American Journal of Theoretical and Applied Statistics
2026; 15(3): 88-103
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20261503.12

Modelling Malaria Incidence Using Poisson Generalized Linear Model: A Case Study Of Mbita Sub-County, Kenya

Samuel Okinyi* , Joseph Eyang'an Esekon, Martin Mutwiri Kithinji

Department of Pure and Applied Sciences, Kirinyaga University, Kerugoya, Kenya

Email address:

sokinyi@kyu.ac.ke (Samuel Okinyi), jesekon@kyu.ac.ke (Joseph Esekon), mkithinji@kyu.ac.ke (Martin Kithinji)
*Corresponding author

To cite this article:

Samuel Okinyi, Joseph Eyang'an Esekon, Martin Mutwiri Kithinji. (2026). Modelling Malaria Incidence Using Poisson Generalized Linear Model: A Case Study Of Mbita Sub-County, Kenya. *American Journal of Theoretical and Applied Statistics, 15*(3), 88-103. https://doi.org/10.11648/j.ajtas.20261503.12

Received: 12 May 2026 ; **Accepted:** 22 May 2026 ; **Published:** 12 June 2026

Abstract: Malaria continues to pose a major threat to public health across the world, particularly in African countries where infection rate s remain high. Nearly half of the global population is exposed to the risk of contracting the disease. Malaria is caused by parasites belonging to the Plasmodium family and affects both humans and other warm-blooded animals. Over the years, researchers have explored different causes of malaria transmission using techniques such as spatial analysis, time series methods and regression models. Although these methods are useful, they are less suitable when the data set involved are categorical or count variables. This study used Poisson Generalized Linear Model to investigate factors associated with malaria incidence in Mbita Sub-county. The variables considered included treated mosquito bed net use, age group, educational level of household heads and access to healthcare services. The Poisson Generalized Linear Model was fitted having estimated its parameters. The findings showed that treated mosquito net usage, age group, and access to healthcare facilities were statistically significant. The educational background of the household head was not significant. The association between the exploratory variables and the response variable was determined by use of the Chi-square test.The results indicated that there was an association between mosquito bed net use and malaria, age group and malaria, education level of family head and malaria and finally healthcare access and malaria. The goodness of fit was conducted by the use of deviance statistic. A comparison between the null model and the full model was done and this resulted into a p-value of 0.001871, which was below the 0.05 significance threshold. As a result, the null hypothesis was rejected, indicating that additional exploratory variable improve the model. This suggests that the full model together with additional parameters significantly improves the fit of the model to the data. The study's findings reinforce existing evidence that the use of treated mosquito bed net use plays an important role in lowering malaria infections. organizations that handle matters in relation to health and environment such as World Health Organization and United Nations may apply the outcome to aid in developing mechanisms to lower the spread of malaria within Mbita Sub-county and other parts of the world with similar settings.

Keywords: Malaria, Poisson Generalized Linear Model, Deviance Statistic, Chi-square Test

1. Introduction

Malaria over a great period of time has showed to be one of the most outstanding worldly public well-being burdens with about half of the entire world population being at risk of being infected. Malaria is not a monolithic disease and thus can be caused by various pathogens such as *Plasmodium* genus, [20]. Malaria is not only restricted to human beings but also the warm-blooded vertebrate species. The species that have greatly led to the spread of malaria has triumphantly and extremely diversify across the globe.

The *Plasmodium falciparum* is the main cause of malignant malaria which is among the greatest severe human contagious diseases. The *Plasmodium reichenowi* which is a chimpanzee parasite is one of the nearest relative to *Plasmodium*

falciparum. The hypothesis test conducted suggest that the two parasites evolved differently from one common origin about five to seven million years ago. This was corresponding to the divergence of the major host, hominin and the chimpanzee lineages. An analysis was performed on *Plasmodium reichenowi* across Cameroon and results showed that it is the most widespread and genetically distinct chimpanzee parasite.

The genus *Plasmodium* is composed of approximately 200 species which are parasitic to various categories of vertebrates such as mammals, birds and reptiles. In addition to the plasmodium parasite, the *Toxoplasma*, *Cryptosporidium* and *Babesia* cause disease in human beings [32]. There is no specific fossils record of the *apicomplexans* that gives a clear origin of the phylum. The *apicomplexans* is the most ancient species and can only be compared to multicellular kingdoms of animals, fungi and plants. The *Plasmodium* category had deviated from *apicomplexans* about 700 million years ago. This is assumed to have taken place before the origin of Cambrian and vertebrates from their hereditary invertebrate lineage.

There are several believes concerning the origin and evolution of human being malaria pathogens *Plasmodium falciparum* and *Plasmodium vivax* in Africa. Approximately 95% of the total malaria infection is accounted for by both *Plasmodium falciparum* and *Plasmodium vivax* which cause a great challenge to the health sector [25]. Scientists believe that the *Plasmodium falciparum* evolved the same time with the human ancestors about more than several millions of years ago. An assumption was made that the *Plasmodium vivax* emerged from southeastern Asia based on the cross-species transmission of the pathogen from macaque. However the multitude *Plasmodium* species contradicts that and suggest that *Plasmodium falciparum* and *Plasmodium vivax* originated from pathogens infecting wild animals such as African apes.

Microscopic characterisation identified parasites from apes that resembled *P. falciparum*, *Plasmodium malariae*, and either *Plasmodium ovale* or (the similar) *Plasmodium vivax* in humans, suggesting the existence of distinct Plasmodium species. Which were classified as *Plasmodium reichenowi*, *Plasmodium rhodaini* and *Plasmodium schwetzi*, respectively. The tiny microscopic features recognized parasite from the apes that looked similar to the *Plasmodium falciparum, Plasmodium malariae* and whether *Plasmodium ovale* or *Plasmodium vivax* in human beings suggested the specific *Plasmodium* species that were grouped as *Plasmodium reichenowi*, *Plasmodium schwetzi* and *Plasmodium rhodaini* respectively.

Additionally, the *Plasmodium malariae* is among the leading causes of human being malaria in both Asia and Latin America [24]. However, the *Plasmodium malariae* is not felt in central parts of Africa because of the fixation of mutation that inhibits the free expression of the receptors, Duffy antigen that is found in the human erythrocytes. The non-invasive technique showed that the wild chimpanzees together with gorillas across central parts of Africa are infected with pathogen that are highly and closely linked to the human being *Plasmodium vivax*. The *Plasmodium vivax* is thus assumed to have originated from Africa and selected based on Duffy-negative mutation.

The *Plasmodium vivax* being a human malaria pathogen originated from Africa due to the high frequency of missing Duffy blood category antigen in human beings in sub-Sahara Africa [14]. An investigation was done on phylogenetic association across 10 different species of *Plasmodium* which infected primates through use of 3 genes [16]. The *Plasmodium* pathogens that normally cause malaria are transmitted to human beings by the mosquito vector. Over the past year, the association between insects and mammals has been refined thus making the pathogen to partially elude the human together with pest immune system.

The Generalized Linear Model (GLM) is a mathematical method that is normally applied in modelling the association between the response variable and a group of independent variables [4]. The GLM has great advantages as compared to the Multiple Linear Regression since it enables one to bring together both categorical and continuous variables.

A study was conducted on when the GLM is put into use. Based on the study, the GLM was used in estimating parameters when the independent variables were exposed to various measurement errors [3]. The induced technique for the response variable on the estimated independent variable was not the GLM. Under circumstances when the various distribution of measurement errors are said to be well known, application of GLM recursively weighted the least square method and the adjusted data showcased to yield the maximum quasi likelihood.

Logistic regression is normally used in the modelling of binary variables that gives yes or no outcomes [37]. The Poisson Generalized Linear Model (PGLM) is used in count cases. While dealing with incidences in relation to disease modelling across any study the PGLM becomes one of the best models to be put into use. The individuals under study are followed for a period of time and the count is taken based on factors being considered.

2. Literature Review

A study performed indicated that malaria is one among the most harmful and perilous infections across the globe [12]. The study formulated a GLM that was based on PGLM together with negative binomial regression model in order to predict malaria incidence across three different confined locations in Senegal. The regressor variables taken into consideration were the climate factors which included monthly rainfall, standard temperature, and relative humidity. Additionally, the study looked unto other dependent variables like Insecticide-Treated bed-Nets together with the Artemisinin-based Combination Therapy. Based on the study, a predicting algorithm was established by capturing the meteorological regressor variable X_j at a given time t_{ℓ_j}, where ℓ_j represented the lag at which X_j maximized its correlation together with malaria cases. Rainfall was saturated so as to

minimize over-prediction. Results indicated that the Poisson regression model was more efficient than negative binomial regression model in predicting correctly the malaria cases having taken into account at least two of the independent variables.

A study was conducted in order to assess malaria propagation in association with ITN distribution in Nigeria by use of a Bayesian spatial generalized linear mixed model together with a Leroux prior [31]. Exploratory analysis technique was used to asses the variation in malaria spread in association with ITN dispersion in 1325 Demographic and Health review groups within Nigeria. In addition to that, a Bayesian Spatial Generalized Linear Mixed Model with a Leroux conditional auto-regressive prior for the cluster-level effects was applied in the modelling of spatial and contextual fluctuation in malaria occurrence and Insecticide-Treated Net(ITN) allocation having adjusted for the environmental variables. Based on the study the allocation of ITN displayed considerable variation across the 6 political regions providing a p-value less than 0.05. The ITN allocation was so high in rural areas compared to urban areas. Based on the Bayesian hierarchical regression, the outcome revealed a statistically non-significant negative association between malaria dominance and ITN range. However, a significant spatial structured latent effects and non-hierarchical random effect was observed. The main associates of malaria spread included, ultimate temperature, rainfall and proximity to water bodies. A conclusion was drawn that malaria spread was insignificant in association to ITN allocation.

A research aimed at obtaining the core contribution of Intermittent Preventive Therapy (IPT) of malaria in relation to malaria pharmaceuticals in Ghana [8]. Data was obtained from a public hospital in Ghana over a period of about 8 years, from December 2008 till January 2017, the data being captured on a monthly basis. A GLM was applied with a negative binomial approach to examine the contribution of IPT. In the study, a log-linear time-series formulation was implemented and later used in the validation of the results acquired from logistic model. Based on the study, it was shown that maternal health registration of expectant mothers, male partner participation in maternal health facility attendance and number of visits prenatal clinics were the core contribution of IPT uptake.

A study was conducted to asses the effect of climatic conditions [23]. In particular, the variables confined were wind, rainfall, humidity and temperatures on malaria cases in Uganda. Monthly data from 2020 to 2024 was obtained from Uganda Ministry of Health. The study applied Auto-regressive Generalized Linear Model (AR-GLM) together with lagged covariates. This method was used so as to obtain instantaneous and delayed factors. The outcome showed that rainfall and humidity increased with specific delays being positively linked with the malaria incidence. However the changes in temperature showed a complicated and volatile relationship. The AR-GLM model outperformed the standardized GLM. This was indicated by the diminished Akaike Information Criterion. The AR-GLM efficiently covered temporal dependencies on the malaria data set thus indicating insignificant auto-correlation in the residuals.

A study was conducted to determine the use GLM in the spatial assessment of small-area malaria cases rates in KwaZulu, South Africa. The authors used a method of adjustment of regression analysis outcome that indicated a strong spatial association in the proportions by use of generalized linear mixed models and variograms [22]. The results of spatially adjusted, and multiple linear regression assessment indicated a strong positive relationship between winter rainfall and maximum temperature. There existed a negatively significant relation with increased distance from the water bodies. The statistical framework used in the production of the map in prediction of malaria cases in the area having considered various variation from model forecasting to determine whether the variation was endorsed by the collected data. The input variables indicated that minor variations in weather had significant effect in malaria transmission even in regions that had been subjected to various ways of controlling malaria for a good period of time.

Based on the study conducted to examine short term impacts of rainfall and temperature together with the social economic factors on malaria cases within Rwanda and Uganda from 2002–2011, generalized additive mixed model was applied to estimate the nonlinear impacts on temperatures together with rainfall data set [11]. From the study, a time series cross validation technique was used for selection of the most suitable subset of the socioeconomic inputs to establish the magnitude of smoothing the climate variables. The outcome indicated that trends in malaria cases varies based on both temperature and rainfall. The assessed short term impacts of temperature together with the precipitation were nonlinear. The outcome supported what previous researchers had done.

A study was conducted using Generalized Seasonal Auto-regressive Integrated Moving Average Model (GSARIMA). Count data was used in implementing malaria time series with minimum number of cases [9]. The GSARIMA model was used for parsimonious evidence-based modelling of the non Gaussian and non stationary seasonal time series of the data count. The two models were applied on the monthly malaria incidence time series in Lanka that had experienced decline drastically in the recent past. The outcome showed that malaria incidences had indicated a long term fluctuation averagely and unstable variance together with seasonality. The posterior predictive distribution showed that the negative-binomial model gave a suitable forecasting as compared to the Gaussian models more so when the count was low. The auto-correlation that was available in the series was captured by the GSARIMA model. The GSARIMA model was of great importance towards the elimination of malaria due to the seasonality and non stationarity more so when the rate of control was increased.

A study was conducted on malaria incidence in Jubek; South Sudan using the GLM [19] . The study focused on the various trends of malaria cases and data was on the weekly basis from January 2011 to October 2015. Data obtained from the Directorate of Preventive Health Services of Ministry of Health, South Sudan was used in the assessment. In the

analysis of the data, Poisson and negative binomial regression models were applied. The outcome showed that malaria incidences had increased by 0.3% per week.

A study aimed at examining the prevalence rate and the relation of various factors such as demographic, socio-economic and geographic state in relation to the quick spread of malaria in Ethiopia [6]. The GLM was applied on data with either presence or absence of malaria being the independent variable. Based on a baseline malaria survey done by Carter Centre, Rapid Diagnostic Test was used. The results of rapid diagnostic test indicated a significant association between the age together with gender. In addition to that, there were meaningful explanatory variables such as water sources and latrine facilitation together with roofing materials. Houses that used clean water had a less malaria prevalence. Mud thatched house and cow-dunged floor houses had a high malaria diagnosis. The water sources, housing state, distance from the water source, age and gender of household members had impact on malaria prevalence. Children and the female participants were the most likely to suffer from malaria. Household members with archaic socioeconomic status were positively linked with the infection of malaria.

A research study performed to asses spatial variation in malaria cases at a sustained spatial scale together with estimated zone in which prevalence surpasses the required benchmark of 5% across Rwanda [30]. The socioeconomic and atmospheric conditions were also put into consideration. Data was sourced from Rwanda Demographic and Health Survey. An application of geostatistical modelling approach was used. The approach was based upon the Stochastic Partial Differential Equation to evaluate the geospatial incidence of malaria in the midst of children below five years in Rwanda. The application of Bayesian inference was based upon the integrated nested Laplace approximation. Having done the assessment, the outcome showcased uneven spatial fluctuation of malaria incidence with some of the districts within Rwanda such as Kirehe, Huye, Rusizi and Kayonza. The malaria incidence was observed to be increasing with the rising temperatures but decreasing with increased millimeters of rainfall collected by the rain gauge. Based on the results, it also showed a meaningful relation between the spread of malaria and demographic determinants such as educational attainment of the mother, area of domicile, pediatric age and gender.

A study to determine the efficiency of utilization of the local spatial deviations to reveal the statistical association between malaria occurrence rate and the environmental factors within various counties in Kenyan [15]. The Data set that was used was retrieved from Kenya Demographic and Health Survey and the nationwide malaria indicator survey of 2015. The outcome of spatial assessment indicated clusters of various counties with greater incidence rate of malaria in Lake Victoria region, east coastal region and Mombasa. The cold areas such as Nairobi had low cases of malaria. The geographical weighted regression was applied for assessment of the obtained data sets. The applied technique aids in modelling how the environmental together with social variables associate with malaria case rate while considering the confounding factors of non-stationarity. Factors such as closeness to water sources, rainfall and population size had different impacts on malaria cases in Kenya. El-Nino-Southern Oscillation (ENSO) that occurred in 2015 had a great significant impact towards malaria more so in southern parts of Lake Victoria as compared to previous times. The spatial multivariate clustering assessment indicated the relevance of both behavioral and social participant feedback.

A study on spatial and periodic dynamics of malaria transmission in rural Western Kenya. The study was aimed at understanding the relationship between *Plasmodium falciparum* malaria spread and the health outcomes [1]. The study examined the extent of spread together with drivers of regional difference in malaria encountered in habitual zones. Data was modelled using the Geostatistical zero inflated binomial and negative binomial technique to obtain the various locations of house approximation of sporozoite frequency together with mosquito densities. The techniques' predictions were escalated so as to estimate spatial trend of yearly entomological inoculation incidence. The model used were environmental factors together with climatic determinants obtained from satellite data, harmonic cyclical variations and indicators depicting space-time association. The outcome indicated that the *Anopheles gambiae* was the central vector type accounting for a total of 86% mosquitoes gathered. Out of the total surveyed houses, 68% had no mosquitoes. The distance from water sources, daily temperatures and the vegetation had a strong positive associated with the mosquito prevalence. The monthly mosquito prevalence changed during the study time and was high during the month of May. Based on the predictions and observation, prevalence of mosquito Entomological Inoculation Rate indicated a prominent seasonal and spatial trend in the study zone.

A study based on the spatial dynamic risk determinants for malaria across geographically heterogeneous terrain in western parts of Kenya. Through the application of various techniques such as spraying of houses with insecticides together with sleeping under treated mosquito bed nets, this has had a great impact to control malaria [18]. In various parts of the country such as Rusinga Island, malaria has been proven to be so rapid in spreading whereby a population of about 3632 from 790 households was found to suffer from malaria. Data used for analysis was collected through use of questionnaires and the ecological factors were obtained through Quick-bird aerial imagery. Data was analyzed by fitting a multiple linear regression model that contained various variables to asses the level of total spatial variation in the malaria incidence that would be justified by both socio-demographic and environmental indicators. In addition to that, a Geographically-Weighted Regression (GWR) was conducted having assumed non-stationarity of the various risk factors. The GWR model was as follows:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ki} + \varepsilon_i$$

$y_i =$ Represents the cluster prevalence.

$\beta_0 =$ Represents the intercept.
$\beta_k =$ Represents the coefficient estimates, $k = 1, 2, \cdots, p$.
$X_{ki} =$ Represents the co-variate values for observation i.
ε (error term) $\sim \mathrm{N}(0, 1)$.

Diagnostic test for multicollinearity was conducted before doing the analysis. The outcome indicated that the prevalence rate was at 24% and that 2 groups had meaningfully increased numbers of malaria incidence. The correlation coefficient suggested that there was an association among living standards, outside jobs, population size and increased malaria cases. The geographically-weighted regression technique helped in improving the model goodness of fit and the association of malaria with various risk factors varied spatially over Rusinga Island.

A longitudinal non-experimental study of infants and toddlers who had been admitted at Kilifi Hospital, Kenya [28]. The non-experimental study was linked to data of population of individuals that slept under treated mosquito nets. Data was retrieved from 69104 juveniles between the age bracket of 3 months to 13 years. The data was for a period of 25 years from 1990 to 2014. Data set was analyzed using the logistic regression model with the exploratory variables being: age of the child, place of dwelling and vegetation cover metric. The number of patients admitted in the hospital went on rising and falling as years went by. The older juveniles accounted for a greater increase in the years 2009 to 2014. The results showed a nonlinear association between the risk of malaria and its occurrence. The older juveniles arose with increased vulnerability to malaria.

A study conducted on the assessment on transmission of malaria using the multilevel logistic regression in Kenya [4]. The key objective of the research was to evaluate the effect of Sample Enumeration Areas (SEAs) together with SEA attributes on personal malaria quick screening test. The data set used was secured from Kenya National Malaria Indicator Survey. The data set comprised of 301 aggregates of which 134 were from urban areas and 167 from rural areas and was then evaluated using the multilevel logistic regression technique. The weights were considered for the adjustments of irregular probabilities of selection within groups. There existed a malaria indicator survey hierarchical structure that was due to distinct data. The used technique did not take into consideration any association between various data points and an assumption was built that personal malaria conditions were distinct of their drivers. The outcomes indicated that geographical regions, residential area, people who sleep under treated mosquito nets, location of water sources, household asset index and size were statistically related to malaria's prevalence. Having taken into consideration various variables, the SEAs accounted for 47.1% while the rest were the unobserved variations between individuals.

A research concerning the changing malaria prevalence on coastal region of Kenya since 1974 taking into consideration the climatic conditions, drugs and vector control [35]. A total number of 1,147 age-standardized *Plasmodium falciparum* pathogen prevalence evaluations done across rural communities within the coastal region in Kenya were brought together between 1974 to 2014. The data was analyzed with the aid of Bayesian conditional auto-regressive generalized linear mixed model in order to estimate 279 sub areas. A comparison to the sequential predictor variables associated with amount of rainfall, pharmaceuticals resistance and treated mosquito bed nets use was done by fitting a polynomial splines curves of changing *P. falciparum* parasite rate. Based on the outcome it showed that the malaria prevalence kept on fluctuating from time to time. In 1998 going forward, the prevalence began to decrease till 2011 and increased much in 2014. The major decline was before the introduction of treated bed mosquito nets being widely distributed and also the variation in amount of rainfall. The rise and fall of transmission of malaria along the coastal region was due to emerged resistance to quinine derivative.

A study was conducted to examine the association between malaria incidence as an estimate of dissemination and death rate across various age sets. The Parasitical data was obtained from the yearly cross-sectional assessments from Kisumu [21]. The Health and Demographic Surveillance Systems data used from 2007 to 2015 and was used in the determination of malarial parasite incidence together with clinical malaria. The family unit surveys together with Verbal Autopsy were applied in order to obtain data on most causes and malarial death rates. The Bayesian negative binomial geo-statistical regression model was applied so as to asses the link between clinical malaria and death rate across different age sets. A comparison was made based on the annual data and aggregated data for a period of about 5 years. The results showed an association between pathogen incidence and malaria-specific death rate.The Bayesian credible interval was found to be 95%. The clinical malaria was linked to both all-causes and malaria specific death rate in grouped ages. There was no association between infants and adults. The distance from various health centres, social economic state and study year were the most essential elements for all-causes and malaria death rates. The malaria pathogen from the study was linked with death rate across the age set of 4 to 5 years with the clinical malaria having a strong relation with death rate as compared to malaria prevalence. The effect appeared more stronger among young adults between 5 and 14 years as compared to other age categories.

Over several decades, malaria has been of great concern to the society according to [26]. The close monitoring of scale heterogeneity of malaria cases has been a burden and measures have been taken to monitor the spread of malaria. The Malaria Indicator Survey has been of great help in assessing the progress of malaria mitigation strategies. The geo-statistical techniques enable the interpolation of country wide data to elaborate regional disease impact which is essential in informing targeted population. The binomial geostatistical technique applied the Markov Chain Monte Carlo parameter evaluation methods. This was adopted for the understanding of spatial drivers of malaria exposure across Kenya in the prediction of malaria exposure and location of concentration

zones. The data set included 11,549 juveniles of age group 6 months to 14 years. The malaria prevalence was 8.4% with lake region having the greater part of about 18.1%. The results showed that treated mosquito nets distribution, amount of rainfall and temperature were all meaningfully independent variables of malaria spread. The Lake Victoria zone, coastal areas and shores of Lake Turkana were the most affected with malaria. Based on the malaria incidence maps produced from the analysis, it indicated that Kenya had increased malaria zones especially in areas least expected.

The Ethiopian government provided treated mosquito nets to cub the spread of malaria [27]. The spread of malaria and its trend were so high during 2001 to 2011. The treated mosquito bed nets distribution record was used to identify the ratio of population at risk protected by the treated mosquito bed nets. Time series analysis was applied to the data set from 41 healthcare institutions in malaria prone zones to examine the trend of malaria incidence and mortality. The ratio of population protected from malaria due to sleeping under treated mosquito nets increased by 51% in 2011. The confirmed malaria incidence reduced by 66% in 2011. Among juveniles below 5 years, the malaria admission together with mortality rate decreased by 81%. In addition to that, the threshold value for the trend-lines took place between January and June in 2006. Across the same duration, non-malaria incidence and mortality neither increased nor remained constant. The various malaria diagnostic that was conducted aids in the reduction of malaria cases. The amount of rainfall contributed greatly to the spread of malaria.

A study was performed to asses the shift in varied trends in malaria time series within the western parts of Kenyan Highlands [10]. Inferential temporal findings were obtained from malaria cases time series within Kericho County. Data visualization was conducted so as to explore the various dynamics of time series. The auto-correlation function of malaria inpatient time series was examined. The assessment showed that a greater portion of malaria time series was of 1^{st} order seasonal auto-regressive process. The observations were correlated at one month together with a seasonal time lag of twelve months. A second order seasonal auto-regressive process was fitted and it was of the form:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \alpha\text{trend} + \varepsilon_t$$

where

y_t = Seasonal auto-regressive process.
μ = Mean value for the time series.
ϕ_i = Non-linear trend estimate, $i = 1, 2$
α = The effect of time on the model.
ε_t (error term) $\sim \text{N}(0, 1)$.

A comprehensive assessment of a five monthly time series spanning between fifteen and forty one years was done from the western parts of Kenya incorporating an altitude slope across the shores of Lake Victoria. The results indicated a declining but varied malaria trends in the 1980s at low-elevation zones and 2000s at elevated areas. The regime variation was experienced in 3 series and were simultaneous in two time series from elevated areas. At low-elevated zones, regime variation was linked to a variation from the rising and falling of malaria transmission together with decreased variability. At the highlands, regimes variation showed an increase in the transmission of malaria variability. The non-uniformity in malaria trends reflected the multitude of various factors which contributed to spread of malaria and highlighted the importance of the spatial and temporal fine-grained set of data to make a suitable inference concerning the effect of climatic changes as well as control various interventions on the transmission of malaria.

A study to assess the impact of climate changes on malaria trends across Baringo County in Kenya [2]. The key variables were: amount of rainfall, highest temperature together with vegetation on the seasonal trends across selected health institutions in Baringo County. The climatic factors were retrieved from both International Research Institute and Lamont-Doherty Earth Observatory climatic database together with malaria incidence recorded in 10 healthcare institutions within four environmental zones. Time series analysis was conducted on lowland, highland, stream-side and mid-altitude from 2004 to 2014. The negative binomial regression technique with lagged climatic factors was used in modelling the long term per month malaria cases. In order to detect the general monotonic trends in malaria incidence, seasonal Mann–Kendall test was performed. The outcome showed that malaria incidences increased more significantly across the highland together with midland areas. The various variation in malaria incidences were due to fluctuations in the amount of rainfall and highest temperatures. The amount of rainfall at a time lag of two months led to increased malaria transmission across four regions. In addition to that, increased temperatures at time lag 0 and 1 month respectively led to increased malaria incidence in highland areas. The transmission of malaria was at its peak during the short rains and after long rainy seasons.

A research was performed to asses the impact of climate diversity on malaria cases after scale up interventions in the western parts of Kenya using time series analysis [29]. The Bayesian negative binomial approach was applied on the monthly malaria cases data from 2008 to 2019. The data set was secured from patient records that had febrile sickness visiting Lwak Mission Hospital for treatment. The data containing individual who slept under treated mosquito nets, and social and economic situation was retrieved from family unit survey. Additionally, climatic proxy attributes obtained from geospatial monitoring were also included as explanatory variables within the model. The Bayesian indicator selection was applied to determine the time that elapsed between climate significance and malaria cases. The outcome indicated that malaria cases had increased by 50% between 2008 to 2010 and later declined by an average of 73% by 2015. The low temperatures led to fall of malaria incidence giving an incidence rate of 0.70 and the Bayesian credible interval range of 0.59–0.82. The treated mosquito bed net use led to the fall of malaria incidence. The change in climatic variable indicated a strong significant effect on malaria cases than treated bed net

use.

A study was conducted by Kenya Medical Research Institute (KEMRI) to determine the relationship between weather changes and children under 5 years mortality in western part of Kenya from 2003 to 2008 [33]. The data set was modelled by use of time series. The general additive approach together with Poisson linked function was fitted to the model of per week association of the lagged total rainfall together with mean temperatures on malaria death rate at KEMRI. The time trend together with seasonality was controlled through introduction of the trend function. The outcome indicated that malaria death rate was linked to temperature and amount of rainfall in KEMRI's Health and Demographic Surveillance System (HDSS).

A study performed on remote sensed environmental factors and malaria death rate in three different malaria prone zones in western parts of Kenya [34]. It was established that malaria depends on various conditions such as amount of rainfall and the environmental temperatures. The lagged relationship of the normal difference environmental condition, surface temperatures together with precipitation on malaria death rate were considered. The lagged impact of each environmental factor per week death rate was modelled through distributed lag non linear technique. In each and every variable, the natural spine with three degrees of freedom was constructed for the variable and lag dimension respectively. A consideration of the lag period was made till the twelfth week. The impact of environmental conditions fluctuated between different zones containing longer lags. Across the three different zones, malaria death rate was linked to precipitation. The risk arose with increased weekly sum precipitation above 20 millimetres and was at peak at 80 millimetres. The threshold for increased death rate was approximately 0.3 to 0.4 at smaller lags. The results showed that amount of rainfall had the greatest consistent predictive trends to malaria spread.

3. Methodology

3.1. Informative Statistics

The entire summary of the data set was carried out using informative statistics. This study aimed at around 400 household respondents, 4 predictor variables together with 1 dependent variable. The 4 independent variables are treated mosquito bed net use(X_1), age group(X_2), education level for household head(X_3) and access to good healthcare facilities(X_4). The response variable is the number of malaria incidence within every household in Mbita Sub-County.

3.2. Target Population

Mbita Sub-County is the main region of interest and the targeted population comprised of people who reside within Lake Victoria basin in Mbita.

3.3. Data Collection

The data set that was utilized for this study was primary data collected with the help of KoboCollect having drafted the questionnaire. There was an extensive coverage of various questions to ensure objectives were well covered for the achievement of the research. The major considerations were the number of malaria incidence, treated mosquito bed net use, age group, education level for household head and access to healthcare facilities. After collection of the data, it was imported to excel and then a further importation to R statistical software for intensive data cleaning and analysis was done.

3.4. Validity of the Data

To ensure that the obtained data was valid enough, consistency of the collected data was considered. The reliability and completeness were also done together with missing values. The data set was based on clinical malaria incidences and not just symptomatic cases. The independent variables such as treated mosquito bed net use (X_1), age group(X_2), education level for household head(X_3) and access to good healthcare facilities(X_4) were considered to be categorical and the dependent variable, malaria incidence was treated as count.

3.5. Stratified Random Sampling

Based on the research, stratified random sampling was used. This was a two-step procedure where the entire population was partitioned into sub population called stratum. The strata were mutually exclusive and exhaustive in such a way that every element of the population was assigned to one and only one stratum and no population element was omitted. The elements of the study were selected from each stratum by simple random sampling.

3.6. Sample Size

The expected general formula that was implemented to determine the sample size was of the form:

$$n = Z^2 \times P \times \frac{1-P}{d^2}$$

where:
$n =$ Sample size
$Z =$ The Z-value from the statistical table
$P =$ The targeted population proportion
$d =$ The amount of error to be tolerated

After substituting the values in Equation (1) we get

$$(1.96)^2 \times 0.5 \times \frac{1-0.5}{0.05^2} = 384$$

3.7. Generalized Linear Model Assumptions

The assumptions for both linear regression modelling and generalized linear model are similar. These assumptions need to be satisfied before conducting any analysis in either regression modelling or Poisson generalized linear model [36].

The assumptions of a given model are on the basis of the ordinary least squares technique as suggested [13]. The assumptions include:

1. On the basis that all factors are held constant, the relationship between each exploratory variable and the dependent variable is assumed to be linear.
2. The Distribution for the error term (ε) is normally expected to have a constant variance.
3. The residuals must have a normal distribution with mean $\mu = 0$ and variance σ^2 ($\mu = 0$,var= σ^2).
4. The observations are expected to be independent so as to guarantee that the independent variable are not highly correlated.

3.8. The Diagnostic Tests

The forecast on the diagnostic test is to be performed so as to ensure that the assumptions of the ordinary least squares were upheld [17]. The various diagnostic tests were conducted to find out if the data is fit for the appropriate analysis. The major diagnostic tests that were performed under this study included the following:

3.8.1. Test for Normality

The Q-Q plot together with Shapiro–Wilk test were used to verify the assumption for normality. The residuals were checked to find out if they are not deviating from the line of best fit. The residuals were assumed to have a normal distribution only if the p-value is more than the α level of significance which was 0.05 based on Shapiro–Wilk test. The dependent variable was assumed to be normally distributed when it met this condition. The hypotheses for this test are:

H_0 : The population(or residual) are normally distributed,

H_1 : The population(or residual) are not normally distributed.

3.8.2. Heteroscedasticity Test

The Breusch-Pagan test was used for testing heteroscedasticity since it changes the normal standard errors by raising the likelihood function thus contributing to Type II error. Based on the Breusch-Pagan test, when the p-value is greater than the significance level which is 0.05, it shows presence of heteroscedasticity.

3.8.3. Multicollinearity Test

The test for multicollinearity was facilitated through use of Variance Inflation Factor (VIF). When the VIF value was observed to be less than 1, that showed non existence of association among the exploratory variables. When the VIF value was greater than 5, that suggested a great correlation. In mathematical form, each and every response variable X_q was regressed against the other independent variables. The equation is of the form:

$$X_q = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_q X_q + \varepsilon$$

$\beta_i, i = 1, 2, \ldots, q$

ε (error term)

Letting R_j^2 be the coefficient of determinant, the VIF is given by:

$$VIF_j = \frac{1}{1 - R_j^2}, 0 \leq R_i^2 \leq 1$$

In some instances, these assumptions may fail be to upheld as suggested thus making it difficult to fit the required model [7]. The diagnostic tests therefore have to be performed and satisfied before fitting the model.

3.9. The General Poisson Generalized Linear Model

The Poisson Generalized Linear Model is applicable when dealing with explanatory variables containing a mixture of continuous data and categorical data set. The general equation of the Poisson Generalized Linear Model is of the form:

$$\log(E[Y]) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_q X_q$$

where

$Y =$ Response variable

$X_i =$ Independent variable, $i = 1, 2, \ldots, q$

β_0 = Unknown regression coefficient of the response variable (Y) having considered other factors being held constants (X_i's=0)

$\beta_1, \ldots, \beta_q =$ Unknown parameters that can be explained by the independent variables.

3.9.1. The Specific Poisson Generalized Linear Model

The specific equation for the Poisson Generalized Linear Model was of the form:

$$\log(E[Y]) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \qquad (1)$$

where

$Y =$ Number of malaria cases,

$X_1 =$ Treated mosquito bed net use,

$X_2 =$ Age group,

$X_3 =$ Education level for household head,

$X_4 =$ Access to good healthcare facilities,

3.9.2. Testing For Goodness of Fit of the Model

When testing for the goodness of fit for the Poisson Generalized Linear Model either a deviance statistic or Pearson chi-square statistic can be conducted. Either of the two tests can be conducted to evaluate how good the model is.

3.9.3. Deviance Statistic

The deviance statistic was put into use to determine how fit the model was. The main goal of the deviance statistic in testing for goodness of fit was to measure the major difference between the model to be fitted (null model) and the saturated model. The appropriate equation for the deviance statistic is given by:

$$D_s = 2\sum_{j=1}^{n}\left[Y_j \log\left(\frac{Y_j}{\mu_j}\right) - \left(Y_j - \mu_j\right)\right] \tag{2}$$

where
$D_s =$ Deviance Statistic.
$Y_j =$ observed malaria cases.
$\mu_j =$ Expected number of malaria cases.

The deviance normally follows a chi-square distribution with $(n-k)$ degrees of freedom, that is

$$D_s \sim \chi^2_{(n-k)}$$

where k and n are number of parameters and sample size respectively.

4. Data Analysis, Results and Discussions

4.1. Informative Statistics

The informative statistics was carried out by the use of tables, graphs together with summary measures. The summary of the given data set is given in Table 1 in relation to maximum, minimum and mean. This was based on the number of individuals who had been diagnosed to have suffered from malaria from January 2025 to December 2025.

Table 1. *Informative statistics based on maximum, minimum and mean.*

Statistics	Total household members	Malaria incidence in every household
Maximum	10	6
Minimum	1	0
Mean	5	3

Based on the summary statistics in Table 1, the highest number of household members was 10. The maximum number of individuals diagnosed to have suffered from malaria was 6. The lowest number of total household members was 1 with 0 malaria incidence. The mean for the aggregate members was 5 and the diagnosed mean members was 3. This indicated that at least half the members of each and every total household members had been diagnosed to have suffered from malaria.

4.2. Incidence Rate

The malaria incidence rate was considered over the last one year from January 2025 to December 2025. Incidence rate is given by

$$\text{Incidence Rate} = \frac{N}{P} \times K$$

where N, P and K are number of new cases, population at risk and scaling constant respectively. When the values were substituted, the incidence rate was obtained to be

$$\frac{1430}{2477} \times 1000 = 577.3112$$

This suggests that in every population of 1000 people, there would be 577 new cases of malaria in Mbita Sub-county. Based on the result, this indicates that more than half of the population is at a risk of being affected by malaria.

4.3. Summary of Malaria Incidence

The distribution of the malaria cases is shown in table form indicating the number of malaria incidence against frequency. This is used to show the distribution of the malaria incidence.

Table 2. *Distribution of malaria Incidence.*

Malaria Incidence	Frequency	Total Incidence
0	1	0
1	45	45
2	60	120
3	225	675
4	50	200
5	48	240
6	25	150
	Total	1, 430

Table 2 shows the frequency of how every incidence occurred. The results clearly indicate that 3 malaria cases were the most witnessed across many households. The households that had 1, 2, 4 and 5 cases of malaria incidence were also fairly distributed. The highest number of malaria cases was 6 and was only witnessed in 25 homesteads. One household never experienced any malaria case within the study period.

4.4. The Diagnostic Tests

The various diagnostic tests were performed before fitting the Poisson Generalized Linear Model. This was done to ensure that all the assumptions are met.

4.4.1. Assumption for Normality

To test for the assumption of normality, Q-Q plot was applied.

4.4.2. Test for Normality

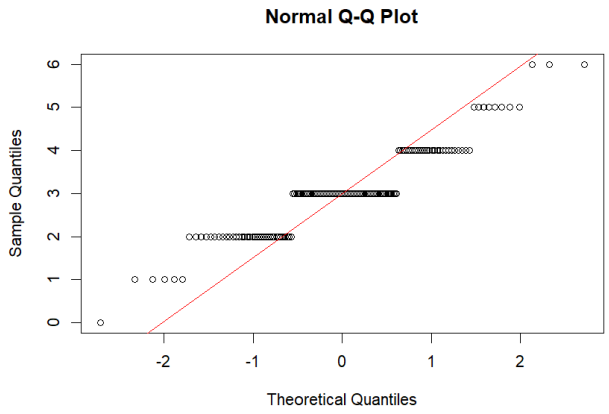

Figure 1. Q-Q plot of theoretical quantiles to check for normality.

Based on Figure 1, the residuals did not deviate so much from the line of best fit thus suggesting normality.

The assumption for normality was also confirmed by carrying out the Shapiro–Wilk test.

Table 3. Normality Test based on Shapiro–Wilk test.

level of significance	p-value
0.05	0.823

In addition to the Q-Q plot, we plotted the Shapiro–Wilk test to confirm for normality. Based on Table3, the p-value was 0.823 which is greater than 0.05 which is the α level of significance thus suggesting normality.

4.4.3. Test for Heteroscedasticity

To test for heteroscedasticity, Breusch-Pagan test was conducted to compare the p-value and the level of significance.

Table 4. Test for Heteroscedasticity.

level of significance	p-value
0.05	0.6441

The value was obtained to be 0.6441 which is greater than 0.05 level of significance thus suggesting presence of heteroscedasticity as indicated in Table 4.

4.4.4. Test for Multicollinearity

The variance inflation factor was considered to assess if there exists any association between the independent variables.

Table 5. TVIF Values.

Independent Variables	VIF
Bed Net Use	1.168173
Age Group	1.047315
Education Level	1.168644
Healthcare Access	1.086885

The VIF values are indicated in Table 5. They are less than 5 which is the reference point and thus suggests no multicollinearity.

4.4.5. Relationship Between Malaria Incidence and Bed Net Use

To asses the relationship between malaria incidence and bed net use, box plots were used to aid in the assessment.

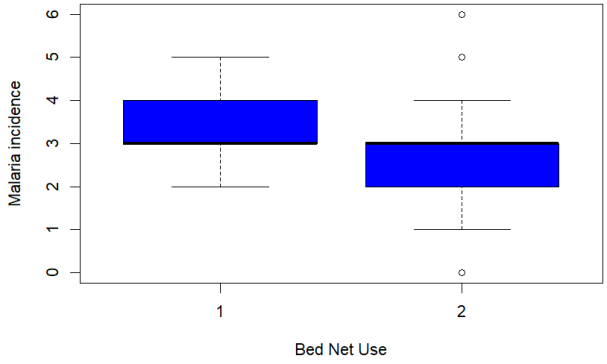

Figure 2. Box plots of Malaria Incidence versus Bed Net Use.

Based on Figure 2, it is clear that there is a relationship between those who sleep under treated mosquito nets and the malaria incidence. The first box-plot represents those who failed to sleep under the treated mosquito nets while the second box-plot represents those who slept under treated mosquito nets. The visualization indicates that individuals who failed to sleep under the treated mosquito nets lead to high rate of malaria incidence as compared to those who sleep under treated mosquito nets.

4.4.6. Relationship Between Malaria Incidence and Age Group

Under this section the visualization of malaria incidence and age group was considered based on box plots.

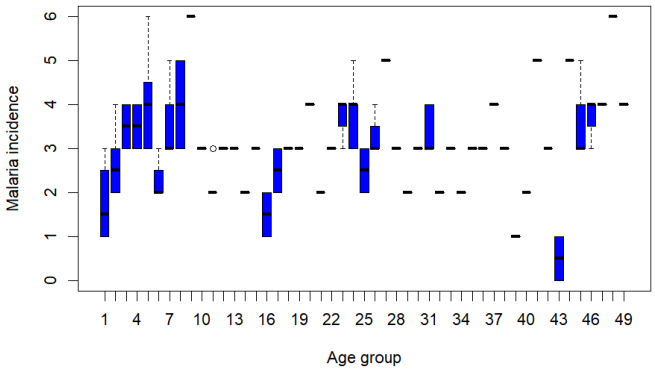

Figure 3. *Box plots of Malaria Incidence versus Age Group.*

On the visualization of age group against malaria incidence, the individuals aged 0–9 years were highly affected probably due to the immune system. The malaria incidence of individuals between ages 10–18 were moderate as compared to age groups between 0–9 and 28–36 years. Individuals from age 36 years onwards had a higher number of malaria incidence which could be linked to the immune system.

4.4.7. Relationship Between Malaria Incidence and Education Level

Based on the 6 selected educational levels that were: primary, secondary, certificate, diploma, bachelors and master; we get the result in the figure below.

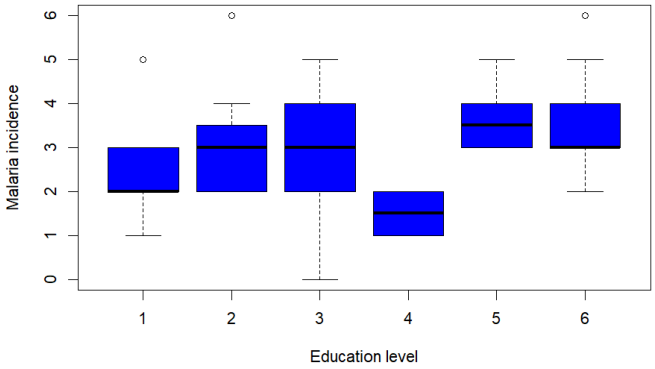

Figure 4. *Box plots of Malaria Incidence versus Education level.*

The distribution reveals the importance of education of the household heads. Level 6 represents households whose head of the family's highest level of education was primary and had the highest malaria incidence. This is due to low level of education and lack of information concerning malaria. Homes headed by those whose highest level of education was secondary also had a registered higher number of malaria incidence. Homesteads whose household heads had certificate, diploma, degree and masters had a low malaria incidence. This was due to them being informed concerning malaria and how to control it.

4.4.8. Relationship Between Malaria Incidence and Access to Healthcare Facilities

To show the relationship between malaria incidence and access to good health, the scatter plot and summary table in response to access to good healthcare were considered.

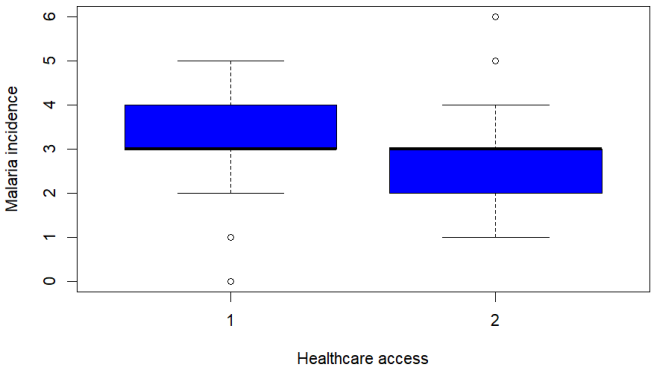

Figure 5. *Box plots of Malaria Incidence versus Access to Healthcare.*

Table 6. *Health access and malaria cases.*

Access to healthcare	Total
No	255
Yes	145

Homes where individuals have easy access to good healthcare facilities had less incidence of malaria as indicated in Table 6 and Figure 5. This is because they are able to access treatment and sensitization on matters related to malaria, how to control the spread and also receive treated mosquito nets from various health facilities such as level 4 and 5 hospitals for free. Those who stay away from the hospital experience high incidence rate since access to health facilities makes it difficult for them to receive the treated mosquito net during sensitization program organized by the governmental and non-governmental organizations.

4.5. Fitting the Poisson Generalized Linear Model

The unknown parameters β_0, β_1, ..., β_q were obtained by the use of the least squares technique before fitting of the Poisson Generalized Linear Model.

4.6. Estimation of The Parameters

The estimated parameters of the model are $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$. The analysis was applied in modelling the association between mosquito bed net use (X_1), age group

(X_2), education level for household head (X_3) and access to good healthcare facilities (X_4). There was great importance of assessing the intensity of resulting association among the independent variable and utilization of the results in determining prediction according to the resulting Poisson Generalized Linear Model as indicated in Table 7.

Having assessed the independent variables X_1, X_2, X_3 and X_4, the research opted for the model which would grant an outstanding fit in elaborating the association between the independent variables and the dependent variable. The best fitted specific Poisson Generalized Linear Model is in Equation (1).

Table 7. *Parameter estimates for the Poisson Generalized Linear Model.*

Coefficients	Estimates	std error	t-value	Pr$(>\|t\|)$.
Intercept	1.4353	0.71765	3.997	0.0064
Bed Net Use	−0.1618	0.08091	−1.502	0.0133
Age Group	0.0034	0.00164	1.013	0.0311
Education Leve	0.0293	0.01465	1.125	0.261
Healthcare Ac	−0.0793	0.03965	−0.777	0.0437

The estimated parameter values were fitted to the model to get the following:

$$\log(E[Y]) = 1.4353 - 0.1618X_1 + 0.0034X_2 + 0.0293X_3 - 0.0793X_4 \tag{3}$$

To eliminate the log on the left hand side, introduction of the exponential is done on both sides resulting in:

$$\mathrm{E}[Y] = \exp(1.4353 - 0.1618X_1 + 0.0034X_2 + 0.0293X_3 - 0.0793X_4) \tag{4}$$

The outcome from Equation (4) indicates that the intercept $\hat{\beta}_0 = 1.4353$.

$$\exp(1.4353) = 4.2009$$

Obtaining the exponential value of $\hat{\beta}_0$ results to 4.2009. This suggests an approximate of 4 malaria cases per household on assumption that mosquito bed net use (X_1), age group (X_2), education level for household head (X_3) and access to good healthcare facilities (X_4) are all held constant. This means that all the independent variables are assumed to have no effect on the dependent variable.

The estimated value in relation to bed net use had a $\hat{\beta}_1 = -0.1618$. When the exponential of $\hat{\beta}_1$ is obtained, it results to 0.8506. This indicates that families which do not sleep under treated mosquito nets have an approximate of 85.06% expected malaria incidence having held all other variables constant. Based on the results, 15% decline in malaria incidence is linked to the treated mosquito bed net use. $\hat{\beta}_2 = 0.0034$ leading to 1.003 done the exponential. This suggests that a single unit increase in age category is linked to 0.3% of malaria incidence having held all other factors constant. The effect tends to be small, thus suggesting minimal influence in the model.

The estimate of household head education level is $\hat{\beta}_3 = 0.0293$. On getting the exponential of estimated $\hat{\beta}_3$, it results to 1.03 thus suggesting an increase on malaria cases by 3% having considered all other factors constant. The reference point was based on highest education level of household head being primary. Additionally, the estimated value of access to good healthcare facilities was obtained to have a $\hat{\beta}_4 = -0.0793$. The resulting exponential value is 0.9238, thus indicating that individuals who lack access to good healthcare have 92% expected malaria incidence having held all other variables constant.

Table 8. *Chi-square test for mosquito bed net use and malaria incidence.*

Statistic	Value
χ^2	18.886
df	1
p-value	0.004361

4.7. Assessing the Association Between the Explanatory Variables and the Response Variable

To determine whether malaria incidence and the independent variables were significantly associated, application of the chi-square test was done. A statistical hypothesis was conducted to test on the independence of the variable. The hypothesis was as follows:

H_0 : There exists no association between the independent variable and the response variable,

H_1 : There exists an association between the independent variable and the response variable.

Table 8 shows the outcome of the chi-square test concerning the relationship between mosquito bed net use and malaria incidence.

Based on the results, the p-value is 0.004361($<$0.05). Thus we reject the null hypothesis that there is no association between mosquito bed net use and malaria incidence. This statistically implies that the spread of malaria is significantly

linked with mosquito bed net use.

Table 9. *Chi-square test for age group and malaria incidence.*

Statistic	Value
χ^2	4.521
df	5
p-value	0.002141

Additionally, a chi-square test was conducted to assess the association between age categories and malaria incidence. Table 9 indicates the findings. The resulting p-value is 0.002141 (<0.05). Thus we reject of the null hypothesis that there is no association between age group and malaria incidence. The results therefore suggest that the distribution of malaria cases is significantly associated with age group based on the data used.

The relationship between Education level and malaria incidence was also looked into.

Table 10. *Chi-square test for education level and malaria incidence.*

Statistic	Value
χ^2	62.919
df	5
p-value	0.04023

Based on the results in Table 10, the p-value is 0.04023 (<0.05) thus suggesting that the distribution of malaria cases is significantly associated with education level of household head.

Table 11. *Chi-square test for healthcare access and malaria incidence.*

Statistic	Value
χ^2	10.004
df	1
p-value	0.001245

Finally a chi-square test was conducted to asses the association between healthcare access and malaria incidence. As per the results in Table 11, the outcome based on the p-value is 0.001245 (>0.05). Based on the results, we fail to reject the null hypothesis and conclude that there exists an association between healthcare access and malaria incidence.

4.8. Testing for Goodness of Fit

To determine the goodness of fit for the Poisson Generalized Linear Model, the deviance statistic was conducted. This was done having taken into consideration Equation (2). Based on the hypothesis test,

H_0 :The additional exploratory variable do not improve the model,

H_1 :The additional exploratory variable improve the model.

gave the results in the table below.

Table 12. *The Deviance Analysis.*

Model	Residual. df	Residual. Dev	df	Deviance	Pr($>$ |Chi|)
Null model	399	495.531	–	–	–
Full Model	384	94.732	12	420	0.001871

Based on the analysis of deviance statistics as indicated in Table 12, the comparison of two different models to assess the impact of inclusion of more predictor variables was done. The resulting p-value was obtained as 0.001871, which was less than the 0.05 level of significance and therefore we reject the null hypothesis and conclude that the additional exploratory variable improve the model. This suggests that the full model together with additional parameters significantly improves the fit of the model to the data.

5. Conclusion

Based on the study, Poisson Generalized Linear Model was applied to model Malaria incidences in Mbita Sub-county. The Response variable was count and independent variables being categorical thus leading to the use of PGLM. Parameters of Poisson Generalized Linear Model were estimated in reference to a certain baseline being selected for each and every independent variable due to the categorical nature. The significance of each and every independent variable was determined based on the p-value.

Assessing the association between the exploratory variables and the response variable was determine by use of the chi-square test. A chi-square test was conducted on each and every independent variable and the response variable. The results indicated that there was an association between mosquito bed net use and malaria, age group and malaria and education level of family head and malaria and finally there was an association between healthcare access and malaria.

To determine the goodness of fit the deviance statistic was applied. Based on the tested hypothesis to determine whether The fitted model provides an adequate fit to the data, a comparison was done between the null model and the full model. The resulting p-value was obtained as 0.001871, which is less than the 0.05 level of significance and therefore we

reject the null hypothesis and conclude that the additional exploratory variable improve the model.

ORCID

0009-0000-6010-2525 (Samuel Okinyi)
0009-0000-8033-9175 (Joseph Eyang'an Esekon)
0000-0002-0449-3776 (Martin Mutwiri Kithinji)

Abbreviations

AR-GLM	Auto-Regressive Generalized Linear Model
df	Degrees of Freedom
ENSO	El-Nino-Southern Oscillation
GLM	Generalized Linear Model
GSARIMA	Generalized Seasonal Auto-regressive Integrated Moving Average
GWR	Geographically-Weighted Regression
HDSS	Health and Demographic Surveillance System
IPT	Intermittent Preventive Therapy
KEMRI	Kenya Medical Research Institute
P.falciparum	Plasmodium Falciparum
PGLM	Poisson Generalized Linear Model
P.malariae	Plasmodium Malariae
SEAs	Sample Enumeration Areas
VIF	Variance Inflation Factor

Author Contributions

Samuel Okinyi: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing.

Joseph Eyang'an Esekon: Supervision, Project administration.

Martin Mutwiri Kithinji: Supervision, Project administration.

Acknowledgments

We wish to express our sincere gratitude to Jane Awuor Kimori, Roseline Achieng Oburu and Kirinyaga University fraternity for their support.

Conflicts of Interest

We hereby declare no conflicts of interest.

References

[1] Amek, N., Bayoh, N., Hamel, M., Lindblade, K. A., Gimnig, J. E., Odhiambo, F., Laserson, K. F., Slutsker, L., Smith, T., and Vounatsou, P. (2012). Spatial and temporal dynamics of malaria transmission in rural western kenya. Parasites & vectors, 5(1): 86. https://doi.org/10.1186/1756-3305-5-86

[2] AAnyona, D. (2017). Effect of climatic variability on malaria trends in Baringo County, Kenya. Malaria Journal. https://doi.org/10.1186/s12936-017-1848-2

[3] Armstrong, B. (1985). Measurement error in the generalised linear model. Communications in Statistics-Simulation and Computation, 14(3), 529-544. https://doi.org/10.1080/03610918508812457

[4] Atkinson, P., Jiskoot, H., Massari, R., & Murray, T. (1998). Generalized linear modelling in geomorphology. Earth Surface Processes and Landforms: The Journal of the British Geomorphological Group, 23(13), 1185-1195. https://doi.org/10.1002/(SICI)1096-9837(199812)23:13<1185::AID-ESP928>3.0.CO;2-W

[5] Ayele, D. G., Mohammed, M. O. M., Abdallah, A. S. R., and Wacho, G. A.(2024).Assessment of malaria transmission in kenya using multilevel logistic regression. Heliyon, 10(21).

[6] Ayele, D. G., Zewotir, T. T., & Mwambi, H. G. (2012). Prevalence and risk factors of malaria in Ethiopia. Malaria journal, 11(1), 195. https://doi.org/10.1186/1475-2875-11-195

[7] Ayinde, K. (2012). Estimators of linear regression model and prediction under some assumptions violation. Open Journal of Statistics. http://dx.doi.org/10.4236/ojs.2012.25069

[8] Boateng, E. Y., Anyormi, G. E., Otoo, J., & Abaye, D. A. (2018, December). Drivers of intermittent preventive treatment of malaria during pregnancy in Ghana: a generalized linear model with negative binomial approach. In Applied Informatics (Vol. 5, No. 1, pp. 1-12). SpringerOpen. https://doi.org/10.1186/s40535-018-0057-6

[9] Briët, O. J., Amerasinghe, P. H., & Vounatsou, P. (2013). Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. PloS one, 8(6), e65761. https://doi.org/10.1371/journal.pone.0065761

[10] Chaves, L. F., Hashizume, M., Satake, A., & Minakawa, N. (2012). Regime shifts and heterogeneous trends in malaria time series from Western Kenya Highlands. Parasitology, 139(1), 14-25. https://doi.org/10.1017/S0031182011001685

[11] Colón-González, F. J., Tompkins, A. M., Biondi, R., Bizimana, J. P., and Namanya, D. B. (2016). Assessing the effects of air temperature and rainfall on malaria incidence: an epidemiological study across rwanda and uganda. Geospatial health, 11(S1):18-37.

[12] Diao, O., Absil, P. A., & Diallo, M. (2023). Generalized linear models to forecast malaria incidence in three endemic regions of Senegal. International Journal of Environmental Research and Public Health, 20(13), 6303. https://doi.org/10.3390/ijerph20136303

[13] Ernst, A. F., & Albers, C. J. (2017). Regression assumptions in clinical psychology research practice systematic review of common misconceptions. PeerJ, 5, e3323. https://doi.org/10.7717/peerj.3323

[14] Escalante, A. A., Cornejo, O. E., Freeland, D. E., Poe, A. C., Durrego, E., Collins, W. E., & Lal, A. A. (2005). A monkey's tale: the origin of Plasmodium vivax as a human malaria parasite. Proceedings of the National Academy of Sciences, 102(6), 1980-1985. https://doi.org/10.1073/pnas.0409652102

[15] Gopal, S., Ma, Y., Xin, C., Pitts, J., & Were, L. (2019). Characterizing the spatial determinants and prevention of malaria in Kenya. International journal of environmental research and public health, 16(24), 5078. https://doi.org/10.3390/ijerph16245078

[16] Hemingway, J., & Craig, A. (2004). New ways to control malaria. Science, 303(5666), 1984-1985. https://doi.org/10.1126/science.1096548

[17] Hickey, G. L., Kontopantelis, E., Takkenberg, J. J., & Beyersdorf, F. (2019). Statistical primer: checking model assumptions with regression diagnostics. Interactive cardiovascular and thoracic surgery, 28(1), 1-8. https://doi.org/10.1093/icvts/ivy207

[18] Homan, T., Maire, N., Hiscox, A., Di Pasquale, A., Kiche, I., Onoka, K., ... & Takken, W. (2016). Spatially variable risk factors for malaria in a geographically heterogeneous landscape, western Kenya: an explorative study. Malaria Journal, 15(1), 1. https://doi.org/10.1186/s12936-015-1044-1

[19] Jumi, L. G. L. (2017). Generalized linear models of malaria incidence in Jubek State, South Sudan. Science Journal of Applied Mathematics and Statistics, 5(4), 134-138. https://doi.org/10.11648/j.sjams.20170504.12

[20] Keeling, P. J., & Rayner, J. C. (2015). The origins of malaria: there are more things in heaven and earth?. Parasitology, 142(S1), S16-S25. https://doi.org/10.1017/S0031182014000766

[21] Khagayi, S., Desai, M., Amek, N., Were, V., Onyango, E. D., Odero, C., ... & Vounatsou, P. (2019). Modelling the relationship between malaria prevalence as a measure of transmission and mortality across age groups. Malaria journal, 18(1), 247. https://doi.org/10.1186/s12936-019-2869-9

[22] Kleinschmidt, I., Sharp, B. L., Clarke, G. P. Y., Curtis, B., & Fraser, C. (2001). Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal, South Africa. American journal of epidemiology, 153(12), 1213-1221. https://doi.org/10.1093/aje/153.12.1213

[23] Kutegeka, S., Wamwea, C., & Waititu, A. (2025). Modeling the effect of climatic factors on malaria incidences using an auto-regressive generalized linear model with lagged covariates. Commun. Math. Biol. Neurosci., 2025, Article-ID.

[24] Liu, W., Li, Y., Shaw, K. S., Learn, G. H., Plenderleith, L. J., Malenke, J. A., ... & Sharp, P. M. (2014). African origin of the malaria parasite Plasmodium vivax. Nature communications, 5(1), 3346. https://doi.org/10.1038/ncomms4346

[25] Loy, D. E., Liu, W., Li, Y., Learn, G. H., Plenderleith, L. J., Sundararaman, S. A., ... & Hahn, B. H. (2017). Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. International journal for parasitology, 47(2-3), 87-97. https://doi.org/10.1016/j.ijpara.2016.05.008

[26] Mategula, D., & Gichuki, J. (2023). Understanding the fine-scale heterogeneity and spatial drivers of malaria transmission in Kenya using model-based geostatistical methods. PLOS Global Public Health, 3(12), e0002260. https://doi.org/10.1371/journal.pgph.0002260

[27] Midekisa, A., Senay, G., Henebry, G. M., Semuniguse, P., & Wimberly, M. C. (2012). Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. Malaria journal, 11(1), 165. https://doi.org/10.1186/1475-2875-11-165

[28] Mogeni, P., Williams, T. N., Fegan, G., Nyundo, C., Bauni, E., Mwai, K., ... & Bejon, P. (2016). Age, spatial, and temporal variations in hospital admissions with malaria in Kilifi County, Kenya: a 25-year longitudinal observational study. PLoS medicine, 13(6), e1002047.2 https://doi.org/10.1371/journal.pmed.1002047

[29] Nyawanda, B. O., Beloconi, A., Khagayi, S., Bigogo, G., Obor, D., Otieno, N. A., ... & Vounatsou, P. (2023). The relative effect of climate variability on malaria incidence after scale-up of interventions in western Kenya: A time-series analysis of monthly incidence data from 2008 to

2019. Parasite Epidemiology and Control, 21, e00297. https://doi.org/10.1016/j.parepi.2023.e00297

[30] Nzabakiriraho, J. D., & Gayawan, E. (2021). Geostatistical modeling of malaria prevalence among under-five children in Rwanda. BMC Public Health, 21(1), 369. https://doi.org/10.1186/s12889-021-10305-x

[31] Okunlola, O. A., Oyeyemi, O. T., & Lukman, A. F. (2021). Modeling the relationship between malaria prevalence and insecticide-treated bed net coverage in Nigeria using a Bayesian spatial generalized linear mixed model with a Leroux prior. Epidemiology and Health, 43, e2021041. https://doi.org/10.4178/epih.e2021041

[32] Rich, S. M., & Ayala, F. J. (2006). Evolutionary origins of human malaria parasites. In Malaria: genetic and evolutionary aspects (pp. 125-146). Boston, MA: Springer US. https://doi.org/doi.org/10.1007/0-387-28295-5-6

[33] Sewe, M., Rocklöv, J., Williamson, J., Hamel, M., Nyaguara, A., Odhiambo, F., & Laserson, K. (2015). The association of weather variability and under five malaria mortality in KEMRI/CDC HDSS in Western Kenya 2003 to 2008: a time series analysis. International journal of environmental research and public health, 12(2), 1983-1997. https://doi.org/10.3390/ijerph120201983

[34] Sewe, M. O., Ahlm, C., & Rocklöv, J. (2017). Remotely sensed environmental conditions and malaria mortality in three malaria endemic regions in western Kenya. In American Journal of Tropical Medicine and Hygiene (Vol. 95, No. 5, pp. 119-119). American Society of Tropical Medicine and Hygiene. https://doi.org/10.1371/journal.pone.0154204

[35] Snow, R. W., Kibuchi, E., Karuri, S. W., Sang, G., Gitonga, C. W., Mwandawiro, C., ... & Noor, A. M. (2015). Changing malaria prevalence on the Kenyan coast since 1974: climate, drugs and vector control. PloS one, 10(6), e0128792. https://doi.org/10.1371/journal.pone.0128792

[36] Uyanik, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. Procedia-Social and Behavioral Sciences, 106, 234-240.

[37] Zou, Y., Geedipally, S. R., & Lord, D. (2013). Evaluating the double Poisson generalized linear model. Accident Analysis & Prevention, 59, 497-505. https://doi.org/10.1016/j.aap.2013.07.017