Prevalence of and factors associated with malaria in children under five years of age in Malawi, using malaria indicator survey data

Rugiranka Tony Gaston*, Shaun Ramroop

School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Pietermaritzburg Campus, Private Bag X01, Scottsville, 3209, South Africa

ARTICLE INFO

Keywords:
Malaria
Children under five years of age
GAMM
RDT
MMIS
Public health
Infectious disease
Nutrition

ABSTRACT

Background: Malaria remains a public health problem in developing countries and Malawi is no different. Although there has been an improvement in reducing malaria in Malawi, it remains a problem, especially in children less than five years old. The primary objective of the study was to assess whether socio-economic, geographic and demographic factors are associated with malaria, using the generalized additive mixed model (GAMM).

Data and methodology: The study used a 2017 dataset from the Malawi Malaria Indicator Survey (MMI) with a total number of 2724 children under five years old. The study also utilized the GAMM to analyze data. The outcome was that either the child had malaria or did not, as detected using the malaria Rapid Diagnostic Test (RDT) (Ayele et al., 2014a).

Results: In this study, more than 37% of the total number of children who were tested showed a positive malaria result. In addition, the results from this study using GAMM indicated that anaemia, mother's education level, wealth index, child's age, the altitude of the place of residence, region, place of residence, toilet facility and electricity were significantly associated with a positive malaria RDT.

Conclusion: The study revealed that socio-economic, geographical and demographic variables are the key factors in improving malaria vectors in children. Improving income levels and supporting the poorer rural community mostly from the Central Region would be a great achievement in reducing malaria vectors in Malawi. In addition, improving health care in rural areas, especially at higher altitudes, would contribute to controlling malaria and reducing anaemia.

1. Introduction

Malaria is a serious disease caused by a protozoan parasite called \textit{Plasmodium} species and the most dangerous is \textit{P. falciparum} which occurs mainly in Africa. There are four other different parasites which are not as dangerous as \textit{P. falciparum}, these being \textit{P. vivax}, \textit{P. ovale}, \textit{P. malariae} and \textit{P. knowlesi} (WHO, 2015). \textit{Plasmodium falciparum} is the most prevalent parasite in Africa especially in Sub-Saharan Africa and is estimated to be the cause of 99% of malaria in 2016, while outside of Africa \textit{P. vivax} is the most predominant vector (WHO, 2017).

The parasites that cause malaria are transmitted to humans through the bite of the female \textit{Anopheles} mosquito and it takes 10–15 days to develop symptoms of the disease after being infected (Perkins et al., 2011). Malaria is more often transmitted during the rainy season and with higher temperatures. The disease mostly affects the poorest countries (Mhunu, 2005; Chirombo et al., 2014; Cibulskis et al., 2016). Malaria is not contagious; however, it is possible to contract the disease from another person through blood transfusions or organ transplants (WHO, 2016).

Despite the interventions and precautions taken against malaria, the disease remains a major health problem worldwide, especially in developing countries (WHO, 2016). Malaria is among the leading causes of morbidity and mortality especially in the Sub-Saharan Africa countries and the most vulnerable are pregnant women and children (WHO, 2013; Semakula et al., 2016). The global estimate of malaria cases was 237 million, 211 million and 216 million in 2010, 2015 and 2016 respectively, with Africa having the most cases each year (WHO, 2016). This shows that there has been a reduction in malaria cases generally, but with, a slight increase globally between 2015 and 2016.

The estimated number of deaths from malaria globally was 655 000, 446 000 and 445 000 in 2010, 2015 and 2016 respectively. It was estimated that approximately 80% of these deaths were from Sub-Saharan
African countries in 2015, while in 2016 the figure was estimated as 91%. Moreover, 70% of all those who died in 2016 were children under five years old (WHO, 2015; WHO, 2016).

Malaria was the fourth highest cause of mortality in Sub-Saharan Africa, accounting for 10% of children’s deaths. This is the equivalent of approximately one child in Sub-Saharan Africa dying of malaria every two minutes (WHO, 2015). In Malawi, as one of the Sub-Saharan Africa countries, where malaria is endemic throughout the country, the disease remains a health problem. In this country, malaria most often affects individuals who live in the rural areas, which are hotter, wetter, more humid and tend to be low-lying, rather than those who live in the dry, urban, highland areas (Kazembe et al., 2006; Dzinjalimala, 2009; Chirombo et al., 2014; Kazembe and Mathanga, 2016).

In 2013, malaria in Malawi was the leading cause of hospital admissions and death in children under five years of age and pregnant women. The disease accounted for 20% of all deaths of children under five (WHO, 2016). The prevalence of malaria in Malawi has decreased from 43% in 2010 to 33% in 2014 and 24% in 2018 (NMCP and ICF, 2010; NMCP and ICF, 2014; NMCP and ICF, 2018). This reduction was due to the efforts of the Malawian government and the international sponsors who put more resources into fighting the burden of malaria among children under five years of age (Mathanga et al., 2012). In 2010, the Malawian government introduced indoor residual spraying (IRS) in various districts around the country as one of the methods of reducing malaria. In 2012 to 2014, the government of Malawi and the sponsors distributed free long-lasting insecticide-treated nets (LLIN) to the children and pregnant women in the whole country (Mathanga et al., 2012; Chanda et al., 2016). Although there has been a great reduction in malaria in Malawi, the disease remains a health problem, especially in children under five (NMCP and ICF, 2012; NMCP and ICF, 2018).

The study by Zgambo et al. (2017) used the 2012 and 2014 Malawi Malaria Indicator Surveys (MMIS) to compare the prevalence of and factors associated with malaria parasitaemia in children under five years of age in Malawi. Their findings showed that the prevalence of malaria had increased from 28% in 2012 to 33% in 2014. This reveals that malaria is still a health problem in the country and more research is needed using different methods to identify the risk factors associated with the disease, especially for children under five years of age in Malawi.

Post-2000 research about malaria in Malawi such as that by Buchwald et al. (2016); Chitungu and Musenge (2016); Kazembe et al. (2006); Lazzanini et al. (2016); Hershey et al. (2017); Kabaghe et al. (2017); and Parvin et al. (2018) have used different parametric methods, such as logistic regression, generalized linear mixed models and other statistical models. These models are an amazing asset in modelling the relationship between the outcome variable and covariates. However, in numerous applications, this relationship between the outcome and some confounding covariates may have an unknown function form (Ayele et al., 2014a, b). Therefore, along these lines, with such kinds of parameters, it is very important to estimate non-parametrically. This prompted an investigation into non-parametric methods, which include semi-parametric additive models. Thus, the current study used the generalized additive mixed model (GAMM) to overcome these challenges. The GAMM is an extension of the generalized additive model (GAM), which includes the random effect. The random effect is used to model the correlation between observations. The GAM does not include random effect, only the model covariate effect (Wang, 1998; Lin and Zhang, 1999). The GAMM is also an extension of the generalized linear mixed model (GLMM) with a parametric model introduced by Breslow and Clayton (1993). The GAMM enables the parametric fixed effects from GLMM to be modelled as a non-parametric model using the additive smooth function (Hastie and Tibshirani, 1990).

According to our knowledge, researchers have not used the generalized additive mixed model (GAMM) nation-wide to identify the factors associated with malaria in children under five years of age in Malawi. The study by Chirombo et al. (2014) used structured additive regression models, which include GAMM and the geo-additive model in their research in Malawi for children under five years of age, using the 2010 MMIS data set. However, it must be stressed that the data set used is different from the 2017 MMIS data set used in the current study.

Therefore, this study aims to investigate the prevalence of and factors associated with malaria in children under five years old, using the MMIS for 2017 with the application of GAMM. The study also investigates the risk factors for malaria and whether or not they had remained the same after the study by Chirombo et al. (2014).

2. Methodology and material

2.1. Study area

Malawi is a Sub-Saharan African country situated south of the equator and is bordered by Tanzania in the north and northeast; by Mozambique to the east and southwest; and Zambia to the west and northwest (NMCP and ICF, 2018). The total area is approximately 118 484 square kilometres, in which 9 4276 square kilometres are land and the remaining area consists of Lake Malawi. The country is split into three regions and twenty-eight districts. The Northern region has six districts, the Central region nine and the Southern region comprises thirteen districts (NMCP and ICF, 2018). Malawi has a tropical continental climate with sea influences, where the variation of rain and temperature depends on altitude and proximity to Lake Malawi. The tropical climate is favourable for the breeding of Anopheles Mosquitoes and the breeding of Anopheles increases in the rainy season from November to April. The weather in Malawi becomes cool and dry from May to August and the transmission of malaria is not as high as in the rainy season (Kazembe, 2007; NMCP and ICF, 2012). The economy of Malawi is based on agriculture and it is one of the poorest countries in the world. Healthcare is poor compared to other African countries (WHO, 2016; Team, 2018).

2.2. Data sources

The study used secondary data from the 2017 Malawi Malaria Indicator Survey (MMIS) and was collected between 15 April and June 2017. The MMIS was implemented by the Malawi National Malaria Control Program (NMCP) through support from the President’s Malaria Initiative (PMI). The United States Agency for International Development (USAID) provided financial support through the President’s Malaria Initiative (PMI). They also funded the project by offering technical assistance in the implementation of population and health surveys as they do in countries worldwide (NMCP and ICF, 2018). The governing body of Malawi provided staff, office space and strategic help. Thereafter, the ICF provided technical support through the Demographic and Health Survey (DHS) program.

The 2017 MMIS data interviewed all residents or visitors who stayed in the selected households the night before the interview. The survey was population-based on a household cluster survey. The data sampling followed the two-stage sampling method. The first stage included a selection of 150 clusters from the enumeration areas (EAs) demarcated in the 2008 population and housing census. Out of the 150 clusters, 60 were selected from urban and 90 from rural areas. The second stage sampling involved the systematic selection of a sample of 3 750 households. Out of the 3 750 households, 25 households were selected from each enumeration area (EA). All women aged 15–49 years who were living in or had visited the selected household the night before the survey were eligible to be interviewed. The children aged 6–59 months in these households were tested for malaria infection with the consent of their parents or guardians. The study used a total number of 2 724 children as a weighted sample in order to ascertain a national-level representation (NMCP and ICF, 2018). The survey sample of this study is representative at the national and regional level, as well as for urban and rural areas.

In the sampling process, the number of women surveyed in each region should contribute to the size of the total sample in proportion to the
size of the region (NMCP and ICF, 2018). However, some regions may have small populations, and this unweighted distribution does not represent the exact population. To resolve this problem, regions with small populations are oversampled. Thus, the weighted sample used in the study to get statistics that are representative of the country and to account for the complex sample design from MMIS data set (NMCP and ICF, 2018). The SD Bioline Malaria Ag P.f/P, a rapid diagnostic test (RDT) was used to collect a blood sample from children’s finger or heel-prick. This test is appropriate in detecting the histidine-rich protein II (HRP II), an antigen of Plasmodium falciparum and common Plasmodium lactate dehydrogenase (pLDH) of Plasmodium species in human blood (NMCP and ICF, 2018). The diagnostic test incorporates an expendable sample tool that comes in a standard package. A tiny volume of blood is caught on the instrument and placed in the well of a testing device. The RDTs for malaria offers the possibility to expand the arrangement of exact malaria diagnosis to the region where microscopy services are not accessible; for example in a remote area or after standard laboratory exact malaria diagnosis to the region where microscopy services are not caught on the instrument and placed in the well of a testing device. The sample tool that comes in a standard package. A tiny volume of blood is (NMCP and ICF, 2018) The diagnostic test incorporates an expendable sample tool that comes in a standard package. A tiny volume of blood is

\[ \beta_i + \beta_2 x_2 + \ldots + \beta_k x_k. \]

(1)

Eq. (1) was replaced with the additive form and hence, the general equation for GLM as \( \sum \beta_1 x_1 \) becomes \( \sum f(x) \) in GAM. Thus, the equation of GAM is written as follows:

\[ g(\mu_i) = 6x + f_1(x_1) + f_2(x_2) + f_3(x_3) + \ldots + f_n(x_n). \]

(2)

From Eq. (1), \( \mu_i = E(\gamma_i) \) and \( x_j \) distributed some exponential family distribution, \( X_j \) is the designed matrix, while \( \theta \) is the corresponding parameter vector and \( f_i \) are the smooth functions of covariates, while \( g(.) \) is the monotonic differentiable function (Wood, 2017). If there is no linear component in Eq. (2), the model is known as nonparametric, whilst the models whose predictors have both linear and unspecified nonlinear function are semiparametric. To estimate the parameters, the standardized condition of the smooth functions \( f_i \) should be satisfied such that \( E(\gamma_j) = 0 \), apart from that, each function will have free constants (Hastie and Tibshirani, 1990). When the data has repeated measurement or correlations, the model includes a random variable and this leads to the extension of GAM. Hence, GAM becomes the generalized additive mixed model (GAMM) in the same way as the generalized linear mixed models GLMM are an extension of GLM (Hastie and Tibshirani, 1990). The GAMM was introduced by Breslow and Clayton (1993) to include the random effect in the GAM and model the correlation between the observations. The equation of GAMM can be expressed as follows:

\[ g(\mu_i) = \beta_i + f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n) + Z \beta, \]

(3)

where \( g(.) \) is monotonic differentiable link function, \( X_i \equiv (1, x_{i1}, \ldots, x_{ik}) \) are n covariate associated with fixed effects and \( q \times 1 \) vector of covariates \( Z_i \) associated with random effects. Thus, the given \( q \times 1 \) vector of random effect \( b \), the observations \( y_i \) are assumed to be conditionally independent with means \( E(y_i | b) = \mu_i \) and variance, \( \gamma_i (b) = \gamma \psi \), where \( \gamma(.) \) is the specified variance function and \( \psi \) is a scale parameter. Moreover, \( f_i(.) \) is a centred twice differentiable smooth function and the random effects \( b \) is assumed to be distributed as \( N(0, \psi \gamma (\mu)) \) and \( \rho \) is a \( c \times 1 \) vector of variance components. In addition, when \( f_i \) is a linear function, the GAMM reduces to GLMM (Lin and Zhang, 1999). For a specified variance component, \( \theta \) the log-likelihood function of \( (\beta, f, \theta) \) is expressed by Lin and Zhang (1999) in the following equation:
where \( y_i \) and \( d_i(y, \mu) \) define the conditional deviance function of \((\beta, f_i, \theta)\) given \( b \). The statistical inference for GAMM on nonparametric function \( f_i \) requires the estimate of smoothing parameter \( \tau \) and the inference on variance component \( \theta \). It is known that the smoothing spline estimators and linear mixed models have close connections (Green and Silverman, 1993; Wang, 1998; Lin and Zhang, 1999). Moreover, the natural cubic smoothing spline estimators of function \( f_i \) maximize the penalized log-likelihood for the same given \( \tau \) and \( \theta \) and give the following equation:

\[
\exp \left[ \sum_{i=1}^{k} \frac{1}{2} \int_{\mathcal{H}} y_i - \mu_{i, \theta} \, dx \right] = \exp \left[ \int_{\mathcal{H}} \frac{1}{2} \sum_{i=1}^{k} \tau_i \, f_i(x) \, dx \right]
\]

\[
\text{exp} \left[ \{y; \beta_0, f_1, \ldots, f_k, \theta]\right] = \text{exp} \left\{- \frac{1}{2} \sum_{i=1}^{k} \tau_i \, f_i(x) \, dx \right\}
\]

where the \( s_i \) and \( t_i \) indicate the range of \( b^i \) covariate and \( t_i \) are the smoothing parameters that manage the trade-off between goodness of fit and the smoothness of the estimated functions (Lin and Zhang, 1999; Ayele et al., 2014a). Moreover, \( f_i(\cdot) \) is an \( n \times 1 \) unknown vector of the values of \( f_i(\cdot) \), estimated at \( n \) ordered values of the \( x_i \), where \( k = 1, \ldots, n \) and \( s_i \) is the smoothing matrix (Green and Silverman, 1993). By using the matrix form, the GAMM given in Eq. (3), can be written as:

\[
g(\mu_i) = \beta_0 + \sum_{j=1}^{k} \tau_j \, f_j(x_i) + Zb.
\]

where \( g(\mu_i) = [g(\mu_1), \ldots, g(\mu_n)] \), \( n \times 1 \) the vector of ones, \( N_i = m \times r \) matrix, such that the \( k^\text{th} \) component of \( Nf_i \) is \( f_i(x_i) \) and \( Z = Z_1, Z_2, \ldots, Z_n \). In order to evaluate Eq. (7), the numerical integration is required. Additionally, to calculate the natural cubic smoothing spline estimators of \( f_i \) by maximizing Eq. (6) is sometimes complicated. Consequently, Lin and Zhang, (1999) resolved this problem by suggesting the double penalized quasi-likelihood (DPQL) model as an alternative approach. Hence, the estimation of nonparametric function \( f_i \) can be obtained by applying double quasi-likelihood. The function \( f_i \) is re-parameterized in terms of \( \beta_i \) and \( a_0 \) in one-to-one transformation as:

\[
f_i = X_i^\tau \hat{\beta}_i + \beta_0 X_i a_0.
\]

where \( X_i \) is \( n \times 1 \) vector with the \( r_j \) centred and ordered distinct values of the \( x_i (k = 1, 2, \ldots, n) \) and \( \beta_0 \) is \( L\left(I_{n-k}\right)^{-1} \) and \( I_k \) is an \( r_j (r_j - 2) \) full rank.

### Table 1. Childhood malaria by categorical variables.

| Variable                      | Category       | Malaria (Positive) | Malaria (Negative) | p-value |
|-------------------------------|----------------|-------------------|-------------------|---------|
| Region                        | North          | 62 (20.8%)        | 236 (79.2%)       | 0.000   |
|                               | Central        | 452 (39.8%)       | 683 (60.2%)       |         |
|                               | South          | 500 (38.7%)       | 791 (61.3%)       |         |
| Place of residence            | Rural          | 987 (41.9%)       | 1369 (58.1%)      | 0.000   |
|                               | Urban          | 27 (7.3%)         | 341 (92.7%)       |         |
| Wealth index                  | Poorer         | 611 (48.0%)       | 662 (52.0%)       | 0.000   |
|                               | Middle         | 204 (39.6%)       | 311 (60.4%)       |         |
|                               | Richer         | 199 (21.3%)       | 737 (78.7%)       |         |
| Mother's education level      | No education   | 152 (46.3%)       | 176 (53.7%)       | 0.000   |
|                               | Primary        | 557 (37.5%)       | 930 (62.5%)       |         |
|                               | Secondary      | 61 (17.4%)        | 290 (82.6%)       |         |
|                               | Tertiary       | 1 (3.6%)          | 27 (96.4%)        |         |
| Child's age in months         | 6-23           | 200 (25.6%)       | 581 (74.4%)       | 0.000   |
|                               | 24-31          | 411 (40.6%)       | 602 (59.4%)       |         |
|                               | 32-59          | 403 (43.4%)       | 526 (56.6%)       |         |
| Anemia level                  | Anemic         | 706 (45.6%)       | 843 (54.4%)       | 0.000   |
|                               | Not anemic     | 307 (26.1%)       | 867 (73.9%)       |         |
| Sex of the child              | Male           | 463 (36.5%)       | 804 (63.5%)       | 0.001   |
|                               | Female         | 447 (35.8%)       | 800 (64.2%)       |         |
| Altitude                      | 0-500metres    | 119 (34.3%)       | 226 (65.7%)       | 0.000   |
|                               | 501-1000m      | 492 (43.8%)       | 631 (56.2%)       |         |
|                               | >1000          | 403 (32.2%)       | 850 (67.8%)       |         |
| Toilet facility               | Toilet with flush | 88 (21.1%)   | 330 (78.9%)       | 0.000   |
|                               | Pit latrine    | 832 (39.3%)       | 1287 (60.7%)      |         |
|                               | No facility    | 95 (50.8%)        | 92 (49.2%)        |         |
| Electricity                   | Yes            | 21 (7.3%)         | 265 (92.7%)       | 0.000   |
|                               | No             | 992 (40.7%)       | 1445 (59.3%)      |         |
| Main roof material            | Thatch/Palm leaf | 686 (45.8%)   | 812 (54.2%)       | 0.000   |
|                               | Corrugated &Metal | 283 (25.8%)   | 816 (74.2%)       |         |
|                               | Stick & Mud    | 45 (35.7%)        | 81 (64.3%)        |         |
| Main wall material            | Wood/Mud      | 331 (43.2%)       | 436 (56.8%)       | 0.000   |
|                               | Bricks         | 467 (40.1%)       | 697 (59.9%)       |         |
|                               | Cement/block   | 216 (27.2%)       | 577 (72.8%)       |         |
| Main floor material           | Earth/Sand    | 840 (43.3%)       | 1092 (56.5%)      | 0.000   |
|                               | Mud block/wood | 74 (39.4%)        | 114 (60.6%)       |         |
|                               | Cement/block   | 100 (16.6%)       | 504 (83.4%)       |         |
| Sleep under mosquito bed net  | All children   | 553 (33.6%)       | 1093 (66.4%)      | 0.000   |
|                               | Some           | 88 (34.1%)        | 170 (65.9%)       |         |
|                               | None           | 372 (45.4%)       | 447 (54.6%)       |         |
matrix satisfying \( S_k = LL' \) and \( L_k x_k = 0 \). Thus, the double penalized quasi-likelihood with respect to \((\beta, f)\) and \( b \) becomes:

\[
-\frac{1}{2t} \sum_{i=1}^{n} [d_i(y_i) - \frac{1}{2} G_i^{-1} b - \frac{1}{2} D_i^{-1} d_i] (8)
\]

where \( f_i S_i f_k = a_i a_k, a = (a_1, a_2, \ldots, a_k) \) and \( D = \text{diag}(\rho_1, \rho_2, \ldots, \rho_k) \) with \( \rho_k = \frac{1}{k} \). Note that the small values of \( \rho_k = (\rho_1, \rho_2, \ldots, \rho_k) \) correspond to over smoothing (Breslow and Clayton, 1993; Lin and Zhang, 1999).

5. Results

The present study used survey-weighted data in order to ascertain a national level presentation (NMCP and ICF, 2018). The results from cross-tabulation analysis are summarized in Table 1. The results indicated that all independent variables were significantly associated with childhood malaria (p-value < 0.05). The age of child and altitude were categorized, however, in multivariate GAM are considered as continuous. The results from Table 1 indicated that the prevalence of malaria was 25%, 40.6%, and 43.4% among children aged between 6-23, 24-41 and 42-59 months respectively. The prevalence of malaria in terms of the sex of the child was 36.5% for a male child and 35.8% for a female child.

It was observed that the prevalence of malaria was higher in the Central Region (39.8%), followed by the South Region (38.7) and lastly the North Region (20.8%). The results also showed that the prevalence of malaria was higher in children from poorer households (48.0%), middle income households (39.6%) and wealthy households (21.3%) respectively. It was found that the prevalence of malaria among the children with anaemia (45.6%) was higher compared to those without anaemia (26.1%). The results also showed that the prevalence of malaria in children living in rural areas is higher (41.9%) than in children living in urban areas (7.3%). The prevalence of malaria was highest in children from mothers with no education (46.3%), followed by those whose mothers had primary (37.5%), secondary (17.4) and tertiary education (3.6%) respectively.

5.1. Model fitting

The multivariate study used R software to analyze the data with the application of “mgcv” packages. The GAMM was used to model the effect of age and altitude non-parametrically, while other covariates were used as parametric factors. These factors have a continuous effect and might of age and altitude non-parametrically, while other covariates were used. The study used shrinkage smoothers (splines) to fit the GAM model, due to advantages such as assisting to control the knot placement. Furthermore, the shrinkage smoother is constructed in such a way that the smooth terms are rebuffed away all around (Wood, 2006). The study also considered the fundamental impact and possible two-way interaction effect. The p-value of the individual smooth term and the AIC of each model, together with the inference of smooth were analyzed. The selection of the model was based on the smallest AIC, the higher value of degree of freedom and high statistical significance. Hence, the final model for this study is given in Eq. (9) as follows:

\[
g(\mu_j) = \beta_0 + \beta_1 \text{Anemia}_j + \beta_2 \text{Wealth}_j + \beta_3 \text{Region}_j + \beta_4 \text{Toilet}\_facility_j + \beta_5 \text{Electricity}_j + \beta_6 \text{Education}_j + \beta_7 \text{Residence}_j + f_1 (\text{Age}_j) + f_2 (\text{Altitude}_j) + b_{ij} \tag{9}
\]

where, \( g(\mu_j) \) is the logit link function, \( \beta s \) are the parametric regression coefficients, \( f_1 \) and \( f_2 \) are centred smooth functions, while \( b_{ij} \) is the random effects, which can be written as \( b_{ij} \sim N(0, G(\theta)) \).

5.2. Interpretation of results

The results in model (9) are presented in Tables 2 and 3 and in Figure 1. Table 2 indicates the parameter estimates for the model, standard error, z-value, odds ratio and p-values. The study reported the variables with significant impact for malaria using a RDT such as anaemia, electricity, region, residence, wealth index, the toilet facilities and mother’s education status. The study checked all possible interactions. However, the two-way interaction effect was not included, since it did not add any significant effect to the model with non-significant p-values.

Table 2 shows the children with no anaemia were 0.233 times less likely to test positive for malaria using a RDT as compared with anaemic children. The results also revealed that the odds of positive malaria results in a RDT for children living in the Central Region were 1.936 times more likely than for those who lived in the North Region. Similarly, the odds of positive malaria in a RDT for children living in the South Region were 1.179 times more likely than for those who lived in the North Region. The children living in rural areas were 4.318 times more likely to test positive for malaria in terms of RDT results compared to those living in urban areas. The study also showed that the odds of positive malaria results in a RDT test for children from a household with no toilet facilities were 2.938 times more likely than those with flush toilets. Furthermore, the children from households with pit latrines were 1.389 times more likely to test positive for malaria in a RDT, compared to those with flush toilets. The results indicated that the children from the middle class were 0.743 times less likely to test positive for malaria using a RDT, compared to those from the poorer classes. In addition, the children from the wealthier classes were 0.571 times less likely to test positive for malaria in a RDT than those from the poorer classes. Lastly, the results indicated that the odds of positive malaria in a RDT for children from households with access to electricity were 0.435 times less likely than those from households with no access to electricity. Table 3 shows that the age of a child and the altitude of their region of residence has a significant impact on malaria using a RDT.

The letter \( S \) in Table 3 represents the smooth term and the number in parentheses shows the estimated degree of freedom (edf). The test statistics for child age and altitude of the region of residence (22.340; 90.420 respectively) together with their p-value (0.000; 0.000) shows that there is no linear trend associated either for child age or for altitude. This is confirmed in Figure 1, where the trend shows that the effect of malaria results in a RDT increases with age up to approximately 35 months and thereafter remains constant with no sign of decreasing. The same results indicate that the effect of malaria results in a RDT increases with the altitude of their region of residence up to approximately 750 m above sea level and above that starts decreasing.

6. Discussion

The present study utilized the generalized additive mixed model (GAMM) to investigate the risk factors associated with malaria using nationwide malaria survey data from Malawi. The previous studies had used the parametric models such as the generalized linear mixed model (GLMM) to analyze the malaria results using RDT (Ayele et al., 2013;...
The parametric models are useful to model the relationship between a response variable and covariance. However, non-parametric models are flexible enough to allow non-normal error distributions, modelling continuous predictor variable, relaxes the assumption of normality and linearity in linear regression (Lin and Zhang, 1999). The parametric and non-parametric models should complement each other, and for this reason the combination of the two methods is more useful (Wu and Zhang, 2006). Thus, the study first used the parametric model to create geographical and social-economic status variables such as type of place of residence, region, wealth quintile, mother’s highest education level, type of toilet facility and availability of electricity.

The effect of age and altitude was modelled as non-parametric and was statistically significant. The interaction effect was not included in the model as it was not statistically significant to improve the original model. The results from the parametric part revealed that the probability of increasing a positive malaria RDT was lower in the richer and the middle classes compared to the poorer class. These results confirmed that the prevalence of malaria is linked to socio-economic conditions, where poorer people are more vulnerable (Hay et al., 2004; Chitunhu and...
Musenge, 2016). This is due to the limited access to healthcare and the affordability of treatment (Worrall et al., 2002). The study revealed that the households with access to electricity are less likely to increase the positive malaria RDT rates. Moreover, the households with no toilet facilities are more likely to increase the positive malaria RDT rates. This shows that the households with access to electricity and toilet facilities can more easily access the healthcare and afford the treatment (Hay et al., 2004). Hence, these factors that are indicators of socio-economic status are consistent with the study by Ayele et al. (2014a).

The results from the study also showed that the risk of malaria is lower in children from mothers with higher education. This might be linked to socio-economic status, as educated individuals are more likely to have a better standard of living and better understand health-related issues. Furthermore, the individuals with higher levels of education can more easily access healthcare and afford the use of mosquito nets, indoor residual spray and other preventive measures for malaria. These results were consistent with previous studies such as those by Snyman et al. (2015); Zgambo et al. (2017) and Sultana et al. (2017).

The study revealed that the households from rural areas have a higher prevalence of testing positive to malaria compared to those from urban areas. This may be explained by the individuals living in rural areas not having the same access to the many things that their urban counterparts have; such as proper formal houses, drinking water, education opportunities, access to health care and so forth. The individuals living in rural areas often drink water from rivers which may attract mosquitoes, as river water is often dirty. Moreover, it takes longer to pass through the bush to reach these rivers, which may increase susceptibility to mosquito bites and therefore contracting malaria. In addition, most of the individuals living in rural areas live in poor housing conditions. A particular issue could be holes in the walls of the houses where the Anopheles Mosquitoes could enter the houses increasing the chance of malaria transmission through its bite (Iwetoijera et al., 2013; Jenkins et al., 2015; Kazembe and Mathanga, 2016; Sultana et al., 2017).

The study indicated that children without anaemia have a lower prevalence of testing positive to malaria compared to that of anaemic children. This might be explained by the link between anaemia and malaria, as has been shown by previous studies, such as those of Biemba et al. (2000) and Sultana et al. (2017).

The study indicates a large variation among the three regions, where the households from the Central Region being most likely to test positive for malaria. This is due to the fact that the region is covered by a large plain of land and the low-lying zone along the lake. Moreover, the lake might be an area conducive to the breeding of malaria vectors (Minka et al., 2012; Zgambo et al., 2017).

The results from the non-parametric model indicate that the probability of a positive malaria RDT increases as the child’s age increases. This could be due, in part, to the impact of maternal immunity in the child before one year of age. In addition, the children younger than one year old are more protected and well taken care of and this helps to fight any kind of disease. This reduces as the children get older. These results are consistent with the studies by Ayele et al. (2014a) and Chirombo et al. (2014).

The research reveals that the risk of having a positive malaria RDT result increases as the altitude increases, up to 750 m, and starts showing a decrease as altitude increases thereafter. This may be explained by the very high temperatures at lower altitudes as mosquitoes develop in hotter areas. As the altitude increases, the temperatures decrease and this reduces the risk of having a positive malaria RDT result (Lindsay and Martens, 1998; Chirombo et al., 2014).

7. Conclusion

The aim of this study was to assess the prevalence of and factors associated with malaria in children under the age of five years in Malawi using GMM. The current findings show that the government should consider other factors associated with malaria especially in children under five years of age; such as anaemia, region, residence type, toilet facilities, wealth index, the use of electricity, mothers’ education, children’s age and the altitude of the region of residence.

The findings from this study revealed that malaria is still a major problem and is linked to socio-economic factors as well as geographical location. The government should focus on poorer communities from rural and low altitude areas, especially in the Central Region, as their target group of individuals to educate, support and help change mindsets. In addition, children with anaemia should take priority in receiving the necessary health care and support. The key findings also show that there is a need to educate the population through workshops, mobile clinics and various social media platforms on how to prevent malaria in children under five years of age.

Moreover, the education of mothers should be considered and supported so that they can take better care of and protect their children, especially after the child's first six months from birth, as they are more likely to be exposed to malaria vectors.

The study will help the government and donors to control and possibly eliminate malaria in children under five years of age. The main focus should be on children with anaemia, mother’s education level, wealth index, children's age, the altitude of the place of residence, region, place of residence, toilet facility and electricity facilities. Furthermore, the model used in this study will help other researchers to compare findings.

Future research could use the joint model to model malaria and anaemia simultaneously in order to examine the possible correlation between the two diseases, as there is a link between them (Ayele et al., 2014b; Adebayo et al., 2016).

8. Limitation

The current data set was cross-sectional and consequently cannot address causality. It would have been ideal to have a longitudinal data set to study the change in factors and prevalence over time.

Declarations

Author contribution statement

R.T. Gaston: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

S. Ramroop: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by the Health Economics and HIV/AIDS Research Division (HEARD).

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

The authors acknowledge the National Malaria Control Programme and the ICF through the DHS programme for providing the data.
References

Adebayo, S.B., Gayawan, E., Heumann, C., Seiler, C., 2016. Joint modelling of anaemia and malaria in children under five in Nigeria. Spatial Spatio Temporal Epidemiol. 17, 105–115.

Alegana, V.A., Wright, J.A., Nahat, S.M., Butt, W., Sedigi, A.W., Habib, N., Snow, R.W., Atkinson, P.M., Noor, A.M., 2014. Modelling the incidence of Plasmodium vivax and Plasmodium falciparum malaria in Afghanistan 2006–2009. PloS One 9 (7), e102094.

Ayelo, D.G., Zezotir, T.T., Mwambi, H.G., 2013. The risk factor indicators of malaria in Ethiopia. Int. J. Med. Med. Sci. 5 (7), 335–347.

Ayelo, D.G., Zezotir, T.T., Mwambi, H.G., 2014. Semiparametric models for malaria rapid diagnosis test result. BMC Pub. Health. 14 (1), 31.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models: a Roughness Penalty Approach. CRC Press, London.

Butt, W., Sedigi, A.W., Noor, A.M., 2013. Mapping malaria transmission intensity in Malawi, 2000–2010. Am. J. Trop. Med. Hyg. 89 (5), 840–849.

Biamba, G., Dolmans, D., Thuma, P.E., Weis, G., Gordeuk, V.R., 2000. Severe anaemia in Zambian children with Plasmodium falciparum malaria. Trop. Med. Exp. Int. Health 5 (1), 9–16.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. 88 (422), 9–25.

Buchwald, A.G., Walldorf, J.A., Coome, L.M., Coalson, J.E., Chimbiyia, N., Bauleni, A., Nkanazera, K., Ngwira, A., Kapita-Tembo, A., Mathanga, D.P., Taylor, T.E., 2016. Bed net use among school-aged children after a universal bed net campaign in Malawi. Malar. J. 15 (1), 127.

Chanda, E., Mzilahowa, T., Chipwanya, J., Ali, D., Troell, P., Dodoli, W., Mnzava, A.P., 2009. The impact of malaria transmission in areas with high bed net coverage. Parasites Vectors 2 (1), 57.

Chilhoub, R.E., Alonso, P., Aponte, J., Aregawi, M., Barrette, A., Bergeron, L., Fergus, C.A., Knox, T., Lynch, M., Patouillard, E., Schwarte, S., 2016. Malaria: global progress 2000–2015 and future challenges. Infect. Dis. Poverty 5 (1), 61.

Chrombo, J., Love, R., Kazembe, L., 2014. Using structured regression models to estimate risk factors of malaria: analysis of 2010 Malawi malaria indicator survey data. PloS One 9 (7), e101116.

Chitunzu, S., Musereng, E., 2016. Spatial and socio-economic effects on malaria morbidity in children under 5 years in Malawi in 2012. Spatial Spatio-Temporal Epidemiol. 16, 21–35.

Dzinjalamala, F., 2009. Epidemiology of malaria in Malawi. Epidemiol. Malawi 203, 21.

Gaston, R.T., Ramroop, S., Habyarimana, F., 2018. Determinants of factors associated with Plasmodium falciparum malaria infection in Malawi. Trop. Med. Int. Health 23 (9), 834–3895.

Green, P.J., Silverman, B.W., 1993. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. CRC Press, London.

Hastie, T., Tibshirani, R., 1986. Generalized additive models. Stat. Sci. 1, 297–318.

Hastie, T.J., Tibshirani, R., 1990. Generalized Additive Models. CRC press, London.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M., Snow, R.W., 2005. The global distribution of malaria. Nature 434, 141–147.

Kazembe, L.N., Kleinschmidt, I., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.

Kazembe, L.N., Kleinseim, M., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.

Kazembe, L.N., Kleinseim, M., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.

Kazembe, L.N., Kleinseim, M., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.

Kazembe, L.N., Kleinseim, M., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.

Kazembe, L.N., Kleinseim, M., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.

Kazembe, L.N., Kleinseim, M., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.

Kazembe, L.N., Kleinseim, M., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.