecticide Treated Nets is also recommended especially in the rainy seasons when malaria cases increase.

References

Adebayo, A. M., Akinyemi, O. O., & Cadmus, E. O. (2015). Knowledge of Malaria Prevention among women and female caregivers of under-five children in rural southwest Nigeria. *Peer Journal, 3*, e792. https://doi.org/10.7717/peerj.79*Journal of Health Geographics, 5*(1), 60. https://doi.org/10.1186/1476-072X-5-60.

Davies, M., & Eaton, J. (2018). Malaria: Finding a preventive strategy that African countries can afford.*International Journal of Biometeorology, 46*(2), 81–89. https://doi.org/10.1007/00484-001-0119-6.

Helfenstein, U. (1991). The use of transfer function models, intervention analysis and related time series methods in epidemiology. *International Journal of Epidemiology, 20*, 808–815. https://doi.org/10.1093/ije/20.3.808.

Medical Research Council. (2001). *Malaria advice for southern Mozambique, Swaziland, and South Africa*. Retrieved from http://www.malaria.org.za/Malaria_Risk/riskadv/ General_ Malaria_Info. pdf.

Nobre, F. F., Monteiro, A. B. S., Telles, P. R., & Williamson, G. D. (2001). Dynamic linear model and SARIMA: A comparison of their forecasting performance in epidemiology. *Statistics in Medicine, 20*(20), 3051–3069. https://doi.org/10.1002/(ISSN)1097-0258.

Pascual, M., Cazelles, B., Bouma, M. J., Chaves, L. F., & Koelle, K. (2008). Shifting patterns: Malaria dynamics and rainfall variability in an African highland. *Proceedings of the RoyalSociety B: Biological Sciences, 275*(1631), 123–132. https://doi.org/10.1098/rspb.2007. 1068.

Reindolf, A. (2018). Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997–2003. *Malaria Journal, 6*(1), 129.https://doi.org/10.1186/1475-2875-6-129.

Ture, M., & Kurt, I. (2006). Comparison of four different time series methods to forecast hepatitis A virus infection. *Expert Systems with Applications, 31*(1), 41–46. https://doi.org/10.1016/j. eswa.2005.9.002.

World Health Organization. (1997). Application of the international classification of diseases to neurology: *ICDNA*.World Health Organization.

World Health Organization. (2003). *WHO technical report series No. 892*. Geneva: WHO Expert Committee on Malaria.

World Health Organization. (2013). *Fact sheet on the world malaria report 2013*. Retrieved from http://www.who.int/malaria/media/world_malaria_report_2013/en/.

World Health Organization. (2014). *Malaria* [Internet]. Retrieved from http://www.who.int/mediacentre/ factsheets/fs094/en/.