Spatial Modelling of Malaria Prevalence in Kenya

Morris Mwenda John[1*], Elphas Luchemo[1] and Ayubu Anapapa[1]

[1]*Department of Mathematics and Actuarial Science, Murang'a University of Technology, Kenya.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Original Research Article

Abstract

Malaria is one of the leading causes of deaths in Kenya. Malaria is a vector-borne disease caused by a parasite of the genus plasmodium. Complete eradication of malaria in the country has remained a problem. A lot of effort and resources has been put in the fight against malaria in developing countries which has led to underdevelopment and low human development index. Malaria burden affects the world's poorest countries. About 90% of the malaria burden is reported in sub-Saharan Africa. The disease has led to high mortality cases in children and pregnant women. Despite the massive government eradication campaign, new and resurgent cases have been recorded. The specific objective was to determine the malaria risk factors and spatial distribution in Kenya. The 2015 malaria indicator survey data was used for the study. Demographic and social-economic factors were used as predictor variables. A generalized linear mixed model was used to determine the spatial variation and prevalence of malaria in Kenya. Demographic and social-economic factors were found to have significant impact on Prevalence of malaria in kenya. Most cases of malaria were reported in lake, western and coastal regions. The most prone areas were Kisumu, Homabay, Kakamega and Mombasa. There were less cases in central Kenya counties like Nyeri, Tharaka-Nithi with a significant number reported in arid and semi-arid regions of Northern-Kenya counties of Garissa, Mandera, Baringo. Rural population was more susceptible to malaria compared to those in urban areas. The odds of getting (verse not getting malaria) in places of residence increases by 1.32, which is estimated to .28, CIs 95% (1.01, 1.72), and a p-value .04. Malaria prevalence varied significantly from one region to another. The study established that Spatial autocorrelation exists among regions mostly due to weather patterns, geography, cultural practices and socio-economic factors.

_____

Corresponding author: Email: mwendajm1@gmail.com;

1 Introduction

Malaria is a disease caused by single cell parasites of genus Plasmodium. There are four identified species of the genus plasmodium: *plasmodium vivax, plasmodium malaria, plasmodium ovale* and *plasmodium falciparum*, of these species, *plasmodium falciparum* poses the threat. The disease is introduced into the human circulatory system from bites caused by infected female anopheles' mosquitoes. Of the reported cases, over 90% of deaths resulting from the disease have been identified to be caused by parasites of genus *plasmodium falciparum* [1]. Malaria is one of leading causes of mortality and morbidity in Kenya, the occurrence of malaria in many parts of the world has been correlated with poverty, low economic growth and ranked low in the human development Index. Malaria is associated with stalled growth in many countries with records of up to 1.3% annual growth penalty in many countries where malaria is endemic, especially in most of Sub-Saharan African countries [2].

In Kenya, statistics have shown that over half of the population is at risk of malaria, accounting for more than 40% of outpatients in hospital attendance with over 20% being admissions to health facilities [3]. It has been estimated that a lot of man hours are lost annually to the disease with an estimated 170 million working days lost. According to Mogeni et al. [4] reports from the government show that pregnant women and the children under the age of 5 years are the most vulnerable group to malaria, with an estimated 20% of deaths reported in children below 5 years. Several environmental factors greatly influence transmission of malaria in various regions in the country. Population density, proximity to wetlands among many other factors both social and economic were identified as some of local spatial variations with a significant association with the spread of malaria. In some other regions like Senegal and Ethiopia, there has been a demonstration of a spatial relationship between climatic variabilities like rainfall and the occurrence of malaria [5].

There is a knowledge gap in our communities, and rampant misconceptions on Malaria. However, the risk and transmission intensity of malaria shows reasonable temporal and spatial variations directly related to variations in temperature, rainfall, altitude, human settlement and topography [6]. The biology of mosquitoes is greatly influenced by climatic conditions. Therefore, climate conditions influence its transmission in malaria endemic countries. Major effort has been put in the fight against malaria. For instance, in late 1800, in the United States of America, wire gauzes were introduced in most housing facilities to act as physical barriers to the fight against the parasite. Although crude, by then the method seemed to work. In other areas around the world, like Algeria, oil was spilled on water bodies thought to harbor the parasites as a control mechanism for the mosquito larvae development. However, the continued use of this method rendered large swathes of water bodies not safe for other uses like animal consumption. A factor that made the method less attractive to the locals.

Although associated with some degrees of success, the method was not ideal thus was shunned. Within the course of the early 20th century to its mid years, several methods to combat the problem of larvae that grew to adult anopheles' mosquitoes were adopted. Some of these were reported to be toxic, ranging from the use of Arsenic and its derivatives. Known as the wonder drug, Paris green was extensively used in and around Europe to fight against malaria in spite of the availability of other pesticides like quinine. Following the introduction of the Paris green, most countries in the 1940s had registered elimination of *Anopheles gambiae*. These include countries like Brazil and Egypt. However, it is important to note that there has not been a single effective method to combat malaria world over. This has prompted the use of different approaches around the world to fight malaria [7].

Indoor residual spraying (IRS) has over the years come up as one of the main methods used for protecting persons from malaria especially in malaria endemic regions. For instance, there has been a mega leap in the number of persons protected, from slightly above 10 million mark to over 50 million individuals as a result of using this method [8]. This represents over 8% percent of persons at risk of malaria in Africa. The use of treated mosquito nets, mosquito repellants and some other off-the-shelve sprays has been hailed as the major method of combating the disease in most households especially in Kenya. As observed by Zhou et al. [9], Indoor spraying with insecticides in areas where most of the Anopheles species have been identified to habit has largely been a

success**.** The use of long-lasting mosquito bed nets (LLINs), that have been treated with insecticides in beds has been hailed as one of most effective methods.

2 Literature Review

In the study, [10] carried out research on modelling of the geostatistical malaria risk data in Mali. The Bayesian geostatistical model was used to the malaria risk data in quantifying the environmental-disease relations, identification of significant malaria predictors and transmission in the environment and provided model-based malaria risk predictors together with their precision. The models used in the study were based on the assumption of stationarity. In the study, they implied that the spatial correlation is a function of the distance between location and independence of location. A Bayesian non-stationarity model was used in the study. The model fit for the study as well as its predictions was based on the Markov chain Monte Carlo methods (MCMC). From the study, the results indicated that the stationarity assumption was a very important factor, since it had a major influence on the significance of environmental factors and the corresponding malaria risk maps.

Between the years 2007 and 2015, parasite prevalence was used as a measure to determine the prevalence of malaria in areas with high endemic cases. These studies established that more conclusive investigations were not carried out in order to establish the malaria mortality cases in different age groups. [11] in the study used parasitological data from the annual cross-sectional surveys taken from Kisumu (HDSS) in determining the prevalence of malaria parasites. The study used household surveys and verbal autopsies (VA) in obtaining the data on the causes of malaria specific mortality. The study used Bayesian negative binomial geo-statistical regression models in the investigation of parasite prevalence of clinical malaria and mortality in the various age groups. The estimates were obtained based on yearly data from the aggregated data collected for over a period of one year. The results from the study were that malaria parasitaemia from cross-sectional surveys was associated with mortality in the various age groups. The study showed that clinical malaria was associated with cases of malaria mortality more than the parasite prevalence. The study noted that the effects were much higher in children between the ages of 5 to 14 years in comparison with other groups.

Homan et al. [5], explored the role that local demographics, social and environmental factors played on malaria across various regions in western Kenya. The study explored the spatially varying risk factors for malaria in Kenya in order to obtain insights on how the human and environmental factors play in sustaining the transmission and prevalence of malaria. The study used a standard linear regression model where multiple variables were fitted to help explain how much of spatial variation on malaria prevalence could be explained by the demographic and environmental data. Making the assumption of non-stationarity of the risk factors, the geographically weighted regression walk (GWR) was performed. The effects of local multicollinearity and spatial autocorrelation were given much more attention to detail. The results from the study were that, combining the data from these surveys, a multivariate linear model showed that outdoor occupation, population density, socio-economic status, and population density increased the risk of malaria, the model included environmental and household factors. The local geographically weighted regression (GWR) model made substantial improvements and the model fitted considerably, the results of the study was that the relationship of malaria and the risk factors varied spatially in this region. Depending on the area of the island, outdoor occupation, population density, as well as socio-economic status had a positive or negative association with the prevalence of malaria.

Tuyishimire [12] carried out a study in the southeastern region of Rwanda to model the malaria risk factors. The study aimed at assessing the relationship between the malaria risk factors and its prevalence in this region, the study spatially models malaria risk factors. In the study, spatial statistics was used to obtain the spatial clusters of malaria prevalence in the study regions. The analysis on these clusters showed that zones with high malaria risk, also known as malaria hotspots as well as those zones with low and moderate distribution of malaria risk characterized the distribution of malaria. The study showed that the prevalence of malaria varied from one household to another, the same case for one administrative unit to the other. The study used a logistic regression model to assess the relationship between the malaria risk factors and the prevalence of malaria. The results from the study clearly showed that the rate of infection with malaria increased with proximity to the irrigated farmlands. More so, the size of the household was also seen to increase the rate of malaria infection

considerably. Similarly, poor housing quality was associated with a high risk of malaria infection. These factors (size of household, low housing quality and irrigated farmlands) were seen as the main malaria underlying factors.

A survey was carried out in Uganda to establish the prevalence of malaria among the children below five years, the study explored the progress of the control interventions and then established how malaria is geographically distributed across the country. A Bayesian geostatistical model with statistically varying coefficient was used in determining the effects of interventions at both national and regional level. To identify the most important forms and predictors, spike and slab was used in selecting the variables. The results from the study showed that indoor residual spraying (IRS) and insecticide treated nets (ITN) had a very significant effect overall but varied in terms of protective effect on the prevalence of malaria. Environmental factors like rainfall, land cover, temperatures of day and night were significantly associated with prevalence of malaria. The study found out that malaria was significantly high in rural areas compared to urban areas, on the other hand, malaria prevalence was seen to increase with age of the child while higher education levels and high household socio-economic status was seen to drastically decrease the cases of malaria prevalence [13].

In a study to investigate the spatial relationship between the occurrence of malaria and environmental risk factors in South Sumatra province in Indonesia, the study investigated six potential ecological factors associated with malaria. Ordinary least squares method (OLS) and geographically weighted regression (GWR). The spatial variability and the global patterns of associations between the cases of malaria and a set of selected ecological factors were explored. The results from the study were that different geographical and environmental parameters were important at different global and village levels. Distance from the forest, altitude and rainfall were used as independent variables. In global OLS, these variables were seen to have a significant effect on cases of malaria. However, the relationship between malaria and environmental factors had a strong spatial variation in different regions [14].

Having the right knowledge on how the primary vectors are distributed is crucial since it plays a great role in providing key information in the designing of transmission models so as to target control measures [7]. On the other hand, they are crucial to understand how the existing interventions have an impact on the relative abundance of these vectors, where an alternative control is needed. In this study, predictive species distribution models generated suitability probability, a multinomial generalized additive model was used in producing the relative abundance estimates for *anopheles funestus, Anopheles arabiensis and anopheles gambiae.* The models estimated effects of indoor insecticide interventions on these abundances were made using the post-Intervention maps. The results of the study were that the abundance of several different risk factors differed depending on habitat preference. It also showed that each intervention strategy varied depending on the risk factors. Residual spraying had a great impact on the relative abundance of *anopheles fenustus* and lesser effects on *anopheles gambiae.* However, the results noted that insecticide treated nets reduced the relative abundance of these species drastically and in equal measures [15].

3 Materials and Methods

3.1 Source of data

The study is based on secondary data. Data from malaria indicator survey of 2015 (KMIS 2015) in collaboration with Kenya National Bureau of Statistics (KNBS) was used for the study. The response variable of the study was malaria outcome. The predictor variables (χ_i) variables for the study included social-economic and demographic variables. Some of the best predictor variables included: Place of residence, wealth index, education level, Marital status and age. The dataset has both continuous and categorical variables. This dataset has a binary outcome (response, dependent) variable, malaria status (represented with 0s (no malaria) and 1s (has malaria). The predictor variables are factored to categorical variables—dichotomous and multiple categories except the age of the respondents, which is a continuous variable.

3.2 Statistical method: Logistic regression

A logistic regression model is ideal in cases where outcome variables are categorical. A univariate standard logistic model was fitted between each single covariate with the outcome variables (Malaria status). This model is appropriate in the case where the dependent variables are dichotomous. To investigate the status of malaria; Let γ_{ij} be the status of the disease. Where *1* for individuals with the disease and zero otherwise. For the individual j residing in the region *i*. Therefore, *i=1, 2... 47*. Thus $\gamma_{ij} = 1$ **f**or an individual *j* recorded positive for malaria residing in county *i*. The study assumes that $\gamma_{ij} = 1$ which is a dependent variable is a univariate Bernoulli distributed;

$$\gamma_{ij}|\rho_{ij} \sim Bernoulli(\rho_{ij}). \tag{1}$$

The ρ continuous independent variables are contained in the vector $\chi_{ijk} = (\chi_{ij1}, \chi_{ij2}, \dots \chi_{ij\rho})$ whereas $\omega_{ijk} = (\omega_{ij1}, \omega_{ij2}, \dots, \omega_{ijr})$ contains the r categorical independent random variables. In this model, the first component accounts for the intercept. *(p=1 (age), r = 0...)*

The unknown mean responses in the study: $E(\gamma_{ij}) = \rho_{ij}$ relates to the independent variables as:

$$h(\rho_{ij}) = X^T \beta_1 + \omega^T \gamma \tag{2}$$

The logit link function is shown as h (.), For the continuous independent variables, β represents the regression coefficient vectors of p dimension. The categorical independent variables are represented by γ which is a vector of r dimensions. A random walk model of order 2(RW2) accounts for the spatial autocorrelation and for non-linear effects of the continuous variables. The model relaxes the restrictive linear predictors to more semi-parametric predictors as;

$$h(\rho_{ij}) = \sum_{t=1}^{\rho} ft(\chi_{ijt}) + f_{spat}(S_{i1}) + \omega^T \gamma 1. \tag{3}$$

This model is used to determine the risk factors of malaria in Kenya.

For continuous covariates, Function $ft(.)$ is a twice differentiable non-liner smooth function. $f_{spat}(S_{ik})$ accounts for spatial effects for each of the counties. The study used a convoluted structure, here the assumption is that spatial effects can be decomposed into spatially unstructured and spatially structured Components [16];

$$f_{spat}(S_{ik}), = f_{str}(S_{ik}), (S_{ik}) + f_{unstr}(S_{ik}), \text{k=1} \tag{4}$$

The spatially unstructured random effects cover for the correlation or the unobserved covariates inherent within the counties, like climate, culture or social-economic activity. The spatially structured random effects account for the un-observed covariates varying spatially in these counties. This defines the spatial autocorrelation. Technically, it refers to the dependence due to the geographical proximity. Therefore, the final model is expressed as:

$$h(\rho_{ijk}) = \sum_{t=1}^{\rho} ft(\chi_{ijt}) + f_{str}(S_{ik}) + f_{unstr}(S_{ik}) + \omega^T \gamma 1$$
with k= 1 for malaria $\tag{5}$

The model is important in relaxing the stationarity assumption on the effects of risk factors on the prevalence of malaria in Kenya.

Parameter Estimation:

The study used a fully Bayesian estimation approach: The parameters were assigned prior distributions in the study.

3.3 Non- linear effects

There are several methods that have been extensively used in estimating of smooth functions, $(.)$. Among the most used models is the penalised regression splines model used by Eiliers and Marx. In this model the assumption is that the effects of the continuous covariates can be estimated using the polynomial spline (Eilers & Marx, 1996)**.** In the study, the random walk model was used in the estimation of the smooth function $ft(.)$.

3.4 Spatial model

In Spatial processes, the assumption that the relationship between the explanatory variable and the response variable is constant across the study regions does not hold. This relationship is affected by a number of factors in sampling variation like culture and attitude of people. In some areas different responses to a similar stimulus are observed as we move across the space. The conditional autoregressive (CAR) model relaxes this stationarity assumption by allowing the covariates to vary spatially.

The spatial model specifies the distribution of data conditional on the unknown parameters. Subsequently the Unknown parameters are specified conditional on other parameters under study. Thus:

$$\gamma_{ij1}|\rho_{ij1} \sim Bernoulli(\rho_{ij1})$$

$$h(\rho_{ij}) = X^T \beta_1 + \omega^T \gamma \tag{6}$$

Therefore, for the regression coefficients, $[u_\gamma, \Sigma_\gamma] = N(1_{\eta \times 1} \otimes u_\gamma, \Sigma_\gamma)$ defines the prior distribution:

Where:

$\mu_\gamma = (\mu_{\gamma0}, , \mu_{\gamma1}, ... , , \mu_{\gamma\rho})T$ is the mean vector for regression coefficients corresponding to the explanatory variable ρ**.** The covariance Σ_γ takes into account the spatial dependence where the priors for y^s are put into account/ specified.

3.5 The Conditional Auto-Regressive (CAR) model

The model describes observations varying over discrete sets of indices, like the number of disease cases recorded across regions and counties. They are extensively used to put into account correlations among the observations of interest in hierarchical models.

Consider a random vector $\mu = (\mu_1, ... , \mu_n)$. Assuming that each of the components is univariate and fixed at a fixed site $i \in \{1,2, ... , n\}$, these sites can represent either regions in space or points. The components of μ on the other hand can either be discrete or continuous. The CAR model allows to obtain the multivariate distributions for μ from univariate specifications based on local properties for regions of interest.

Assuming that the conditional distribution $p(\mu_1|\mu_{j \neq i})$ Depends only on those sites that are neighbours of i. Only when j is defined as a neighbour of i if and only if the conditional distribution at site i depends on value at site j . For the Gaussian Conditional Auto-regressions:

A random vector $\mu = (\mu_1, ... , \mu_n)T$ follow the CAR model as described if $\mu|\mu_{-1} \sim N(\sum_{j \in di} a_{ij}u_j,)$

Where

$\mu_{-1} = \{u_j, j \neq i\}$, and di denotes the sets that contain indices which are neighbours of i. $\tag{7}$

3.6 Posterior distribution

The posterior distribution defines the way the parameters are distributed after observing the Data. Hence, once the prior distribution is updated with the observed data, the prior distribution is obtained. Thus, sampling from the posterior distribution, it is then those inferences are made. The most common method for inferencing is the Markov Chain Monte Carlo (MCMC), however the method is very slow and performs poorly when such models are used. Thus, in order to go about these shortcomings, Integrated Nested Laplace (INLA) criteria is introduced [17]. Though a relatively new model, it works well in such cases. Therefore, for the latent Gaussian model, the Posterior Distribution becomes:

$$\pi(\chi, \theta | \gamma) \propto \pi(\theta)\pi(\chi | \theta) \prod_{i \in 1} \pi(\gamma_i | \chi_i, \theta)$$

$$\propto \pi(\theta)|Q(\theta)|^{\frac{n}{2}} exp\left(-\frac{1}{2}\chi^T Q(\theta)\chi + \sum_{i \in 1} log\, log\, \pi(\chi_i, \theta)\right)$$

(8)

The class of latent fields is represented by *x*, 0(theta) represents the set of hyper parameters, *y* represents the data. Therefore, the approach for INLA, the posterior marginal of interest is:

$$\pi(\gamma) = \int \pi(\theta, \gamma)\, \pi(\theta | \gamma) d\theta \text{ and } \pi(\gamma) = \int \pi(\gamma) d\theta_J$$

(9)

Therefore, these are used in the creation of nested approximations:

$$\tilde{\pi}(\gamma) = \int \tilde{\pi}(\theta, \gamma)\, \tilde{\pi}(\theta | \gamma) d\theta \text{ and } \tilde{\pi}(\gamma) = \int \tilde{\pi}(\gamma) d\theta_J$$

(10)

3.7 Odds ratio (OR)

Odds Ratio is the ratio that defines the probability of an event taking place or not happening. In Logistic regression, the Odds ratio refers to the exponential of the estimated coefficients $exp\,(\hat{\beta})$ and for each continuous covariate *j*, where $exp\,(\hat{\beta})$ is the predicted change in Odds for the *j* predictor variable. The study has multiple independent variables; thus, they are categorized into *k* levels and their Odds ratios (OR) is subsequently determined,

$$OR = \left[\frac{P(Y=1/x_i)}{1-P(Y=1/x_i)}\right] = \left[\frac{\pi}{1-\pi}\right] = exp(_o + {}_1X_1 +, \dots, + {}_kX_k)$$

(11)

Logit function= log (OR)

$$OR = Log\left[\frac{P(Y=1/x_i)}{1-P(Y=1/x_i)}\right] = Log\left[\frac{\pi}{1-\pi}\right] = (_o + {}_1X_1 +, \dots, + {}_kX_k) = \sum_{j=0}^{k} {}_jX_{ij}$$

(12)

Where $I = 1,2,\dots, n$ and $j=0,1,2,\dots, k$

$Log\left[\frac{\pi}{1-\pi}\right]$ is important in contingency table analysis ("log Odds") which has two columns and many rows of values of *x*.

4 Results and Discussion

In this section, a logistic regression is used to explain each individual variable that determines the malaria condition. Each independent variable is fitted into the model and considered significant at 0.05 significance level and 95% confidence level. The results are as Presented.

Table 1.

Variable	Estimate	Std. Error	z value	Pr(>|z|)		OR	2.50%	97.50%
(Intercept)	-14.44	227.95	-0.06	0.95		0.00	0.00	5.81E+187
1. age_of_respondent								
age_of_respondent	-0.05	0.03	-1.43	0.15		0.95	0.89	1.02
2. age_in_5_years_group								
20-24	13.09	227.95	0.06	0.95		486142.90	0.00	5.25E+199
25-29	13.36	227.95	0.06	0.95		633485.20	0.00	6.84E+199
30-34	13.61	227.95	0.06	0.95		814393.40	0.00	8.80E+199
35-39	13.51	227.95	0.06	0.95		739822.00	0.00	8.00E+199
40-44	13.94	227.95	0.06	0.95		1138210.00	0.00	1.23E+200
45-49	14.02	227.95	0.06	0.95		1230725.00	0.00	1.33E+200
3. region								
Nairobi	-0.14	0.22	-0.63	0.53		0.87	0.57	1.34
4. type_of_place_of_residence								
rural	0.28	0.14	2.03	0.04	*	1.32	1.01	1.72
5. highest_education_level								
primary	-0.61	0.31	-1.97	0.05	*	0.54	0.30	1.00
secondary	-0.45	0.32	-1.41	0.16		0.64	0.34	1.19
tertiary	-0.10	0.35	-0.28	0.78		0.91	0.46	1.80
6. marital_status								
married	0.14	0.22	0.63	0.53		1.15	0.74	1.79
living together	0.23	0.27	0.88	0.38		1.26	0.75	2.13
windowed	-0.01	0.33	-0.03	0.97		0.99	0.52	1.89
divorced	0.58	0.32	1.81	0.07	.	1.79	0.95	3.34
seperated	0.23	0.27	0.86	0.39		1.26	0.74	2.13
7. wealth_index								
poor	0.21	0.37	0.57	0.57		1.23	0.60	2.55
middle	0.05	0.36	0.13	0.90		1.05	0.51	2.14
rich	0.53	0.36	1.47	0.14		1.69	0.84	3.42
richest	0.16	0.37	0.44	0.66		1.18	0.57	2.45
8. has_mosquito_net								
yes	-0.10	0.13	-0.82	0.41		0.90	0.70	1.16
9. county_of_residence								
Nyeri	0.05	0.21	0.24	0.81		1.05	0.70	1.58
Kirinyaga	0.02	0.20	0.12	0.90		1.02	0.70	1.50
Murang'a	-0.98	0.27	-3.65	0.00	***	0.37	0.22	0.64
Nairobi	0.52	0.23	1.78	0.05	*	0.64	0.52	1.08
10. endemicity_zone								
lake endemic	0.54	0.18	2.97	0.00	**	1.71	1.20	2.44
coast endemic	-0.12	0.26	-0.46	0.65		0.89	0.54	1.47
semi-arid, seasonal	0.34	0.19	1.80	0.07	.	1.40	0.97	2.01
low risk	0.14	0.22	0.63	0.53		1.15	0.75	1.75

The results show different analyses of the dataset using a logistics regression model. Other than the estimated value of individual variables, there are coefficients and their standard errors, the z value, the p-value, the odds ratio, and the confident intervals at 0.05. Since the variables are categorical, some terms have strong statistical significance, weak statistical significance, and others never reached statistical significance at all. Murang'a county and lake endemic zone have strong statistical significance with odds ratios (OR) 0.375 and 1.71 respectively. Rural residence and primary level terms also reached strong statistical significance with 1.32 and 0.543 OR respectively. Divorced and semi-arid seasonal endemicity zone terms reached weak significant levels, which have 1.79 and 1.4 OR respectively. The terms are listed in order of degree of statistical significance, from strong to weak.

The study used the OR to define and conclude the relationships of independent variables to the logistics regression model of the outcome variable, malaria status. Considering the OR description of the individual variables, the study categorized them into three groups. The first group of OR greater than 1 (OR>1), second group OR = 1, and third group OR < 1. The age in 5 years groups all have OR > 1, however, the odds of group respondents aged 40-49 are lowest compared to other age groups on malaria cases (positive vs negative). Age groups 20-24 to 25-29 have a lower chance of being infected by malaria than older people of groups 40-44 and 45-49. The chance of aged respondents contracting malaria is 1 time that of youth. Though, all groups do not reach statistical significance. Respondents in rural areas had odds of 1.32 times higher malaria outcome effects

versus no effect on those living in the urban areas and were statistically significant to the malaria outcome. The odds of getting (verses not getting malaria) in places of residence increases by 1.32, which is estimated to .28, CIs 95% (1.01, 1.72), and a p-value .04. Respondents in urban areas are less susceptible to malaria effects than those in rural areas, this could be attributed to the intervention measures and access of information in the respective areas.

The level of education (highest education level) odds given the presence of malaria versus no malaria are indirectly proportional to the level of education. Respondents of primary level have odds of .543 and were statistical significance, CIs 95% (.295, .998), p-value .049. Respondents of education levels under secondary and tertiary categories do not reach statistical significance while respondents with no education were used as the default reference category with no effect on the outcome variable (malaria cases). In general, less educated respondents are reported to be more prone to malaria and its subsequent effects compared to people in more educated categories, of those that reported access to both secondary and tertiary education.

The odds of malaria cases on marital status (divorced) are 1.79 times to get malaria cases than others terms of marital status (married, windowed, living together, and not married) and was statistical significance, with CIs 95% (.954, 3.34), and a p-value .069. Respondents who are married, living together, widowed, and separated do not reach statistical significance. Widowed respondents are slightly different from other marital status with OR <1 (.907).

The odds of malaria in counties of residence are different; respondents from Murang'a county reached statistical significance while others did not. Murang'a has odds of .375 reaching statistical significance, CIs (.221, .635), and p-value .0002. This implies that the spread of malaria is different in different counties.

The odds of malaria outcome to the endemicity zone varies by zone. Two out of five are statistically significant—lake endemic and semi-arid seasons. Their coefficients are OR 1.71, CIs (1.2, 2.44), p-value .002 and OR 1.40, CIs (.97, 2.01), p-value .07 respectively. Respondents in lake endemic zones are 1.71 times more prone to malaria than those in coastal endemic Zones. Those in semi-arid, seasonal and low risk zones are second in vulnerability to malaria. Those lake endemic zones are Siaya, Busia, Kisumu, Migori, Kakamega, Homa-bay, Bungoma, and Vihiga. As such, these areas recorded more malaria cases than any other region in the country. These findings are consistent with [5], that occupation, population density, social economic status, and population density increases the risk of malaria, the model included environmental and household factors.

The categorical variables in the study of malaria outcome and their terms/groups are not consistent with the dependent variable (malaria cases outcome). For those that reached statistical significance, some have a strong relationship, others weak, and no relationship to the outcome variables at all. The analysis is based on the OR, CIs, and P-value of the variables. The lake endemic zone residence has low protection to malaria with coastal zones following closely, these regions reported more cases compared to other zones. Other people prone to malaria effect are divorced and people with low education levels.

4.1 Spatial distribution of malaria in Kenya

In epidemiological modelling, Bayesian approach is essential as it allows the modelling of these epidemiological problems hierarchically. Employing the approach, in order to attain reliable estimates, the method was essential in determining the spatial distribution of Malaria in Kenya. As proposed by Eilers & Marx [18], in order to perform a semi-parametric modelling to allow the spatial variation of the outcome variables, the study used B-splines, penalized likelihood as well as penalized regression splines, to determine the regional variation of malaria. The assumptions of linearity allowed some covariates like age to have a non-linear effect on prevalence of malaria. Using the random walk model of 2nd order and the assumption of stationarity, this allowed covariates to vary spatially using the conditional autoregressive model. Thus, in order to model the spatial effects in the model, Bayesian geostatistical model was used on the malaria risk data in quantifying the environmental-disease relations, identification of significant malaria predictors and transmission in the environment and provided a model-based malaria risk predictors as suggested by Gosoniu et al. [10]. The models used in the study were based on the assumption of stationarity. In the study, they implied that the spatial correlation is a function of the distance between location and independence of location.
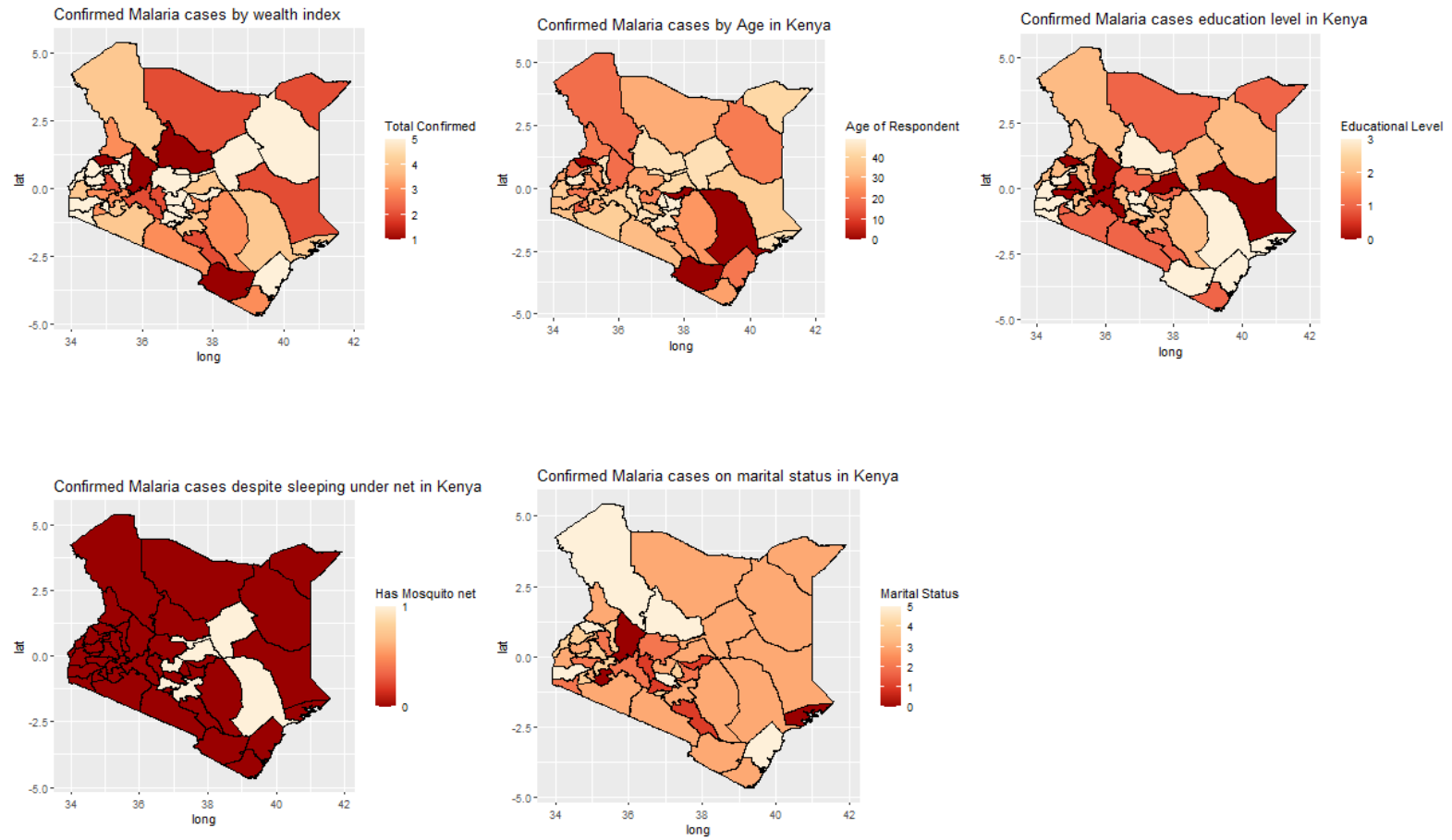
Fig. 1.

The spatial variation of Malaria in Kenya is shown in the maps. These effects are represented geographically as the map of Kenya is segmented into 47 counties. Each map represents each predictor variable under study. These variables are Age, wealth index, education level, Has Mosquitoes nets and marital status. From the first map on wealth index, which shows that as darker the color the lower the wealth index, it is clear that as we increase the wealth bracket the cases of malaria decreases, this may be because the wealthy have ways of prevention of mosquito into human contact either by using mosquito nets of other ways [13]. Counties on the lower index include Samburu, Baringo, Taita-Taveta. Consequently, these counties are experiencing a surge in malaria cases, in more wealthier counties of central Kenya like Nyeri, Kiambu, Murang'a, Meru, Tharaka-Nithi, Kirinyaga as well as the county of Nairobi, there are less cases Nairobi of malaria based on wealth distribution among inhabitants. These findings are consistent with [10] that the results indicate that the stationarity assumption is a very important factor, since it had a major influence on the significance of environmental factors and the corresponding malaria risk maps.

The second map represents the distribution of malaria in Kenya based on Age, it shows darker shade amongst lower ages, most of the counties reported high cases of Malaria among groups of 20-30 compared to the other age groups. These are the youths' age group and as such the cases increase as the age decreases. Also, those between 0-10 had more cases of malaria reported. In counties such as Mombasa, Tana River, Taita-Taveta, Tharaka Nithi, Trans-Nzoia, Kisumu, Homabay, Baringo and Kakamega. However, the severity of these cases varied significantly from one country to the other, with more severe cases being recorded around the lake region counties [19].

From the third map of education level, which shows darker shade from low education levels up to the tertiary education level, many counties reported more malaria cases among those counties where residents had minimal access to education. Fewer cases are recorded in counties registering high access to secondary and tertiary education with more cases recorded in those with primary education or none. Of such, counties in North Rift like Baringo and Turkana and those in North Eastern Kenya like Garissa, Mandera, Wajir Recorded More cases of malaria where access to education is minimal.

From the fourth map of cases despite sleeping under a mosquito net, almost the whole country had cases of people getting malaria by not sleeping under the net. But a few counties like Isiolo, Meru, Nyeri, Machakos reported cases of malaria despite sleeping under mosquito nets. Thus, in those regions where there are minimal cases of access to mosquito nets or where either deliberate or not people chose not to sleep under a treated mosquito net, there were more cases of malaria recorded. However, this variable seemed synonymous in almost all regions of the country. From the fifth map of cases on Marital status, most counties reported cases on level 4 and 5. Counties like Lamu, Makueni, Kericho had many cases of malaria among the single people.

5 Conclusion and Recommendation

The study found out that a significant relationship exists between malaria and socio-economic, demographic and environmental factors. Age, Wealth, Region, place of residence, marital status and education were significant factors in the distribution of malaria in Kenya. Some factors like risk in the rainy season and gender were not considered due to large amounts of missing data. Wealthy households reported far less cases of malaria compared to their poor counterparts. This was largely due to the fact that most of these wealthy households had better capacity and resources to manage and also prevent fatalities caused by malaria. As such, it could be attributed to better housing facilities, access to improved healthcare and living in good sanitary environments, such resources are scarce in poor households. Similarly, education levels in households were seen as another major factor. In those households recording improved access to education, secondary and tertiary institutions, that recorded lower cases of malaria compared to the illiterate households. Education plays a key role in how culture and beliefs are shaped. Those more educated are most likely to adopt new and improved practices compared to less educated households. Such households are most likely to visit health centers to manage conditions in comparison. Thus, as the literacy levels in households increase, they tend to have more ability to access information and control measures as they emerge, this has had a direct impact on lowering the cases of malaria.

Place and region of residence was another important factor that showed a significant relationship with the prevalence of malaria. People living in urban areas recorded fewer cases of malaria compared to those living in rural areas. This observation could be attributed to access to better health facilities in urban areas compared to rural areas. More so, people in urban areas have a better access to information compared to the rural dwellers. On the other hand, region of residence was an important covariate on the spread of malaria. Some regions are more prone and have high cases of malaria in comparison to the others. For instance, coastal and lake region counties recorded more cases of malaria compared to others. This could be to a large part attributed to weather patterns in these counties. They were found to have many ideal places and weather conditions that harbored the breeding of mosquitoes, thus a rise in malaria cases in these regions. However, on the other hand, some other counties like northern Kenya also recorded higher cases of malaria, this could be attributed also to the climatic patterns in those regions, Moreover, lack of access to better health infrastructure and information contributed greatly to many malaria cases in these regions. One major observation was the fact that in most of the regions experiencing high malaria cases, multiple covariates/variables are more pronounced here; some of these includes, minimal wealth distribution, poor access to education, low access to mosquito nets.

It is evident from the study that there is a spatial variation of malaria prevalence in Kenya. A number of factors contribute to this, to the large degree, it can be attributed to uneven distribution of key resources in Kenya. Those counties with improved infrastructure and resources were seen to record minimal cases of malaria. Lack of access to basic infrastructure like healthcare, improved living conditions, better economic environment were seen to largely cause this variation. Similarly, poor sensitization campaigns, diet as well as lack of control to other infectious diseases had a major role on the spread of malaria. Those regions sharing common neighborhoods were also found to have similar malaria patterns. This could be attributed to factors like wealth distribution, access to education and health facilities as well as weather patterns, which to a large degree led to higher prevalence of malaria in those regions, like lake region counties, western highlands and coastal regions.

In order to curb the menace that malaria is, there is a need to have a tailor-made approach from the government and private sector as well. This study recommends that both levels of government increase funding on malaria control initiatives. It also recommends that both levels of government improve on existing and developed new infrastructure in healthcare and sanitization. There should be an initiative to push for access to improved housing facilities. There is also a recommendation for a community level outreach programs for those in rural areas to sensitize them on the dangers of malaria and the best strategies to control and subsequent eradication of malaria. Moreover, this study recommends a custom approach, where region specific initiatives are modelled specific to each region, since a holistic approach has failed to recognize regions-specific challenges. Similarly, the study recommends that more research and resources be put in place, borrowing best practices from regions that have successfully fought the pandemic, for instance China which has recently been declared Malaria free by the Word health Organization (WHO).

The study also established a strong correlation on regions sharing common boundaries. For example, the lake region and coastal region counties. These counties have people with similar cultural practices, weather patterns and socio-economic practices. Therefore, heightened campaigns on adoption and use of treated mosquito nets in these regions as well other endemic regions should be a priority for both levels of government. Malaria control programs, availability of well-trained health personnel to help in managing the problems, improved sanitary facilities, improved wealth distribution and programs to improve literacy levels in households should be prioritized. That way, the knowledge gap between the population and health personnel will be bridged. This will cause a ripple effect overall on governments' effort to curb the prevalence and overall eradication of Malaria.

Competing Interests

Authors have declared that no competing interests exist.

References

1. Carrera LC, Victoria C, Ramirez JL, Jackman C, Calzada JE, Torres R. Study of the epidemiological behavior of malaria in the Darien Region, Panama. 2015-2017. PLoS ONE. 2019;14(11):1–30.
    Available:https://doi.org/10.1371/journal.pone.0224508
2. Robinson LJ, Wampfler R, Betuela I, Karl S, White MT, Li Wai Suen CSN, et al. Strategies for Understanding and Reducing the Plasmodium vivax and Plasmodium ovale Hypnozoite Reservoir in Papua New Guinean Children: A Randomised Placebo-Controlled Trial and Mathematical Model. PLoS Medicine. 2015;12(10).
    Available:https://doi.org/10.1371/journal.pmed.1001891
3. Desai M, Buff AM, Khagayi S, Byass P, Amek N, Van Eijk A, et al. Age-specific malaria mortality rates in the KEMRI/CDC health and demographic surveillance system in Western Kenya, 2003-2010. PLoS ONE. 2014;9(9):5–10.
    Available:https://doi.org/10.1371/journal.pone.0106197
4. Mogeni P, Williams TN, Fegan G, Nyundo C, Bauni E, Mwai K, et al. Age, Spatial, and Temporal Variations in Hospital Admissions with Malaria in Kilifi County, Kenya: A 25-Year Longitudinal Observational Study. PLoS Medicine. 2016;13(6):1–17.
    Available:https://doi.org/10.1371/journal.pmed.1002047
5. Homan T, Maire N, Hiscox A, Di Pasquale A, Kiche I, Onoka K, et al. Spatially variable risk factors for malaria in a geographically heterogeneous landscape, western Kenya: An explorative study. Malaria Journal. 2016;15(1):1–15.
    Available:https://doi.org/10.1186/s12936-015-1044-1
6. Sultana M, Sheikh N, Mahumud RA, Jahir T, Islam Z, Sarker AR. Prevalence and associated determinants of malaria parasites among Kenyan children. Tropical Medicine and Health. 2017;45(1): 1–9.
    Available:https://doi.org/10.1186/s41182-017-0066-5
7. Were V, Buff AM, Desai M, Kariuki S, Samuels AM, Phillips-Howard P, et al. Trends in malaria prevalence and health related socioeconomic inequality in rural western Kenya: Results from repeated household malaria cross-sectional surveys from 2006 to 2013. BMJ Open. 2019;9(9):1–9.
    Available:https://doi.org/10.1136/bmjopen-2019-033883
8. Choi L, McIntyre S, Furnival-Adams J. Indoor residual spraying for preventing malaria. Cochrane Database of Systematic Reviews. 2019;(3).
    Available:https://doi.org/10.1002/14651858.CD013300
9. Zhou G, Lee MC, Githeko AK, Atieli HE, Yan G. Insecticide-Treated Net Campaign and Malaria Transmission in Western Kenya: 2003–2015. Frontiers in Public Health. 2016;4:1–8.
    Available:https://doi.org/10.3389/fpubh.2016.00153
10. Gosoniu L, Vounatsou P, Sogoba N, Smith T. Bayesian modelling of geostatistical malaria risk data. Geospatial Health. 2006;1(1):127–139.
    Available:https://doi.org/10.4081/gh.2006.287
11. Khagayi S, Desai M, Amek N, Were V, Onyango ED, Odero C, Otieno K, et al. Modelling the relationship between malaria prevalence as a measure of transmission and mortality across age groups. Malaria Journal. 2019;18(1):1–12.
    Available:https://doi.org/10.1186/s12936-019-2869-9
12. Tuyishimire J. Spatial Modelling of Malaria Risk Factors in Ruhuha. Rwanda Journal Series - Life and Natural Sciences. 2016;1:17.
    Available:https://www.ajol.info/index.php/rj/article/view/153704/143292%5Cnhttps://www.ajol.info/index.php/rj/article/view/153704
13. Ssempiira J, Nambuusi B, Kissa J, Agaba B, Makumbi F, Kasasa S, Vounatsou P. Geostatistical modelling of malaria indicator survey data to assess the effects of interventions on the geographical distribution of malaria prevalence in children less than 5 years in Uganda. PLoS ONE. 2017;12(4):1–20.
    Available:https://doi.org/10.1371/journal.pone.0174948
14. Hasyim H, Nursafingi A, Haque U, Montag D, Groneberg DA, Dhimal M, Kuch U, Müller R. Spatial modelling of malaria cases associated with environmental factors in South Sumatra, Indonesia. Malaria Journal. 2018;17(1):1–15.
    Available:https://doi.org/10.1186/s12936-018-2230-8

15. Sinka ME, Golding N, Massey NC, Wiebe A, Huang Z, Hay SI, Moyes CL. Modelling the relative abundance of the primary African vectors of malaria before and after the implementation of indoor, insecticide-based vector control. Malaria Journal. 2016;15(1):1–10.
Available:https://doi.org/10.1186/s12936-016-1187-8

16. Ngesa O, Mwambi H, Achia T. Bayesian spatial semi-parametric modeling of HIV variation in Kenya. PLoS ONE. 2014;9(7).
Available:https://doi.org/10.1371/journal.pone.0103299

17. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society. Series B: Statistical Methodology. 2009;71(2):319–392.
Available:https://doi.org/10.1111/j.1467-9868.2008.00700.x

18. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. Statistical Science. 1996;11(2): 89–102.
Available:https://doi.org/10.1214/ss/1038425655

19. Jenkins R, Omollo R, Ongecha M, Sifuna P, Othieno C, Ongeri L, Kingora J, Ogutu B. Prevalence of malaria parasites in adults and its determinants in malaria endemic area of Kisumu County, Kenya. Malaria Journal. 2015;14(1):1–6.
Available:https://doi.org/10.1186/s12936-015-0781-5