



Application of logistic regression model on the spread of malaria infection in Calabar municipality (A case study of university of Calabar teaching hospital)

Deedumbari Charles Gboh-Igbara, Bassey Edet Effiong*, Sebastian Stephen Akpan, Ugbe Thomas Adidaumbe, Enang Ekaette Inyang, Christian Elendu Onwuke, Joy Uket Ogum and Nja Egom Mbe

Department of Mathematics and Statistics, Faculty of Physical Sciences, University of Calabar, Calabar Nigeria.

Open Access Research Journal of Multidisciplinary Studies, 2024, 07(01), 022–038

Publication history: Received on 20 November 2023; revised on 23 January 2024; accepted on 26 January 2024

Article DOI: <https://doi.org/10.53022/oarjms.2024.7.1.0055>

Abstract

Background: The malaria disease is the outcome of the interaction among three elements which includes; man, mosquito and the parasite. The intensity of the disease is being regulated by the physical and socio-economic determinant in the area which interact with these elements. The physical conditions of the region determine the growth and proliferation of mosquito and parasite, while the socio-economic conditions of the people determine the distribution of mosquito and parasites. This piece of research work has been devoted to the study of vector species (*Plasmodium*), spatial-temporal incidence pattern of malaria i.e parasite load, physical and socio-economic determinants responsible for the spread of mosquito and parasites, control measures and risk factor assessment. The favourable explanatory variables considered in the prediction of malaria prevalence such as age group, gender, blood group and genotype presents suitable conditions to determine the validity of malaria prevalence across the metropolis which as well substantially contributed and facilitate the growth and diffusion of malarial incidence in Calabar Municipality. The data obtained was initially entered in Microsoft Excel (2016) and checked for errors after which it was exported to IBM SPSS Statistics 23 software for logistic regression analysis. The metropolis records high incidence of malaria. The predominant parasite *plasmodium falciparum* is considered vital for causing considerable morbidity and mortality in the area. From our analysis, it is observed that fifty nine of our patients were malaria negative and one hundred and one patients were malaria positive. The model predicted in respect to gender that seventy four percent of female population were malaria positive likewise the male gender predicted at eighty percent to be malaria positive.

Keywords: Malaria; Calabar Municipality; *Plasmodium vivax*; *Plasmodium falciparum*; Remote sensing; Spatial analysis; Epidemiology

1. Introduction

The piece of research work has been devoted to the study of vector species (*plasmodium*), spatial-temporal incidence pattern of malaria i.e. parasite load, physical and socio-economic determinants responsible for the spread of mosquito and parasites, control measures and risk factor assessment. The favourable explanatory variables considered in the prediction of malaria prevalence such as age group, gender, blood group and genotype presents suitable conditions to determine the validity of malaria prevalence across the calabar metropolis which as well substantially contributed and facilitates the growth and diffusion of malarial incidence in Calabar Municipality.

Due to data availability, the first four wards of Calabar Municipality were the object of this study and the data used in this study were obtained as secondary data from impatient morbidity and mortality returns register at University of Calabar Teaching Hospital (UCTH), from 1st January, 2008 to 31st December, 2018.

The main objective of this study is to predict malaria infection using gender, age, hematological and other socio_economic parameters. To achieve this goal, a logistic regression model was employed. Therefore, the aim of this

* Corresponding author: Edet Effiong Bassey

study was to quantify the variation in the prevalence of malaria between sample enumeration areas (SEAs) or clusters, the effects of cluster characteristics on the prevalence of malaria using the intra-class correlation coefficient as well as to identify significant factors that affect the prevalence of malaria using the multilevel logistic regression modelling in first four wards of Calabar Municipality.

Assessment of the pattern of the current malaria prevalence and understanding how malaria varies in the community as a result of seasonal, environmental, geographical or year-to-year changes will help to evaluate the effectiveness of proven control interventions of the disease in a locality. So, it is crucial to assess the contributory factors, a study was done by Adigun, Abbas B., *et al.* conforms to other studies where the findings show that malaria is a significant concern in the country. According to Adigun, Abbas, *et al.*, Egidi Ndamuzi *et al.* (2021) malaria symbolizes a significant public health challenge in Nigeria and is a crucial cause of mortality and morbidity. Additionally (Herdicho K. *et al.*, 2021), it gives important insight into the changing malaria situation, which might guide adjustments of malaria program activities and the prioritization of malaria research and the changing malaria situation which requires an updating description of malaria trends. Hence, Mathematical models are tools that have been successfully used to understand the transmission of infectious diseases (Collins & Duffy, 2020; Collins & Duffy, 2021; Collins & Govinder, 2014; Ibrahim & Dénes, 2021; Ojo *et al.*; Koella & Antia, 2003; Tien & Earn, 2010; Tumwiine *et al.*, 2014).

(Fahad A. *et al.* 2023) in his work, proposed a mathematical modelling of the transmission dynamics of malaria which offer a better idea of the disease's spread and impact. It helps prepare for the future and inform appropriate policy making to control the disease. However, these studies did not reflect the impact of awareness movements on controlling malaria. Awareness movements provide substantial tools for controlling the spread of malaria.

In the past, several mathematical models of the transmission dynamics of malaria following the simple S–I–R model were published. Many researchers have modified these models by incorporating more ideas associated with malaria dynamics and possible control of the disease.

The Mathematical modeling of malaria of his work originated from the works of Ross and MacDonald who did some modification to the Ross's model and included superinfection. This has allowed the scientific community to refine these models until today. In order to model the disease, compartmental models of malaria and differential equations are constructed. Despite the success of his research, he recommended the need of mosquito nets, insecticide utility and other strategies that can reduce these parameters, which will be considered explicitly and in details in this research.

Though, [Misra *et al.* 2021] proposed a mathematical model to measure the impact that awareness social media campaigns have on vector-borne diseases, considering a constant disease transmission rate. They divided the human population into three sub-populations, namely susceptible, infected, and aware-people. In addition, a dispersed population representing the number of media campaigns, was used to measure the importance of the media campaigns. However, these efforts have been unsuccessful in controlling the spread of malaria. Therefore, more model-based research on malaria dynamics and studies on the influence of awareness campaigns are needed (Agusto *et al.*, 2013; Tasman, 2013; Kim *et al.*, 2019; Okuneye & Gumel, 2017; Abiodun *et al.*, 2018; Tumwiine *et al.*, 2014; Herdicho *et al.*, 2021; Tasman, 2015; Koella & Antia, 2003).

Fitting a mathematical model to real data on a particular disease and subsequently estimating important parameters from the real data can enable researchers to make predictions on the future dynamics of the disease (Kim *et al.*, 2020; Mojeeb *et al.*, 2019; Zhao *et al.*, 2022).

There is no doubt that those studies have made valuable input in understanding the dynamics and control of malaria. However, as far as we are aware, none of those studies use a mathematical deterministic model that incorporates drug resistant, control measures together with real data to study and make predictions of the possible future dynamics of malaria in Nigeria. This study aims to fill this gap in the literature. The findings from this study are expected to aid both researchers and policy makers in developing better control strategies for effective control and management of malaria outbreaks.

1.1. Basic definition of terms

Y – intercept: can be interpreted as the log-odds when the predictors are all zero, i.e. if the function is linear and is expressed in slope - intercept form

$$F(x) = \hat{a} + bx$$

X – intercept: is interpreted as the weighted predictors corresponding to even odds.

$\beta_1 = 1$ means that increasing X_1 by 1 increases the log-odds by 1.1, so it multiplies the odds by $10^1 = 10$; sometimes referred to as the effect of the predictor X_1 .

$\beta_1 = 2$ means that increasing X_2 by 1 increases the log-odds by 2.1 so it multiplies the odds by $10^2 = 100$, thus the effect of X_2 on the log-odds is twice as great as the effect of X_1 .

The test statistic $-2L_1/L_0$ for a nested model asymptotically will be Chi-squared (X^2) distributed with degrees of freedom equal to the difference in dimensionality of L_0 and L_1 .

Log-odds: In Statistics, logit function or log-odds is the logarithm of the odds $P/1-P$ where p is the probability value from $[0,1]$ to $[-\infty, +\infty]$.

Odds Ratio: is a Statistic defined as the ratio of the odds of A in the presence of B and the odds of A without the presence of B. This Statistic attempts to quantify the strength of the association between A and B.

Cluster Random Sampling: is a way to randomly select participant when they are geographically spread out. For example, if we want to choose 100 participants from the entire population in Nigeria, it is likely impossible to get a complete list of everyone. Instead, the researcher randomly selects areas (i.e. cities/countries) and randomly selects from within those boundaries.

Probability Sampling: is a sampling techniques in which samples from a larger population are chosen using a method based on the theory of probability. For a participant to be considered as a probability sample, he/she must be selected using a random selection.

Logit Model: is an approach for use, when the dependent variable in Logistic regression model is binary (0 or 1) and which ensures that the estimated probabilities are bounded by 0 and 1.

Variable: is any characteristic, number or quantity that can be measured or counted. A variable may also be called a data item. Age, sex, business income and expenses, country of birth, capital expenditure, class grades, eye colour and vehicle type are examples of variables.

Data: Data can be defined as the collection of facts or information from which conclusions may be drawn. Data means information, we typically use the term to refer to numerical files that are created and organized for Analysis.

Dummy Variable: a dummy variable is a dichotomous (a separation or division into two) variables which has been coded to represent a variable with a higher level of measurement.

Dummy Variable Coding or Dummy Coding: refers to the process of coding a categorical variable into dichotomous variables.

Logit Transformation: Is a common transformation for linearizing sigmoid distribution (logic normal distribution of proportions). The logit is defined as the natural log $\ln(p/1-p)$, where p is a proportion. This transformation is referred to as logit or logistic transformation – which solves the issues of the 0/1 boundaries for the original dependent variable (probability).

Explanatory Variables: Is used to predict or explain differences in the response variable. In an experimental study, the explanatory variable is the variable that is manipulated by the researcher. It is also known as the independent or predictor variables.

Binary Dependent Variable or Response Variable: It is also known as the outcome or dependent variable, its value is predicted or its value is predicted or its variation is explained by the explanatory variable. The response variable is simply a designation of two outcomes (a binary response). Examples – dead or alive, success or failure, approved or denied, etc.

Statistically significant: In Statistical hypothesis testing, a result has statistical significance when it is very unlikely to have occurred given the null hypothesis. Statistically significant is the likelihood that a relationship between two or more variables is caused by something other than chance.

P-value: The P-value or Probability value or asymptotic significance is the probability for a given statistical model that, when the null hypothesis is true, the statistical summary (such as the sample mean difference between two compared groups) would be greater than or equal to the actual observed results.

2. Material and methods

2.1. Data collection and analysis

The data used in this work was a secondary data collected from University of Calabar Teaching Hospital (UCTH). All data were recorded and analyzed using SPSS (version 23) Statistical Software. Categorical data were compared using Pearson's chi-square test set at 5% of significant level. To fulfill the objective we apply logistic regression model by calculating the odds ratios as measures of association including other tests such as Hosmer-Lemeshow tests, Cox and Snell R-squared and Pearson Chi-squared test.

Logistic Regression analysis was also employed to model the presence (positive) or absence (negative) of malaria in a patient.

2.2. Data analysis and spatial modelling

2.2.1. Logistic regression

The model that represents the likelihood of infection by malaria at the Calabar Municipality settlement was generated based on Logistic Regression with the 'backward conditional stepwise' procedure, comparing cases/noncases with multiple explanatory variables. It was opted for a dichotomous modeling approach, as no reliable Statistics on the total number of inhabitants domicile could be obtained, which would a pre-requisite of bias-free estimate of absolute cases or its probability (eg. by Poisson regression or generalized linear models). Model performance for different cut-off values was assessed by its sensitivity and specificity (ROC curve). In logistic regression the canonical link function (logits) for the binomial distribution, of the unknown binomial probabilities are modelled as a linear function of the risk factors (x_i):

$$g(P_i) = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i$$

In which:

$g(P_i)$ = link function

P_i = likelihood of response for the i -th factor (or covariate)

β_0 = intercept

β_i = coefficient

x_i = independent variables

Thus, the logit of $P(x)$ simplifies the sum. The quantity $P(x)$ divided by $1-P(x)$, whose log value gives the logit, describe the odds or odds ratio for a malaria patient being dead, with independent variables specified by x .

$$P(x)/1-P(x) = \text{odds for individual } x$$

The model included variables such as blood groups, environment, genotype, age groups, gender, stagnant water and clear bushes, which is expressed as;

$$\text{Logit } P(x_i) = \beta_0 + \beta_1 \text{Genotype}_i + \beta_2 \text{Gender}_{\text{males}} + \beta_3 \text{Gender}_{\text{females}} + \beta_4 \text{Age Group}_{15-24} + \beta_5 \text{Age Group}_{24-30} + \beta_6 \text{Age Group}_{>40} + \beta_7 \text{Blood Groups}_i + \beta_8 \text{Stagnant waters}_i + \beta_9 \text{Clear Bushes}_i + \beta_{10} \text{Use of Insecticidal nets}_i$$

The method of including variables in the model can be carried out in a stepwise manner going forward or backward, testing for the significance of inclusion or elimination of the variable at each stage. The tools are based on the change in likelihood resulting from including or excluding the variable or using the p-value or other test statistic.

2.2.2. Odds Ratio (OR)

Since this study considered several predictors in deciding the eventual outcome, the joint effect of all the predictors (independent variables) put together would be mathematically expressed by:

$$\text{Odds} = p/(1-p) = e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}$$

But in logistic regression model, e^β is the change in the odds ratio of a success at level of the explanatory variable one unit apart.

The odds of the event occurring is defined as the ratio of probability of success to probability of failure as follows:

$$\begin{aligned} O &= p/1-p \\ O(x) &= p(x)/1-p(x) \\ &= 1/1+e^{-(\alpha+\beta x)} \div 1-1/1+e^{-(\alpha+\beta x)} \\ &= 1/1+e^{-(\alpha+\beta x)} \div 1+e^{-(\alpha+\beta x)} / 1+e^{-(\alpha+\beta x)} \\ &= 1/1+e^{-(\alpha+\beta x)} \times 1+e^{-(\alpha+\beta x)} / 1+e^{-(\alpha+\beta x)} \\ &= 1/1+e^{-(\alpha+\beta x)} = e^{\alpha+\beta x} \end{aligned}$$

Similarly,

$$O(x+1) = e^{\alpha+\beta(x+1)} = e^{\alpha+\beta x} e^\beta$$

Then the ratio of the odds at $x+1$ to the odds at x (the odds ratio) can be written as:

$$\begin{aligned} OR(x+1,x) &= O(x+1)/O(x) \\ &= e^{\alpha+\beta x} e^\beta / e^{\alpha+\beta x} = e^\beta \end{aligned}$$

2.2.3. Test and Confidence intervals for the parameters

The Wald Statistic

A Wald test is used to test the statistical significance of each coefficient in the model. A Wald test calculated a Z Statistic, which is;

$$Z = \beta/SE$$

This Z value is then squared, yielding a Wald Statistic with a Chi-square distribution as follows;

$$Z^2 = [\beta/SE]^2$$

The procedure typically used to obtain a large sample of confidence interval for the parameter by computing the estimate of the parameter plus or minus a percentage point of the normal distribution multiplied by the estimated standard error, such that

$$100(1-\alpha)\% \text{ CI for } \beta_i$$

$\beta_i \pm Z_{1-\alpha/2} S_{\beta_i}$ where β_i and S_{β_i} are obtained from printout and Z from N(0,1) tables.

CI for Odd Ratio; exp (CI for β)

Thus, if we consider model, and X_i denotes a (0,1) exposure variable of interest than a 95% confidence interval for the adjusted odds ratio is given by exp ($\beta \pm 1.96 S_{\beta_i}$)

Likelihood-Ratio Test

The likelihood ratio, test for a particular parameter compares the likelihood of obtaining the data when the parameter is zero (l_0) with the likelihood (l_1) of obtaining the data evaluated at the Maximum Likelihood Estimate (MLE) of the parameter. The test statistic is calculated as follows;

$$-2x \ln(\text{Likelihood ratio}) = -2 X (\ln L_0 - \ln L_1)$$

It is compared with a X^2 distribution with 1 degree of freedom.

Goodness of fit of the model

The goodness of fit or calibration of a model measures how well the model describes the response variable. Assessing goodness of fit involves investigating how close values predicted by the model are to the observed values. For example, support the X^2 Statistic is 2.68 with $9-2 = 7$ degrees of freedom, giving $P=0.91$ suggest that the numbers of the deaths have no significantly different with the predicted model.

The X^2 test statistic is given by:

$$X^2 = (\text{Observed} - \text{Expected})^2 / \text{Expected}$$

Cox and Snell Nagelkerke R-squared

This is also known as Pseudo R-squared, which is mathematically expressed by;

$$R^2 = 1 - \{L(M_{\text{intercept}}) / L(M_{\text{full}})\}^{2/N}$$

Such that,

$L(M)$ is the conditional probability of the dependent variable given the independent variables. If there are N observations in the dataset, then n th root of the product $L(M)$ provides an estimates of the likelihood of each y -value.

Thus;

$$R^2 = 1 - \{L(M_{\text{intercept}}) / L(M_{\text{full}})\}^{2/N} \div 1 - L(M_{\text{intercept}})^{2/N}$$

To achieve this, the Cox and Snell R-squared is divided by it's maximum possible value, $1 - L(M_{\text{intercept}})^{2/N}$

Then, if the model perfectly predicts the outcome and has a likelihood of 1, Nagelkerke/Cragg & Uhler's

R-squared = 1. When $L(M_{\text{Full}}) = 1$, then

$R^2 = 1$; when $L(M_{\text{Full}}) = L(M_{\text{intercept}})$, then $R^2 = 0$

The Hosmer-Lemeshow Test

The Hosmer-Lemeshow test (HL test) is a goodness of fit test for logistic regression, especially for risk prediction models. A goodness of fit test tells you how well your data fits the model. Specifically, the HL test calculates if the observed event rates match the expected event rates in population subgroups.

The Hosmer-Lemeshow test statistic is calculated with the following formula (which is for the 10-group case—modify for your specific number of groups):

$$\begin{aligned}
 & \mathbf{G} \\
 \mathbf{H} &= \sum_{\mathbf{g=1}}^{\mathbf{G}} ((\mathbf{O}_{1g} - \mathbf{E}_{1g})^2 / \mathbf{E}_{1g} + (\mathbf{O}_{0g})^2 / \mathbf{E}_{0g}) \\
 & \mathbf{g=1} \\
 & \mathbf{G} \\
 &= \sum_{\mathbf{g=1}}^{\mathbf{G}} ((\mathbf{O}_{1g} - \mathbf{E}_{1g})^2 / \mathbf{N}_g \pi_g + (\mathbf{O}_{0g})^2 / \mathbf{N}_g \pi_g) \\
 & \mathbf{g=1}
 \end{aligned}$$

Where:

X^2 = chi squared.

N_g = number of observations in the gth group.

O_{1g} = number of observed cases in the gth group.

E_{1g} = number of expected cases in the gth group.

Σ = summation notation. For the above formula, we're summing from 1 to 10. Modify the summation for your number of groups.

Here: O_{1g} , E_{1g} , O_{0g} , N_g and π_g denote the

Observed $y = 1$ events, expected $y = 1$ events, Observed $y = 0$ events, Expected $y = 0$ events, total observations, predicted risk for gth risk decile group, and G is the number of groups. The test statistic asymptotically follows a X^2 distribution with $G-2$ degrees of freedom.

3. Result and Discussion

This research work session deals with the presentation of collected data, analysis of the data and the interpretation of the data using SPSS Statistical Software Package.

A total number of 160 samples ($N=160$) were selected from the Inpatient morbidity and mortality returns register at University of Calabar Teaching Hospital (UCTH), from 1st January, 2008 to 31st December, 2018. The following frequency tables shows the distribution of data of patients according to gender, age, blood group, genotype etc, as obtained from SPSS as shown below:

3.1. Bivariate and multivariable analysis of factors associated with malaria infection concerning gender, age group, blood group and regional variation

3.1.1. Malaria Prevalence by Gender among patients ($n = 160$)

The Gender distribution among the patients in table 1 below shows that the proportion of males was 102 compared to the females of population size of 58. 102 male patients representing 63.8% shows that malaria prevalence is on the increase as compared to the female representing 36.3%. Collectively, these data suggest that sex-specific differences may exist in naturally acquired immunity to malaria, with males being less able to control parasite densities (anti-parasite immunity), leading to higher malaria parasite prevalence among males, and females being less able to tolerate higher parasite densities.

Table 1 Frequency Table showing the distribution of patients according to gender ($n=160$)

	Gender of Patient			
	Frequency	Percent	Valid Percent	Cumulative Percent
Male	102	63.8	63.8	63.8
Female	58	36.3	36.3	100.0
Total	160	100.0	100.0	

3.1.2. Malaria Prevalence by Age Group among patients undergoing blood film examination at UCTH ($n=160$)

In Table 2 as shown below, indicates that the Age Group also plays a major role in malaria prevalence. Patients of population size of 43 within age group (30–39) shows high level of malaria dominance at 26.9% which is 10 times higher as compare to patients within the age brackets (below 9) and (70–80) representing 1.3% which is far lesser. Malaria Prevalence among age group (30–39) may be attributed to poor housing conditions and low socio-economic status across the population.

Table 2 Frequency Table showing the distribution of patients according to age group

	Age group			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Below 9	2	1.3	1.3	1.3
10-19	24	15.0	15.0	16.3
20-29	20	12.5	12.5	28.8
30-39	43	26.9	26.9	55.6
40-49	15	9.4	9.4	65.0
50-59	33	20.6	20.6	85.6
60-69	21	13.1	13.1	98.8
70-80	2	1.3	1.3	100.0
Total	160	100.0	100.0	

3.1.3. Malaria Prevalence Concerning Blood Group (n=160)

The Table 3 as shown below, indicates that the chance of developing complicated falciparum malaria was least in blood group O at 5.0% compared to blood groups A, B and AB. Antigens of blood groups A and B have been suggested to play important roles in cytoadherence. Due to the absence of A and B antigens on the surface of blood group O erythrocytes, cytoadherence, and hence rosetting and sequestration, is reduced in individuals with blood group O. It has been observed that blood group O individuals are less likely to suffer from complicated falciparum malaria. In this study, we observed that only 8 (5.0%) of the 160 participants with complicated disease had blood group O. The low parasitaemia observed in this study, together with reduced rosetting and cytoadherence observed by others, may give the blood group O individuals suffering from falciparum malaria a good prognosis compared with those with other blood groups.

Table 3 Frequency Table showing the distribution of patients according to blood group

	Blood group			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
O	8	5.0	5.0	5.0
AB+	37	23.1	23.1	28.1
AB-	12	7.5	7.5	35.6
A+	45	28.1	28.1	63.8
A-	32	20.0	20.0	83.8
B+	26	16.3	16.3	100.0
Total	160	100.0	100.0	

3.1.4. Malaria Prevalence by blood group (n=160)

From the frequency distribution table 4 below, indicates that patients with genotype AA had a prevalence of malaria at 48.10%; a prevalence that was 10 times higher than patients with genotype OO with lower prevalence at 5.0%. The malaria parasite (*Plasmodium falciparum*) has a high rate of oxygen consumption and utilizes large amounts of haemoglobin (the oxygen-carrying pigment and predominant protein in the red blood cells) during the blood stage of replication and duplication. This thus makes the malaria parasite thrive better in AA genotype since genotype AA red blood cells contain normal haemoglobin both in quantity and structure. Hence, it is important to avoid mosquito bites

by taking necessary precautionary measures of getting rid of their breeding sites, especially as it regards to the present weather condition.

Table 4 Frequency Table showing the distribution of patients according to Genotype

	Genotype			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
OO	8	5.0	5.0	5.0
AA	77	48.1	48.1	53.1
BB	26	16.3	16.3	69.4
AB	49	30.6	30.6	100.0
Total	160	100.0	100.0	

3.1.5. Regional variation in prevalence of malaria among patients across Calabar metropolis (n = 160)

Cases of malaria prevalence as shown in Table 5 below was much higher in Ward 2 at 36.3% compared to Ward 3 with lower prevalence at 15%. Larger distances from water bodies correlates with lower malaria prevalence. Water bodies are breeding sites, and mosquitoes do not move away from the breeding site when they can feed. Human settlements closer to the water would be exposed to more mosquitoes than those further away. Furthermore, poor housing conditions within this geographic space and low socio-economic status across the population are all contributory factors in consideration for malaria prevalence.

Table 5 Frequency Table showing the distribution of patients according to Location of residence

	Location of wards in calabar municipality			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Ward 1 (Goldie/Mayne Ave., Govt Primary School Akim, Federal training centre)	47	29.4	29.4	29.4
Ward 2 (Atimbo/Edim Otop, Malabor Male/Female Hostel, Old Library Unical)	58	36.3	36.3	65.6
Ward 3 (Navy Town, Ediba Health Centre)	24	15.0	15.0	80.6
Ward 5 (Essien town, Ekorinim village, Navy base)	31	9.4	9.4	100.0
Total	160.0	100.0	100.0	

3.2. Bivariate and multivariable analysis of factors associated with malaria infection concerning ITN availability, ITN usage, presence of stagnant water, indoor residual spray, history of malaria treatment, and residence at UCTH

3.2.1. Malaria Prevalence Among Patients Concerning Insecticide Treated Net (ITN) Usage

On Frequency Table 6 below, shows that 45.6% of households owned at least one insecticide treated net (ITN). The ownership of nets was 10% higher in the Municipality than in Sub-urban areas of the town. Majority of the households in the Sub-urban do not owned insecticide treated nets at 54.4%, unlike the Municipality with an ownership rate of 45.6%. An inverse relationship between wealth quintile and ownership of nets was also observed. The lowest wealth quintile had an ownership rate of 86%, and households in the highest wealth quintile had an ownership rate of 58%. Though, there has been a scale up of efforts to prevent malaria transmission in Calabar over the years and effective

malaria prevention methods identified are the use of insecticide treated mosquito nets and indoor residual spraying (IRS).

Table 6 Frequency Table showing the responses of patients on the use of long lasting insecticidal nets

	Use of long lasting insecticidal nets			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
No	87	54.4	54.4	54.4
Yes	73	45.6	45.6	100.0
Total	160	100.0	100.0	

3.2.2. Malaria Prevalence Concerning Removal of Shrubs and Bushes

Clearing of bushes and pruning of flowers and shrub within the calabar metropolis decreased the local mosquito vector population by nearly 62.50% as shown on Table 7. It was observed that areas where they removed the bushes shows a drop of mosquito numbers collected falls from an average of 11 to 4.5 for females, and 6 to 0.7 for male mosquitoes. The total number of mosquitoes across these metropolis decreased by nearly 60% after removal of bushes. After bush removals, the number of older more dangerous vector females in the population dropped to levels similar to those recorded in areas that had no presence of bushes and shrubs.

Table 7 Frequency Table showing the responses of patients on clearing bushes

	Clear bushes			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
No	100	62.5	62.5	62.5
Yes	60	37.5	37.5	100.0
Total	160	100.0	100.0	

3.2.3. Malaria Prevalence Concerning Removal of Stagnant Water among Patients.

In Table 8 below, indicates a valid percent at 65% which indicates that 104 patients practice malaria control measures of removing stagnant water from within there surroundings as compared to 35% of the patients that do not practice it. As a result of this practice, dependent variable such as the larval habitats may be destroyed by filling depressions that collect water, by draining swamps or by ditching marshy areas to remove standing water. Hence, mosquitoes that breed in irrigation water can be controlled through careful water management practices over time.

Table 8 Frequency Table showing the responses of patients on the removal of stagnant water

	Removal of stagnant water			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
No	56	35.0	35.0	35.0
Yes	104	65.0	65.0	100.0
Total	160	100.0	100.0	

3.2.4. Malaria Prevalence Concerning Indoor Residual Spraying (IRS) Utility among Patients

Table 9 shows that the population 45% of households were protected by Indoor residual Spraying compared to 55% of the household that do without the indoor residual spraying. Indoor residual spraying (IRS) is a core vector control

intervention that can rapidly reduce malaria transmission as it involves the application of a residual insecticide to internal walls and ceilings of housing structures where malaria vectors may come into contact with the insecticide.

Table 9 Frequency Table showing the responses of patients on the use of indoor residual spraying

	Use of indoor residual spraying			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
No	88	55.0	55.0	55.0
Yes	72	45.0	45.0	100.0
Total	160	100.0	100.0	

3.2.5. Malaria Prevalence Concerning Climate Conditions

In Table 10 below, indicates that malaria incidence shows a clear seasonal pattern with 65–70 % of all new malaria cases occurring between August and January, with a peak in October corresponding with the end of the rainy season. The incidence was very low from November to March (five to 40 per 1000) but never reached zero. Seasonal transmission of Plasmodium parasites, that is, low transmission during the dry season and higher transmission during the wet season, is common in malaria endemic in Calabar Municipality. During the dry season, the transmission rate is very low, in some areas it is near zero. Malaria is more common during the rainy season—this is because waterlogged and damp places provide suitable breeding environments for mosquitoes.

Table 10 Frequency Table showing the distribution of patients according to the season of the year when admitted

	Season of the year when admitted			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
wet season from April to October	101	63.1	63.1	63.1
Dry season from November to march	59	36.9	36.9	100.0
Total	160	100.0	100.0	

3.2.6. Malaria Prevalence Among Registered Patients from Data Collected (n=160)

Of the total (160) febrile patients, 36.9% were malaria negative and 63.1% (overall prevalence) were malaria confirmed cases. Most of the infections were caused by Plasmodium falciparum (72.5%) followed by Plasmodium vivax (23.7%) and mixed-species (3.8%). This is shown on Table 1.11 as it concern malaria presence.

Table 11 Frequency Table showing the presence of malaria In patient

	Presence of malaria In patient			
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Negative	59	36.9	36.9	36.9
Positive	101	63.1	63.1	100.0
Total	160	100.0	100.0	

4. Data analysis and interpretation

4.1. logistics regression

Logistics regression was employed to predict the spread of malaria over a given population in study (Calabar Municipality). The predictor variables were patient's gender, age group, blood group and genotype.

4.1.1. Case Processing Summary

Table 12 Classification Test for Final Model—displays the percentage of classification for the final model.

Unweighted Cases ^a	N	Percent
Included in Analysis Missing Cases	160	100.0
Selected Case	0	0
Total	160	100.0
Unselected Cases	0	0
Total	160	100.0

a. If weight is in effect, see classification table for the total number of cases.

4.1.2. Dependent Variable Encoding

Table 13 Type III tests for fixed effects

Original Value	Internal Value
Negative	0
Positive	1

4.2. Block 0: Beginning Block

Table 14 Classification table with only the intercept which displays the baseline classification for the model with only the intercept with a predictive ability of 63%.

Classification Table ^{a,b}			
Observed	Predicted		
	Presence of Malaria in the Patient		Percentage Correct
	Negative	Positive	
Negative Presence of Malaria in the Patient	0	59	0
Step 0 Positive	0	101	100.0
Overall Percentage			63.1

a. Constant is included in the model; b. The cut value is .500

The Block 0 output is for a model that includes the intercept (which SPSS calls the constant). Given the base rates of two decision options ($59/160 \times 100 = 37\%$ were malaria negative, while 63% were malaria positive), and no other information, the best strategy is to predict, for every case, that the malaria will continue to spread.

4.2.1. Variables in the Equation

The variables in the equation used, shows that the only intercept model is $\ln(\text{odds}) = 0.538$. If we exponentiate both sides of this expression, we find our predicted odds $[\text{Exp}(B)] = 1.712$. That is the predicted odds of presence of malaria

in a patient is 1.712. Since 59 of our patients were malaria negative and 101 patients were malaria positive, our observed odds are $101/59 = 1.712$.

Table 15 Parameters estimates of odds ratio for the interaction effects

	B	S.E.	Wald	Df	Sig.	Exp.(B)
Step 0 Constant	0.538	0.164	10.763	1	0.001	1.712

Block 1: Method= Enter

4.3. Omnibus Tests of Model Coefficients

Table 16 Parameters estimates of odds ratio for the main effects

		Chi-square	df	Sig.
Step 1	Step	16.842	4	0.002
	Block	16.842	4	0.002
	Model	16.842	4	0.002

Omnibus Tests of Models gives us a Chi-square of 16.842 on 4 df, significant beyond.001

Table 17 Cox & Snell R-square and odds ratio of predictor variables in the final model

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	193.809 ^a	0.100	0.137

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

The table above contains the Cox & Snell R-square and Nagelkerke R-square values which are both methods for calculating the explained variation. Therefore, the explained variation in the dependent variable based on our model ranges from 10% to 13%.

Under the model summary, we see that -2 log likelihood statistics is 193.809. This statistics measures how poorly the model predicts the decisions, the smaller the statistics the better the model. Adding more predicted variables reduces the -2 log likelihood statistics which makes it fit.

4.4. Hosmer and Lemeshow test

Table 18 Hosmer and Lemeshow Test of significance and odds ratio of predictor variables in the initial model

Step	Chi-square	df	Sig.
1	42.549	8	0.000

4.5. Contingency Table for Hosmer and Lemeshow Test

The Hosmer-Lemeshow tests the null hypothesis that predictions made by the model fit perfectly with observed group. Cases are arranged in order by their predicted probability on the criterion variable. These ordered cases are then divided into ten (usually) groups of equal and near equal size ordered with respect to the predicted probability of the target population. For each of these groups we then obtain the predicted group membership and the actual group. This result in a 2×10 contingency Table as shown above. A Chi-square statistic is computed comparing the observed frequencies with those expected under the linear model. A non-significant chi-square indicates that the data fit the model well.

Table 19 Hosmer and Lemeshow Test of significance and odds ratio of predictor variables in the final model

		Presence of Malaria in the patient = Negative		Presence of Malaria in the patient = Positive		Total
		Observed	Expected	Observed	Expected	
Step 1	1	4	8.376	8	3.624	12
	2	6	8.680	9	6.320	15
	3	8	6.560	4	5.440	12
	4	16	7.793	1	9.207	17
	5	8	5.266	8	10.734	16
	6	6	7.057	18	16.943	24
	7	8	4.272	7	10.728	15
	8	0	4.305	16	11.695	16
	9	3	4.022	15	13.978	18
	10	0	2.670	15	12.330	15

Table 20 Summary of various tests conducted to evaluate the model

Classification Tablea				
	Observed	Predicted		
		Presence of Malaria in the Patient		Percentage Correct
		Negative	Positive	
Step 1	NegativePresence of Malaria in the Patient	27	32	45.8
	Positive	21	80	79.2
	Overall Percentage			66.9

a. The cut value is .500

Firstly, note that the table has a subscript which states, "The cut value is .500". This means that if the probability of a case being classified into the "Positive" category is greater than .500, then that particular case is classified into the positive category. Otherwise, the case is classified as in the Negative. The classification table appears to be very simple, it actually provides a lot of important information about your binomial logistics regression result, including:

- The percentage accuracy in classification (PAC), which reflects the represent of cases that can be correctly classified as "Negative" malaria prevalence with the independent variables added.
- Sensitivity, which is the percentage of cases that had the observed characteristics ("Positive" for the presence of malaria disease) which were correctly predicted by the model.

a. Variable(s) entered on step 1: Gender, Age Group, Blood Group, Genotype.

We can now use this model to predict the odds that a patient of a given gender will be malaria positive. The odds prediction equation is $ODDS = e^{a+bx}$. If our patient is a female (Genotype = 0), then the

$$ODDs = e^{1.046+0.328(0)} = e^{1.046} = 2.846$$

That is, a female is 2.846 likely to be malaria positive (than negative).

If our patient is a male (Gender = 1), then the;

$$\text{ODDS} = e^{1.046+0.328(1)} = e^{1.046+0.328} = e^{1.374} = 3.951$$

That is, a male is 3.951 times more likely to be malaria positive than negative.

We can easily convert odds to probabilities;

For Female;

$$Y = \text{ODDS}/1+\text{ODDS} = 2.846/1+2.846 = 2.846/3.846 = 0.74 \sim 74\%$$

That is the model predicts that 74% of the female will be malaria positive.

For Male;

$$Y = \text{ODDS}/1+\text{ODDS} = 3.951/1+3.951 = 3.951/4.951 = 0.798 \sim 80\%$$

The model predicts that 80% of the male will be malaria positive.

The variables in the equation output also gives us the $\text{Exp}(B)$. This is better known as the odds ratio predicted by the model. This odds ratio can be computed by raising the base of the natural log to the both power b is the slope from our logistic regression equation.

For our model $e^{0.328} = 1.388$

The odds ratio is $3.951/2.847 = 1.388$

The results of our logistic regression can be used to classify subjects with respect to what decision information to classify subjects, we need to have a decision rule. Our decision rule will take the following form.

If the probability of the event is greater than or equal to some threshold, we shall predict that the event will take place. By default, SPSS sets this threshold to 0.5. While that seems reasonable, in many cases we may need to set it higher or lower than 0.5. Using the default threshold, SPSS will classify a subject into the "higher malaria prevalence (Positive)" category if the estimated probability is 0.5 or more. SPSS will classify a subject into "Lower malaria prevalence (Negative)" category if the estimated probability is less than 0.5.

For Age Group;

The ODDS ratio is $0.853 > 0.5$ which will be considered as high spread of malaria according to age.

For Blood Group;

The ODDS ratio is $1.177 > 0.5$

For Genotype;

The ODDS ratio is $0.623 < 0.5$ which is considered as low spread of malaria according to a particular genotype.

5. Summary

The malaria disease is the outcome of the interaction among three elements i.e. man, mosquito and the parasite. The intensity of the disease is being regulated by the physical and the socio-economic determinant in the area which interact with these elements. The physical conditions of the region determine the growth and proliferation of mosquito and parasite, while the socio-economic conditions of the people determine the distribution of mosquito and parasites.

This piece of research work has been devoted to the study of vector species (Plasmodium), spatial-temporal incidence pattern of malaria i.e parasite load, physical and socio-economic determinants responsible for the spread of mosquito and parasites, control measures and risk factor assessment. The favourable explanatory variables considered in the prediction of malaria prevalence such as age group, gender, blood group and genotype presents suitable conditions to

determine the validity of malaria prevalence across the metropolis which as well substantially contributed and facilitate the growth and diffusion of malarial incidence in Calabar Municipality.

The metropolis records high incidence of malaria. The predominant parasite *Plasmodium falciparum* is considered vital for causing considerable morbidity and mortality in the area. From our analysis, it is observed that fifty nine of our patients were malaria negative and one hundred and one patients were malaria positive. The model predicted in respect to gender that seventy four percent of female population were malaria positive likewise the male gender predicted at eighty percent to be malaria positive.

Since our odds ratio for the age group and blood group is greater than the p-value at 0.05, we conclude that there is a high spread of malaria according to age and blood group.

Nevertheless, there were still a low spread of malaria according to a particular genotype since its odds ratio is lesser than the probability value.

Recommendation

I recommend that increased funds be made available so that research on malaria can be broadened according to the priorities addressed in the report, including laboratory and field research on the biology of malaria parasites, their mosquito vectors and their interaction with humans.

I recommend that local government areas be given support to orient malaria surveillance away from the mass collection and screening of blood slides towards the collection and analysis of epidemiologically relevant information that can be used to monitor the current situation sequentially, for proper monitoring, evaluation and identification of high risks group and to detect potential epidemics early in their course.

Compliance with ethical standards

Disclosure of conflict of interest

I declare that I have no conflicts of interest, financial or otherwise.

References

- [1] Bakare, E.A. and Nwozo, C.R. (2015) Mathematical Analysis of the Dynamics of Malaria Disease Transmission Model. *International Journal of Pure and Applied Mathematics*, 99(5), 411-437. <https://doi.org/10.12732/ijpam.v99i4.3>
- [2] Baiden F, Webster J, Owusu-Agyei S and Chandramohan D. (2011) Would rational use of antibiotics be compromised in the era of test-based management of malaria? *Trop Med Int Health Journal*. 142-4. doi:10.1111/j.1365-3156.2010.02692.x. pmid:21087379
- [3] Poyer S, Shewchuk T, Tougher S, Ye Y, Group A and Mann G. (2015) Availability and price of malaria rapid diagnostic tests in the public and private health sectors in 2011: results from 10 nationally representative cross-sectional retail surveys. *Trop Med Int Health Journal*. 20(6):744–56. doi: 10.1111/tmi.12491. pmid:25728761
- [4] Tumwiine, J., Mugisha, J.Y.T. and Luboobi, L.S. (2008) Threshold and Stability Results for a Malaria Model in a Population with Protective Intervention among High-Risk Groups. *Mathematical Modelling and Analysis*, vol.13, 443-460. <https://doi.org/10.3846/1392-6292.2008.13.443-460>
- [5] Olaniyi, S., Okosun, K., Adesanya, S. and Lebelo, R. (2020) Modelling Malaria Dynamics with Partial Immunity and Protected Travellers: Optimal Control and Cost-Effectiveness Analysis. *Journal of Biological Dynamics*, 14(1), 90-115. <https://doi.org/10.1080/17513758.2020.1722265>
- [6] Yeung S, Patouillard E, Allen H and Socheat D. (2019) Socially-marketed rapid diagnostic tests and ACT in the private sector: ten years of experience in Cambodia. *Malaria Journal* 22(1); 10:243. doi: 10.1186/1475-2875-10-243. pmid:21851625
- [7] Littrell M, Gatakaa H, Phok S, Allen H, Yeung S and Chuor C. (2011). Case management of malaria fever in Cambodia: results from national anti-malarial outlet and household surveys. *Malaria Journal*; 10:328. doi: 10.1186/1475-2875-10-328. pmid:22039922

- [8] Ansah EK, Narh-Bana S, Affran-Bonful H, Bart-Plange C, Cundill B and Gyapong M. (2022) The impact of providing rapid diagnostic malaria tests on fever management in the private retail sector in Ghana: a cluster randomized trial. *BMJ Public Health*; 350:h1019. doi: 10.1136/bmj.h1019. pmid:25739769
- [9] Saydam FN, Erdem H and Ankarali H. (2020) Vector-borne and zoonotic infections and their relationships with regional and socioeconomic statuses: An ID-IRI survey in 24 countries of Europe, Africa and Asia. *Travel Med Infect Dis* ; 44:102174.
- [10] Olaniyi, S., Okosun, K., Adesanya, S. and Lebelo, R. (2020) Modelling Malaria Dynamics with Partial Immunity and Protected Travellers: Optimal Control and Cost-Effectiveness Analysis. *Journal of Biological Dynamics*, 94(5), 90-115. <https://doi.org/10.1080/17513758.2020.1722265>
- [11] Chitnis, N., Cushing, J.M. and Hyman, J.M. (2006) Bifurcation Analysis of a Mathematical Model for Malaria Transmission. *SIAM Journal on Applied Mathematics*, 67(5), 24-45. <https://doi.org/10.1137/050638941>
- [12] Kimoloi S and Rashid K. (2015) Potential role of Plasmodium falciparum-derived ammonia in the pathogenesis of cerebral malaria. *Front Neuroscience*; vol. 9. 9. 13
- [13] Heeringa SG, West BT, Berglund PA. (2017) Applied survey data analysis. 2nd ed. New York: Chapman & Hall/CRC.
- [14] Georges, M. (1957) The Epidemiology and Control of Malaria. Oxford University Press, Oxford.m up
- [15] 8 Breman JG, Alilio MS, Mills A. (2004) Conquering the intolerable burden of malaria: What's new, what's needed: A summary. *Am J Trop Med Hyg* 2004; 71: 1–15.
- [16] Ramdzan A, Ismail A., and Zanib Z.M., (2020) Prevalence of malaria and its risk factors in Sabah, Malaysia. *International Journal of Infectious Diseases*. 2020; 91: 68–72. Available online at www.elsevier.com/locate/ijid. doi: 10.1016/j.ijid.2019.11.026 [PubMed] [CrossRef] [Google Scholar]
- [17] Dev V, Sangma BM and Dash AP.(2010) Persistent transmission of malaria in Garo hills of Meghalaya bordering Bangladesh, north-east India. *Malaria Journal*; 9:263. doi: 10.1186/1475-2875-9-263.
- [18] Ladjumadil Ahmad Tiu (2021) Impact Of Temperature And Rainfall On Incident Malaria. *IOP Conference Series: Earth and Environmental science*; IOP Conf. Ser.: Earth Environ. Sci. 75(5) 012084 DOI 10.1088/1755-1315/755/1/012084
- [19] Labadin, J., Kon, C.M.L. and Juan, S.F.S. (2009) Deterministic Malaria Transmission Model with Acquired Immunity. *World Congress on Engineering and Computer Science*, Vol. 2, San Francisco, 20-22 October 2009, 779-784.
- [20] Ruth E. Lagerberg (2011) Malaria in Pregnancy: A Literature Review *Journal of Midwifery & Women's Health* Volume 53, Issue 3 p. 209-215
- [21] Baruch DI, Gormely JA and Ma C. (2015). Plasmodium falciparum erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. *Proc Natl Acad Sci USA* 1996; vol. (1)93: 3497–3502.
- [22] Turner L, Lavstsen T and Berger SS (2013) Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature J Am Stat Assoc*; 49(2): 502–505.
- [23] Rulisa S, Kateera F, Bizimana J. (2013) Malaria prevalence, spatial clustering and risk factors in a low endemic area of Eastern Rwanda: a cross-sectional study. *PloS one*. 2013; 8(7). Available online at doi: 10.1371/journal.pone.0069443 [PMC free article] [PubMed] [CrossRef] [Google Scholar]