RESEARCH ARTICLE

Modeling the trend of reported malaria cases in Kisumu county, Kenya [version 1; peer review: 1 approved with reservations, 1 not approved]

Elsen Achieng ¹, Vincent Otieno ², Joseph Mung'atu³

¹Institute of Tropical and Infectious Diseases, University of Nairobi, Nairobi, Nairobi, 00200, Kenya

²Department of Public Health, University of Nairobi, Nairobi, Nairobi, 00200, Kenya

³Department of Biostatistics, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Nairobi, 00200, Kenya

V1 First published: 12 Jun 2020, 9:600
<https://doi.org/10.12688/f1000research.23259.1>

Latest published: 12 Jun 2020, 9:600
<https://doi.org/10.12688/f1000research.23259.1>

Abstract

Background: Although there has been an extensive scale-up of malaria interventions in Kenya, malaria infections persist at unacceptably high levels in some of the regions. Even with renewed calls to eradicate the disease through increased international donor assistance and country-specific government involvement, malaria is still a cause of worry in endemic regions. The objective of this study was to determine the factors associated with the incidence of malaria in Kisumu County over time.

Methods: The study conducted secondary analysis of data from a cross-sectional survey of routinely reported malaria cases. The population of interest were patients confirmed to have malaria by laboratory test. A sample size of 384 was randomly selected from all laboratory-confirmed malaria cases as reported by health facilities in Kisumu County from January 2014 to December 2017. The analysis involved descriptive, trend analysis and time series analysis (ARIMA). A negative binomial regression model was used to measure the effect of each of the selected predictor variables on incidence of malaria and the incidence rate ratio, was reported. Frequency distribution of each of the categorical variables was calculated.

Results: The overall pattern of the reported malaria cases had seasonal variations for weekly cases. The best-fitting time series model developed for predicting the number of weekly reported cases of malaria was ARIMA (2, 0, 1). It was observed that the negative binomial was actually the best model to fit the incidences of malaria because the dispersion parameter given by Poisson regression model had been reduced from 70.292 to 1.103.

Conclusion: There is a need to encourage health professionals to regularly review and report cases of malaria in their facilities. This is because reporting rates, completeness and the consistency of malaria reported cases remain extremely low.

Open Peer Review

Approval Status

	1	2
version 1		
12 Jun 2020	view	view

1. **Samuel Akech** , KEMRI-Wellcome Trust Research Programme, Nairobi, Kenya

2. **João L Ferrão** , Instituto Superior de Ciências de Educação, Beira, Mozambique

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Malaria, Trend Analysis, Modeling, Reported cases, ARIMA model, ACF and PACF plot, Negative Binomial, Poisson regression

This article is included in the **Emerging Diseases and Outbreaks** gateway.

Corresponding author: Elsen Achieng (elsen_a@outlook.com)

Author roles: **Achieng E:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation; **Otieno V:** Supervision; **Mung'atu J:** Supervision

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2020 Achieng E *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Achieng E, Otieno V and Mung'atu J. **Modeling the trend of reported malaria cases in Kisumu county, Kenya [version 1; peer review: 1 approved with reservations, 1 not approved]** F1000Research 2020, 9:600 <https://doi.org/10.12688/f1000research.23259.1>

First published: 12 Jun 2020, 9:600 <https://doi.org/10.12688/f1000research.23259.1>

Introduction

Malaria is among the major killer diseases within tropical regions. Sub-Saharan Africa bears its burden than any other region across the globe. The parasites are spread to humans through bites from infected *Anopheles* mosquitoes, introducing the protozoans into the human body. Even with renewed calls to eradicate the disease through increased international donor assistance and country specific government involvement, malaria is still a cause of worry in the endemic regions.

According to the world malaria report, anti-malaria progress has slowed down in many parts of the world and it is unlikely that the world will achieve targets for 2020 set by the World Health organization (WHO) Global Technical Strategy in reducing malaria incidences. A year on from this recognition, still there is no significant progress in global malaria control (WHO, 2018). Globally, 219 million cases of malaria were estimated to have occurred in 2017 as compared to 239 million cases that occurred in 2010. Even though there were 20 million fewer cases by in 2017 than in the year 2010 worldwide, from 2015 to 2017 there was a slight upward change in trend, suggesting that progress had generally stalled (World Malaria Report, 2018). The WHO statistics of 2017 reveal that Sub-Saharan Africa is still suffering greatly from malaria deaths, with 200 million cases comprising of 92% of the global burden. Malaria has impacted the health and economic development of Sub-Saharan African countries negatively and is considered a major impediment to sustainable development in the world's poorest regions (Gallup & Sachs, 2001).

The 2030 Agenda adopted in September 2015 by world leaders was to ensure sustainable development globally. The aim is to eradicate poverty, improving health, reduce inequity and address challenges related to climate change by the year 2030. In order to track progress, they created a number of Sustainable Development Goals (SDGs), each having a specific target to be realized in the next 15 years. One of the set targets is to end epidemics of neglected tropical diseases and malaria by 2030. A set of indicators to track success on malaria eradication is, the number of malaria incidence experienced in a population and the number of people affected by neglected tropical diseases and are seeking interventions against these diseases (SDG Indicators, 2018)

The inclusion of universal health coverage (UHC) reaffirms the position of the SDGs on global health priorities (Barasa *et al.*, 2018). The purpose of UHC is to make sure that all citizens of a state are able to access curative, rehabilitative and preventive health services at minimum costs, thereby eliminating disparities in access to health care. This also includes free consultation, diagnosis services for normal ailments and getting drugs for free. The Kenyan UHC model of dispensation by the current government calls for the enhancement of HIV, tuberculosis and malaria treatment in the country, with the aim of eradication of these diseases.

More than 70% of the entire population in Kenya is at risk of malaria which is the major cause of ill health (KNBS, 2015). In 2017, confirmed malaria incidences were 2,783,846 as reported

by the public health facilities. The diagnosed cases contribute correspondingly to the high number of malaria deaths in the Kenyan population, which increased from 2016 to 2017 as a result of stalled progress in malaria prevention.

With the aim of achieving the largest reduction rate in morbidity and mortality, the President's Malaria Initiative (PMI) has always considered the regions of Kenya experiencing the high incidences of malaria since 2013. The counties of Kisumu, Vihiga, Migori, Bungoma, Homa Bay, Kakamega, Busia and Siaya all together have a projected population of about 9.6 million in 2018 and bear the highest incidence of malaria in Kenya (President's Malaria Initiative, 2018a). The quarterly surveillance bulletin released in December 2016 revealed that, 73% of all confirmed malaria cases were reported from these counties (President's Malaria Initiative, 2018b). Execution of malaria control interventions focuses on endemic areas and are usually guided by the epidemiological zones. They include community managing the cases as they occur and intermittent preventive treatment in pregnancy (Musuva *et al.*, 2017)

It is in the light of this discourse that the Kenyan government and other developing partners in the health sector have remained dedicated in improving the delivery of health service in various parts of the country, with high priority being placed on prevention and control measures of malaria in endemic regions with eventual elimination. This being one of the objectives that were to be realized by 2018, together they aimed at reducing malaria burden by two-thirds (US President's Malaria Initiative, 2019).

Malaria situation with regard to facility reported cases and distribution of malaria remains unclear within the recent past in the study area, despite the above-mentioned studies in various parts of the country. Therefore, this study aimed at mapping the corresponding factors contributing to the incidence of malaria with the trend data on malaria cases across the region.

Methods

Study design and setting

A cross-sectional survey using routinely reported national Programme data on malaria cases in the DHIS2 platform, as reported by health facilities in Kisumu County was conducted from January 2014 to December 2017.

Kisumu County is located in the former Nyanza Province and its headquarters are in Kisumu City, which is situated approximately 370 km west of the Kenyan capital, Nairobi. According to 2009 census, Kisumu County had a population of 968,879 people and covers an area landmass of 2085.9 km² and 567 km² covered by water.

Study population and eligibility of participants

The study population included individuals at risk of malaria in Kisumu County.

All patients confirmed to have malaria by laboratory test were recruited in the study and patients diagnosed with ailments other than malaria were excluded.

Sample size determination

The Fisher *et al.* (1998) formula was used to determine the sample size for the study as follows:

$$n = \frac{Z^2 pq}{d^2}$$

Whereby:

n = is the desired sample size (if the target population is over 10,000) and Z = 1.96 which corresponds to 95% confidence level. p = is proportion of patients confirmed to have malaria by laboratory test while q = is (1-p) is proportion of patients confirmed **not** to have malaria by laboratory test. d = standard error at 95% confidence limit (0.05). Since the proportion of patients confirmed to have malaria by laboratory test (P) was not known, it was estimated to be 50% (Mugenda & Mugenda, 2013)

Therefore, based on the given figures

$$n = \frac{1.96^2 \times 0.5 \times 0.5}{0.05^2}$$

n = 384 laboratory-confirmed malaria patients

Data collection and study variables

Any patient who was referred to the laboratory by the clinical officer/nurse after presumptive diagnosis of malaria and fulfilled the criteria for inclusion was enrolled in the study.

Secondary data of all laboratory-confirmed malaria cases as reported by all Health Facilities was downloaded from the District Health Information System (DHIS2) using the standard application programming interface provided. The data included weekly number of malaria cases. Patients testing Positive for malaria was the dependent variable and the predictors were locality, Time in years and Total number of patients tested for malaria. The predictors were assessed as given in Table 1. A conceptual framework depicting the predictor-outcome relationship is displayed in Figure 1.

Ethical considerations

Considering that secondary data was used, there was no interaction with human participants in this study. However, all the personal information generated from this study was treated

Table 1. Dummy table for analysis.

Variable	Definition
Dependent variable	
Patients testing positive for malaria	Patients confirmed to have malaria by laboratory test
Independent variables	
Locality	Categories: (0 "Kisumu West", 1. "Kisumu East", 2. "Kisumu Central", 3. "Muhoroni", 4. "Nyando", 5. "Seme", 6. "Nyakach")
Time in years	Categories: 0 "2014", 1 "2015", 2 "2016" and 3 "2017"
Total tested	Number of patients tested for malaria at the health facility

Independent Variables

Dependent Variable

Figure 1. Conceptual framework.

with a lot of confidentiality. Approval to carry out the study was granted by the Ethics Research Committee at Kenyatta National Hospital/University of Nairobi (P511/05/2019). Permission to use the DHIS2 data was also granted Health Management Information Systems (HMIS) at the Ministry of Health Kenya (Ref. No.MOH/F/DHSMEI/01 Vol.I(17)).

Minimization of biases

Data was downloaded from the DHIS2 using the standard application programming interface provided. Each facility's data was visually cross-checked subjected to range and limit tests to confirm obvious outliers or errors in transcription.

Statistical analysis

Weekly facility-level data on malaria parasitological testing were extracted from the Integrated Disease Surveillance and Response (IDSR) registry through the DHIS2 platform. The analysis involved checking for completeness, consistency and accuracy of the data. This include descriptive, trend analysis and time series analysis (ARIMA). The ARIMA model was used and applied to time series data of malaria reported cases in Kisumu County.

Based on the Box-Jenkins approach, model building involved ascertaining the order of the AR and MA components which was guided by the Auto Correlation Function and partial correlation coefficients plots. Dicker Fuller and Shapiro-Wilk test were used to test for stationarity and normality respectively. Additionally, Pearson's Chi-square was used to test the goodness of fit of both Poisson and negative binomial regression models.

The influence of time in years, total number of patients tested and locality on the number of malaria cases was evaluated.

Frequency distribution of each variable and each category for each of the categorical variables was calculated and presented using R statistical software version 3.5.1. A significant count model negative binomial regression model was used to measure the effect of each of the predictor variables on the number of malaria cases and the incidence rate ratio (IRR), was reported. The statistical significance was measured using the 95% confidence interval and P-values. Missing data was addressed using single imputation method.

Results

Descriptive statistics of the study data

The study included 384 reported malaria cases in the final analysis, for the period January 2014 to December 2017. A line graph illustrating reported malaria cases from 2014–2017 is shown in [Figure 2](#). [Figure 3](#) shows the trend in weekly reported malaria cases.

Model identification

The procedure for model building involved ascertaining the order of the AR and MA components. This was guided by the auto correlation function and partial correlation coefficient plots based on the Box-Jenkins approach.

The autocorrelation function in [Figure 4](#) of malaria reported cases from 2014 to 2017 at 95% confidence level spikes significantly at lag 1, 2, 3 and 4. After lag 4 it cuts off, implying that the moving average component of the ARIMA model of order 4 (MA= 4) would be needed to describe this data set.

The partial autocorrelation function in [Figure 5](#) at 95% confidence level significantly spikes at lag 1 and then quickly cut off, implying that the autoregressive component of the ARIMA model is of order 1 (AR=1).

Figure 2. Line graph illustrating reported malaria cases from 2014–2017.

Figure 3. Trend in weekly reported malaria cases.

Figure 4. The autocorrelation function of reported malaria cases.

Figure 5. The partial autocorrelation function of reported malaria cases.

The ACF and PACF plots in [Figure 4](#) and [Figure 5](#) respectively suggest that a MA = 4 and AR =1 would be needed in describing this data set as coming from a moving average and autoregressive process, respectively.

Testing for stationarity

A constant mean, constant variance and a constant autocorrelation structure signifies a stationary series. The test for stationarity using augmented Dicker Fuller test as shown in [Table 2](#) depicts that the time series is stationary with a significant p-value of 0.01.

In comparing the Akaike information criterion (AIC) values of the five likely ARIMA models depicted in [Table 3](#), it was deduced that model ARIMA (2, 0, 1) is the best model for the data since it has the lowest AIC value of 5581.917.

Shapiro-Wilk normality test results ($p\text{-value} \leq 0.0001$) in [Table 4](#), depicts that the residuals are normally distributed. Consequently, it can be inferred that the true mean of the residuals is approximately equal to zero and there is a constant variance among residuals of the selected model.

Modeling incidence of malaria cases results of regression analyses

[Table 5](#) shows the residual deviance for the fitted Poisson regression was given as 26219 on 373 degrees of freedom. To check

Table 2. Testing for Stationarity.

Tests	Statistics	df	P-Value
Augmented Dicker Fuller Test	-4.7214	Lag = 5	0.01

Table 3. ARIMA model estimations.

Model	AIC
ARIMA (1, 0, 4)	5585.055
ARIMA (2, 0, 3)	5583.113
ARIMA (3, 0, 2)	5583.516
ARIMA (3, 0, 3)	5585.377
ARIMA (2, 0, 1)	5581.917

Table 4. Test for normality.

Shapiro-Wilk normality test
data: model5\$resid
W = 0.94476, p-value = 9.009e-11

Table 5. Parameter estimates of the Poisson model.

Covariate	Coefficients	Estimates	Standard errors	z values	pr(> z)
	Intercept	5.589	0.006874	813.061	< 0.001
Locality	Kisumu West (Ref)				
	Nyakach	-0.1208	0.007786	-15.52	< 0.001
	Kisumu East	-0.6122	0.007197	-85.068	< 0.001
	Muhoroni	-0.03054	0.006731	-4.537	< 0.001
	Seme	0.2473	0.006874	35.972	< 0.001
	Nyando	0.2244	0.006604	33.985	< 0.001
	Kisumu Central	-0.4922	0.01265	-38.913	< 0.001
Time in years	2014 (Ref)				
	2015	-0.0829	0.005579	-14.86	< 0.001
	2016	-0.07441	0.005671	-13.12	< 0.001
	2017	-0.2238	0.006251	-35.807	< 0.001
Total tested for malaria	Total Tested	0.0006022	0.000002217	271.587	< 0.001

Significance level <0.05.

the fit of the fitted Poisson model, the value of the residual deviance 26219 on 373 degrees of freedom was considered as observed in Table 5 $\frac{26219}{373} = 70.29223$, a dispersion parameter of 70.29223 is an indication that the data is over-dispersed as the value is far greater than 1. This implies that the model is not fit because the mean and variance of the response variable are not equal. If the mean and variance were equal, the residual deviance should be approximately equal to the degrees of freedom.

In this case, the assumption of mean equal to variance of the Poisson random variable was violated and so, a negative binomial regression model was deemed suitable and practical as they cater for over-dispersion. Also, they allow the likelihood ratio and other standard maximum likelihood tests to be applied. The fitted Poisson model had an AIC value of 29318 and a null deviance of 115777 on 383 degrees of freedom. Table 6 represents parameter estimates after validating the Poisson regression model using negative binomial regression model since the assumptions of Poisson regression was not met.

The AIC of this model is 5345.2, a deviance of 411.59 on 373 degrees of freedom also following the chi-square distribution $D \sim \chi^2 (n-p)$ degrees of freedom. The dispersion parameter was found to be 1.10346 as shown in Table 7. It was observed that the negative binomial was actually the best model which fit the incidences of malaria as the dispersion parameter given by Poisson regression model had been reduced from 70.29223 to 1.10346, an indication that the assumptions of Poisson regression were not met.

$\ln(\text{Incidence}) = 5.185 - 0.1155(\text{Nyakach}) - 0.4613(\text{Kisumu East}) - 0.0829(\text{Muhoroni}) + 0.3357(\text{Seme}) + 0.3001(\text{Nyando}) - 0.375(\text{Kisumu Central}) - 0.1022(\text{Year2015}) - 0.09154(\text{Year2016}) - 0.3084(\text{Year2017}) + 0.0008181(\text{Total Tested})$

Table 5 below shows the goodness of fit results, which clearly indicates that negative binomial regression model is a better fit for the incidence of malaria compared to the Poisson model. To begin with, the ratios of deviance and Pearson chi-square to degree of freedom $\frac{D}{DF} \sim \chi^2 (n-p)$ for the Poisson model are far greater than 1, which is an indication of over-dispersion and thus Poisson is not adequate for modeling such data.

Subsequently, $\frac{D}{DF} \sim \chi^2 (n-p)$ negative binomial model ratios are both close to one which exhibits a good fit to the data. Additionally, Pearson chi-square, the lower deviance, and the larger log likelihood values of negative binomial are against those of Poisson. Together all of them share a similar conclusion of favoring the negative binomial model.

From the Table 4, Adjusting for Kisumu East, Muhoroni, Seme, Nyando, Kisumu Central, years 2015, 2016, 2017 and the total number of patients tested for malaria, the average number of reported malaria cases in Nyakach is $e^{-0.1155(0.8909)}$, 11% less than those in Kisumu West Sub-County.

Adjusting for Nyakach, Muhoroni, Seme, Nyando, Kisumu Central, years 2015, 2016, 2017 and the total number of patients tested for malaria, the average number of reported malaria cases in Kisumu East is $e^{-0.4613(0.6305)}$, 37% less than those in Kisumu West Sub-County.

Adjusting for Kisumu East, Nyakach, Seme, Nyando, Kisumu Central, years 2015, 2016, 2017 and the total number of patients tested for malaria, the average number of reported malaria cases in Muhoroni is $e^{-0.0829(0.9204)}$, 8% less than those in Kisumu West Sub-County.

Adjusting for Muhoroni Kisumu East, Nyakach, Nyando, Kisumu Central, years 2015, 2016, 2017 and the total number of

Table 6. Negative binomial regression model parameter estimates.

Covariate	Coefficients	Estimates	Standard errors	z values	pr(> z)
	Intercept	5.185	0.08489	61.076	< 0.001
Locality	Kisumu West (Ref)				
	Nyakach	-0.1155	0.09111	-1.268	0.204866
	Kisumu East	-0.4613	0.09243	-4.991	< 0.001
	Muhoroni	-0.0829	0.0867	-0.956	0.338993
	Seme	0.3357	0.08801	3.814	0.000137
	Nyando	0.3001	0.08835	3.396	0.000684
	Kisumu Central	-0.375	0.124	-3.024	0.002491
Time in years	2014 (Ref)				
	2015	-0.1022	0.07161	-1.428	0.153335
	2016	-0.09154	0.07199	-1.272	0.203518
	2017	-0.3084	0.07215	-4.274	< 0.001
Total tested for malaria	Total Tested	0.0008181	0.00002961	27.633	< 0.001

Significance level <0.05.

Table 7. Assessment criteria for Poisson and negative binomial regression models.

Assessment Parameter	Poisson Regression model	Negative Binomial Regression Model
AIC	29318	5345.2
Residual Deviance	26219	411.59
Degrees of Freedom	373	373
Dispersion parameter	70.29223	1.10346

patients tested for malaria, the average number of reported malaria cases in Seme is $e^{0.3357(1.3989)}$, 40% more than those in Kisumu West Sub-County.

Adjusting for Seme, Muhoroni Kisumu East, Nyakach, Kisumu Central, years 2015, 2016, 2017 and the total number of patients tested for malaria, the average number of reported malaria cases in Nyando is $e^{0.3001(1.3499)}$, 35% more than those in Kisumu West Sub-County.

Adjusting for Nyando, Seme, Muhoroni Kisumu East, Nyakach, years 2015, 2016, 2017 and the total number of patients tested for malaria, the average number of reported malaria cases in Kisumu Central is $e^{-0.375(0.6873)}$, 31% less than those in Kisumu West Sub-County.

Adjusting for Kisumu Central, Nyando, Seme, Muhoroni Kisumu East, Nyakach, years 2016,2017 and the total number of patients tested for malaria, the average number of reported malaria cases in the year 2015 is $e^{-0.1022(0.9028)}$, 10% less than the year 2014.

Adjusting for Kisumu Central, Nyando, Seme, Muhoroni Kisumu East, Nyakach, years 2015,2017 and the total number of patients tested for malaria, the average number of reported malaria cases in the year 2016 is $e^{-0.09154(0.9125)}$, 9% less than the year 2014.

Adjusting for Kisumu Central, Nyando, Seme, Muhoroni Kisumu East, Nyakach, years 2015, 2016 and the Total number of patients Tested for malaria, the average number of reported malaria cases in the year 2017 is $e^{-0.3084(0.7346)}$, 27% less than the year 2014.

Adjusting for Kisumu Central, Nyando, Seme, Muhoroni Kisumu East, Nyakach, years 2015, 2016 and 2017, the average number of reported malaria cases $e^{0.0008181(1.0008)}$ increases by 0.08% for every increase in the number of patients tested for malaria.

Significance of the predictors

Kisumu East Sub-county is statistically associated with the incidence of malaria while adjusting for Nyakach,

Muhoroni, Seme, Nyando, Kisumu Central, Years 2015, 2016, 2017 and the Total number of patients Tested for malaria (p-value ≤ 0.0001)

Seme Sub-county is statistically associated with the incidence of malaria while adjusting for Kisumu East, Nyakach, Muhoroni, Nyando, Kisumu Central, Years 2015, 2016, 2017 and the Total number of patients Tested for malaria (p-value = 0.0001)

Nyando Sub-county is statistically associated with the incidence of malaria while adjusting for Kisumu East, Nyakach, Muhoroni, Seme, Kisumu Central, years 2015, 2016, 2017 and the total number of patients tested for malaria (p-value = 0.0006)

Kisumu Central Sub-county is statistically associated with the incidence of malaria while adjusting for Kisumu East, Nyakach, Muhoroni, Nyando, Seme, years 2015, 2016, 2017 and the Total number of patients Tested for malaria (p-value = 0.0024)

Controlling for all other covariates, Nyakach and Muhoroni localities do not show any significant association with the incidence of malaria in Kisumu County (p-value = 0.205, 0.339) respectively.

The year 2017 is statistically associated with the incidence of malaria while adjusting for Kisumu East, Nyakach, Muhoroni, Nyando, Seme, Kisumu Central, years 2015, 2016 and the total number of patients tested for malaria (p-value = < 0.0001)

Controlling for all other covariates, the years 2015 and 2016 do not show any significant association with the incidence of malaria in Kisumu County (p = 0.153 and p = 0.2014, respectively).

The total number of patients tested is statistically associated with the incidence of malaria, while adjusting for Kisumu East, Nyakach, Muhoroni, Nyando, Seme, Kisumu Central, years 2015, 2016 and 2017 (p-value ≤ 0.0001)

Discussion

Trend analysis

Preliminary and further analysis in this study, investigated the change in the yearly trends of malaria incidence in Kisumu County from 2014 to 2017 to determine if there is any increase or decrease in malaria reported cases. A trend analysis was applied to the weekly malaria reported cases and the appropriate ARIMA model fitted. The yearly cases have exhibited a continuous increasing trend except in 2017, where there was a decrease. The recorded yearly malaria cases for this study were also cyclical, for instance, the year 2016 represent an increase in cases over the years 2014 and 2015, whilst the year 2017 exhibited a decline compared with all other years.

This increase may be due to improvements in the reporting system by health facilities over time, non-usage of mosquito nets by the residents and low socioeconomic status of the population (Adenomon, 2014). This fluctuating incidence recorded yearly could also be attached to the people not being careful about the illness or inconsistencies by the health care authorities in the management of the illness. The decreasing trend in 2017 could be due to an increase in the use of mosquito nets, general improvement in the malaria awareness by the population of Kisumu owed to the intensive campaigns in malaria education by several organizations. Climatic changes could have also attributed to this (Anokye *et al.*, 2018).

The overall pattern of the reported malaria cases as reported by this study had seasonal variations. Weekly cases towards midyear recorded the highest number of cases, whereas the first weeks of the year recorded the lowest number of cases. This could be as a result of heavy rainfall experienced in the months of April, May, June and July, which is usually the rainfall season in Kisumu, and the temperature associated with rainfall seasons.

This confirms the findings of Craig *et al.* (1999) and Zhou *et al.* (2004) who revealed that that climate is a major factor in explaining the incidence of malaria. According to Zhou *et al.* (2004), the incidence of malaria is influenced by rainfall since mosquitoes need stagnant water for a complete life cycle. At the same time, Mabaso *et al.* (2007) also discovered that rainfall seasonality as well as minimum temperature are related to the number of *Plasmodium falciparum*-infective bites got by an individual annually or during a season.

The best fitted time series model developed for predicting the number of weekly reported cases of malaria was ARIMA (2, 0, 1). This suggests that ARIMA (2, 0, 1) can be utilized as a forecasting model to predict the future values of a series. ARIMA works best when data exhibits a constant pattern over time with a minimum amount of outliers (Labys, 2006). Researchers can use this model to forecast malaria reported cases (Mabaso *et al.*, 2001). Nonetheless, it should be updated from time to time with the incorporation of current data.

Significance of predictor variables on malaria incidences

Accordingly, the study investigated the association between time in years, locality and total number of patients tested for malaria in Kisumu using weekly malaria data from 2014 to 2017.

The study results revealed that residing in Kisumu East, Seme, Nyando and Kisumu Central sub-counties were statistically associated to the incidence of malaria as compared to residing in Kisumu West sub-county. Residing in Muhoroni and Nyakach sub-counties were not statistically associated with the incidence of malaria as compared to residing in Kisumu West sub-county. The different localities have demonstrated inequalities in malaria incidence mainly due to variations in access to health services, urbanization and wealth distribution. This is similar to findings by Galactionova *et al.* (2017), who identified

regional inequalities in the coverage of malaria interventions because of the inequality of the wealth distribution within and across many countries.

Many studies have also established that there is less access to services in the rural areas compared to the urban areas, leading to less reported malaria cases from the rural sub-counties. This makes this variable a useful determinant of malaria incidence since some localities are more urban than rural and vice versa.

Generally, there was fluctuation in malaria incidences during the last four years. The year 2017 was statistically associated to the incidence of malaria as compared to the year 2014. Whereas 2015 and 2016 were not statistically associated with the incidence of malaria as compared to the year 2014. Several factors may be responsible for seasonal changes, for example climatic changes, ecologic and environmental factors, social and economic determinants such as change in health care infrastructure. Availability of health facilities and drug resistance also have an impact on incidence of malaria. Although there were different malaria control activities in each year, such as activities to decrease incidence of malaria, insecticide indoor spraying, distribution of ITNs, massive malaria control campaigns and distribution of ITNs the prevalence is still constant. This is similar to findings by [Alemu *et al.* \(2012\)](#).

There was a significant association between the total number of patients tested for malaria and the incidence of malaria. Though urban areas have normally been at lower risk of malaria, unpredictable and unplanned population growth has been a key factor in making urban or peri-urban transmission an increasing problem. This could explain the high malaria incidence in 2016. This finding is consistent to a study carried out by [\(Knudsen & Slooff, 1992\)](#).

Conclusions

Trends analysis

The objective of this research was to determine the factors associated with the incidence of malaria in Kisumu County over time, given the locality where the patients sought health facility intervention, time in years and the total number of patients who were tested for malaria. The study utilized weekly malaria reported cases data from IDSR Disease Surveillance registry, for the period January 2014 to December 2017 in Kisumu County. Reported malaria cases confirmed by laboratory test was used in the analysis and modeling.

A seasonal pattern was observed in the malaria incidences in Kisumu County. ARIMA (2, 0, 1) model was found to be the best fit statistical model to predict malaria incidences in Kisumu County. The results found from this study offer useful information for policy makers to be able to effectively implement timely and effective malaria preventive and control measures.

Significance of predictor variables on malaria incidences

Goodness of fit model assessment criteria were used in selecting which model will fit the malaria reported cases better by

exhausting both Poisson and negative binomial regression models. The negative binomial regression model was found to fit the data better than the Poisson regression model, based on the results. In modeling the incidence of malaria, the quality of the Poisson model was estimated using the AIC. The AIC value produced was 29,318, the deviance was 26,219 and a dispersion parameter of 70.29 showing an over-dispersion in the data leading to a violation of one of its main assumption of the equality of mean and variance parameters. It became a necessary to validate the Poisson regression models using negative binomial because of over-dispersion.

The results obtained suggested that Kisumu East, Seme, Nyando, Kisumu Central localities, the year 2017 and the total number of patients who underwent a laboratory confirmation test for malaria were significant factors for incidence of malaria in Kisumu County over time. Whereas, Nyakach, Muhoroni localities, the years 2015 and 2016 were not significant factors for malaria incidences. The findings provide better insight of environmental and socio-economic effects on malaria and provide important information for malaria prediction. Nonetheless, there is need for additional studies to consider which other factors influence the incidence of malaria notwithstanding environmental and socio-economic factors. Likewise, health professionals practicing in Kisumu County should be encouraged to regularly review and report cases of malaria in their facilities. This is because, reporting rates, completeness and the consistency of malaria reported cases remain extremely low and importantly was poorer in localities where malaria is endemic.

Data availability

Source data

The data can be obtained from the Integrated Disease Surveillance and Response (IDSR) registry through the DHIS2 platform, using the standard application programming interface provided <https://hiskenya.org/dhis-web-commons/security/login.action>. Weekly facility-level data on malaria parasitological testing were extracted from Kenya DHIS2 between January 2014 to December 2017.

Access to the raw dataset is possible upon placing a formal request to the Ministry of Health, Kenya (emailing pshealthke@gmail.com), since the platform is kept under restricted access.

Acknowledgments

I thank the Almighty God for his guidance through my study period, without his provisions and guidance, my participation in this program of study would have been futile. Secondly, special thanks to the University of Nairobi administrators and staff at large for regular guidance on the whole research process. I also sincerely thank my friends and family for supporting, advising and encouraging me throughout the time I have worked for my project.

References

- Adenomon MO: **Modeling the prevalence of malaria in Niger State: An application of Poisson regression and negative binomial regression models** *Univariate and Multivariate Time Series Analysis View project Econometrics Analysis in Nigeria View project*. 2014.
[Reference Source](#)
- Alemu A, Muluye D, Mihret M, *et al.*: **Ten year trend analysis of malaria prevalence in Kola Diba, North Gondar, Northwest Ethiopia.** *Parasit Vectors*. 2012; 5(1): 173.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Anokye R, Acheampong E, Owusu I, *et al.*: **Time series analysis of malaria in Kumasi: Using ARIMA models to forecast future incidence.** *Cogent Social Sciences*. 2018; 4(1).
[Publisher Full Text](#)
- Barasa E, Nguhiu P, McIntyre D: **Measuring progress towards Sustainable Development Goal 3.8 on universal health coverage in Kenya.** *BMJ Global Health*. 2018; 3(3): e000904.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Craig MH, Snow RW, le Sueur D: **A climate-based distribution model of malaria transmission in sub-Saharan Africa.** *Parasitol Today*. 1999; 15(3): 105–111.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fisher LD: **Self-designing clinical trials.** *Stat Med*. 1998; 17(14): 1551–1562.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Galactionova K, Smith TA, de Savigny D, *et al.*: **State of inequality in malaria intervention coverage in sub-Saharan African countries.** *BMC Medicine*. 2017; 15(1): 185.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gallup J, Sachs J: **The economic burden of malaria.** *Am J Trop Med Hyg*. 2001; 64(1–2): 85–96.
[PubMed Abstract](#) | [Publisher Full Text](#)
- KNBS: **Kenya Malaria Indicator Survey 2015 - Kenya National Bureau of Statistics**. 2015. Retrieved May 2, 2019.
[Reference Source](#)
- Knudsen AB, Slooff R: **Vector-borne Disease Problems in Rapid Urbanization: New Approaches to Vector Control.** *Bull World Health Organ*. 1992; 70(1):1–6.
[PubMed Abstract](#) | [Free Full Text](#)
- Labys WC: **Modeling and forecasting primary commodity prices.** Ashgate. 2006.
[Publisher Full Text](#)
- Mabaso ML, Craig M, Ross A, *et al.*: **Environmental Predictors of the Seasonality of Malaria Transmission in Africa: The Challenge.** *Am J Trop Med Hyg*. 2007; 76(1): 33–38.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mugenda OM, Mugenda AG: **Research Methods: Quantitative and Qualitative Approaches I** Text Book Centre. 2013.
[Reference Source](#)
- Musuva A, Ejersa W, Kiptui R, *et al.*: **The malaria testing and treatment landscape in Kenya: results from a nationally representative survey among the public and private sector in 2016.** *Malar J*. 2017; 16(1): 494.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nobre FF, Monteiro AB, Telles PR, *et al.*: **Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology.** *Stat Med*. 2001; 20(20): 3051–3069.
[PubMed Abstract](#) | [Publisher Full Text](#)
- President's Malaria Initiative: **Kenya Malaria Operational Plan FY 2018**. 2018a.
[Reference Source](#)
- President's Malaria Initiative: **President's Malaria Initiative - Kenya Country Profile**. 2018b.
[Reference Source](#)
- US President's Malaria Initiative: **FY 2019 Kenya Malaria Operational Plan**. 2019.
[Reference Source](#)
- SDG Indicators**. 2018.
[Reference Source](#)
- World Health Organization: **WORLD MALARIA REPORT 2018**. ISBN 978 92 4 156565 3, 2018.
[Reference Source](#)
- Zhou G, Minakawa N, Githeko AK, *et al.*: **Association between climate variability and malaria epidemics in the East African highlands.** *Proc Natl Acad Sci U S A*. 2004; 101(8): 2375–2380.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 01 July 2021

<https://doi.org/10.5256/f1000research.25679.r88444>

© 2021 Ferrão J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

João L Ferrão

Instituto Superior de Ciências de Educação, Beira, Mozambique

This topic is very important and contributes to Malaria eradication efforts. The effort in malaria combat is multidisciplinary and mathematical modelling can greatly contribute. I see this study as good attempt. However, major revision should be addressed to fulfill the intention.

Objective needs a better clarification:

- If the intention is to model, means predicts malaria cases as in the title or if is "to determine the factors associated with the incidence of malaria" as in the abstract or, if is "mapping the corresponding factors contributing to the incidence of malaria". If the intention is mapping, I would expect to see maps.

Literature:

- There is open-source literature in malaria modelling using ARIMA in Africa. I don't see any in this article.
- I don't see any plausible reason for the use of outdated literature for example: (Gallup & Sachs, 2001), Zhou *et al.* (2004), (Mabaso *et al.*, 2001) with so many recent literatures.

Methods:

- Data: since they are historic data, the use of five or more years may improve your precision for trends.
- I don't understand why you used a sample instead of all data.
- 384 data for 208 weeks means less than 2 observations per week. This may not be enough for any trend analysis.
- Independent variables: number of patients tested for malaria is a socio-economic factor? Is location an environmental factor? What factor is time of the year?
- What was the level of missing data?

Results:

- Figure 1. Are you presenting incidence or malaria cases - can you clarify?
- Figure 3. Are you presenting trend on incidence or malaria cases - can you clarify?

Discussion:

- Results are for ARIMA, Negative binomial and, Poisson model. I don't see any discussion on these models compared to other authors.

Overall, I don't see the adding value of this research to science. Adding more data for analysis, defining clearly the objective may result in a better manuscript. Additional file with data used may be useful.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Modelling and mapping malaria risks. Expert in Information Management Systems (Statistics and Econometrics).

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 28 September 2020

<https://doi.org/10.5256/f1000research.25679.r71238>

© 2020 Akech S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Samuel Akech

Health Services Unit, KEMRI-Wellcome Trust Research Programme, Nairobi, Kenya

The study reports that it uses a random sample of 384 laboratory-confirmed malaria cases as reported by health facilities in Kisumu County from January 2014 to December 2017. The study concludes that there is a need to encourage health professionals to regularly review and report cases of malaria in their facilities because of incomplete reporting. The study uses time series analysis approaches. The study as written is quite difficult to follow and is not clear to discern a clear message from the paper. The title, the methods, and the conclusions are not connected. It is not clear how data were obtained, the method section is unclear, the analysis section is not clear and generally, the study is not replicable. Even the introduction section is lengthy but not focused and the aim of the study does not come out clearly.

Sample size calculation shown is for investigation of prevalence, but the study investigates determinants and time trends. It, therefore, seems that the sample size calculation presented is not appropriate. The sample size is not appropriate to address the results and discussion presented by the authors.

The authors state 'Weekly facility-level data on malaria parasitological testing were extracted from the Integrated Disease Surveillance and Response (IDSR) registry through the DHIS2 platform.' It is not clear how the authors did this since data in DHIS2 platform has monthly summaries. It is not clear how the DHIS2 data were combined with data from the cross-sectional survey.

Overall comment: The authors ought to review the data they have and identify a clear message that the data can address and re-write the manuscript to bring out a clear message. This should also involve modifying the title to be appropriate, write methods, and data analysis section clearly.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

No

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** Paediatrics, epidemiology, sepsis, malaria, clinical trials

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research