Malaria Journal

RESEARCH

Open Access

Assessing the influence of socioeconomic and environmental variables on malaria risk in Nigerian children under 5 years: a GLMM approach

Talani Mhelembe[1*], Shaun Ramroop[1] and Faustin Habyarimana[1]

Abstract

Background The study focused on the full population of children from Nigeria, where the dataset was obtained from the demographic and health surveys (DHS). About 10245 children were selected for the current study and based on the rapid diagnostic test (RDT) results, there is about 37% prevalence of malaria in children under 5 years old in Nigeria. Malaria is the leading public health concern, that contributes to child mortality in the African region.

Methods The Nigeria Malaria Indicator Survey (NMIS) 2021 was utilized in this investigation. For the 2021 NMIS, a two-stage sampling technique was used. According to the NIMS study, the children chosen for anaemia and RDT testing were under 5 years of age.

Results A generalized linear mixed model (GLMM) was used to examine malaria RDT findings in conjunction with demographic, geographic, and socioeconomic characteristics. The following underlying risk factors for malaria in children were discovered in the study: altitude, anaemia level, age in months, fever status in the past 2 weeks, toilet facility, main wall material, main roof material, household wealth index, type of place of residence, sex of the child, mother's education level, and knowledge of the preventative measures that can be used to prevent malaria.

Conclusion Missing data were not deleted in this investigation; instead, multiple imputations utilizing chained equations were used to approximate the missing observation. Based on the results found by using the GLMM, the findings of this study may influence how the government combats malaria in Nigeria. The novelty of this study is that the missing values were not dropped. However, imputation techniques were explored, and multiple imputation by chained equations was used.

Keywords Malaria, Anaemia, GLMM, Nigeria, Children, Multiple imputation by chained equations

Background

Malaria is a major contributor to anaemia, with severe anaemia (defined as a haemoglobin level below 8 g per decilitre *(g/dL)*), as the main manifestation of complicated malaria [1, 2]. Both conditions are known to contribute to the huge burden of morbidity and mortality, especially among children under 5 years of age [3, 4]. According to the National Malaria Control Programme (NMCP) of Uganda, malaria alone has been shown

*Correspondence:
Talani Mhelembe
mhelembehht@gmail.com
[1] School of Mathematics, Statistics, and Computer Science, University of Kwazulu-Natal, Private Bag X01, PietermaritzburgScottsville 3209, South Africa

to contribute to between 30 to 50% of outpatient visits, 15–20% of hospital admission, and 20% of hospital deaths with most of this burden borne by children under 5 years and pregnant women [4]. Anaemia alone also affects over 50% of individuals in the same populations [4], making both illnesses' conditions of great public health concern. Malaria remains a public health concern for many developing countries, and Nigeria is no different. According to the World Health Organization (WHO), globally there were an estimated 247 million cases in 2021, an increase from 245 million in 2020, with most of these increases coming from countries in the WHO African Region [5]. Malaria case incidence reduced from 82 in 2000 to 57 in 2019, before increasing to 59 in 2020; there was no change in case incidence between 2020 and 2021 [5]. According to the WHO (5) report, it is believed that the increase in 2020 was associated with disruption to services during the COVID-19 pandemic.

Malaria is a disease caused by a protozoan parasite called *Plasmodium* species and the most prevalent is *Plasmodium falciparum* which occurs mainly in Africa [6]. *Plasmodium falciparum* is the most prevalent parasite in Africa, especially in sub-Saharan African regions [5, 7]. There are other parasites of malaria, but they are not as dangerous as the *P. falciparum* [7]. The parasites that cause malaria are transmitted to humans through the bite of the female *Anopheles* mosquito and it takes 10–15 days to develop symptoms of the disease after being infected [5, 8]. Malaria is more transmitted during the high temperature and rainy seasons. Malaria is not contagious, however, it is possible to contract the disease from another person through blood transfusions or organ transplants [5, 6]. There are several interventions and precautions taken against malaria, but the disease still remains a major health problem globally, especially in developing countries [5]. Twenty-nine countries accounted for 96% of malaria cases globally, and four countries; Nigeria (27%), Democratic Republic of the Congo (12%), Uganda (5%) and Mozambique (4%) accounted for almost half of all the cases globally [5]. The WHO African Region, with an estimated 234 million cases in 2021, accounted for about 95% of all cases. Globally, the malaria mortality rate halved from 30 in 2000 to 15 in 2015; it then continued to decrease but at a slower rate, falling to 14 in 2019. In 2020, the mortality rate increased again, to 15.1, before slightly decreasing to 14.8 in 2021 [5].

The main aim of the current study is to determine the malaria risk factors among children under 5 years of age from Nigeria. The method that will be used to achieve the aim is the generalized linear mixed model, that will be used to model data from the demographic health surveys; particularly the Nigeria Malaria Indicator Survey (NIMS) 2021.

Methods

Study area

This study uses the 2021 Nigeria Malaria Indicator Survey (NMIS), which was implemented by the National Malaria Elimination Programme (NMEP) of the Federal Ministry of Health (FMoH) in collaboration with the National Population Commission (NPC) and the National Bureau of Statistics (NBS). A two-stage sampling strategy was adopted for the 2021 NMIS.

The 2021 NIMS sample was stratified and selected in two stages. Stratification was achieved by separating the 36 states and Federal Capital Territory into urban and rural areas. There were 73 sampling strata since there are no rural areas in Lagos. Sample were selected independently in every stratum through a two-stage selection. Implicit stratification was achieved at the lower administrative levels by sorting the sampling frame before sample selection according to administrative order and using probability proportional to size selection in the first sampling stage.

In the first stage, 568 enumeration areas (EAs) were selected with probability proportional to the EA size. The EA size is the number of households residing in the EA. The sample selection was done in such a way that it was representative of each state. The result was 568 clusters throughout the country, 195 in urban areas, and 373 in rural areas. In the second stage's selection, a fixed number of 25 households were selected in every cluster through equal probability systematic sampling.

Three questionnaires were used in the 2021 NMIS: the Household Questionnaire, the Woman's Questionnaire, and the Biomarker Questionnaire. Based on The DHS Program's model questionnaires, the questionnaires were adapted to reflect the population and health issues relevant to Nigeria. After the questionnaires were finalized in English and translated into Hausa, Yoruba, and Igbo. The National Health Research Ethics Committee of Nigeria (NHREC) and the ICF Institutional Review Board reviewed and approved the survey protocol.

The demographic and health surveys (DHS) data from the 2021 NIMS has been used in this data, and the children sample was selected from the household recode. The children's total population was 12180; according to the NIMS 2021 report, the selected children for anaemia and RDT testing were 6–59 months old. Therefore, after applying the filter only to children between 6–59 months old, the sample used in this study became 10245. There were some missing observations for some of the predictors, and a complete case analysis was performed. The data was then used to model the data using GLMM, and

imputation was adopted to deal with the missing observations. The section below outlines the imputation methods that can be applied to missing data.

Missing data

The presence of missing data affects the analysis of the data, and different methods can be employed in handling missing data. The method employed can affect the outcome of the analysis. This in turn, could compromise the conclusion drawn from the results. It is usual to encounter missing data in epidemiological studies [9–11]. In most cases, missing data is managed by dropping the cases that are not fully measured. However, in this study, a comparison of the complete cases data and the imputed data was compared based on the standard error that resulted from the GLMM model. The comparison has been fully documented under results section.

The assumption made was that the data was missing at random (MAR). When multiple imputations are applied to a MAR dataset, unbiased results with accurate estimates for the standard error (SE) are obtained [12]. In this study, multiple imputation by chained equations was used**.** Chained equations, also known as fully conditional specification (FCS), is the algorithm that SAS Enterprise adopted. This approach is more flexible to imputation since it is designed to handle different types of variables (binary, continuous, categorical, and ordinal) and does not assume the multivariate normality of the data [11, 13]. In practice, the FCS method involves running a series of regression models such that each variable with missing data is regressed on the other variables in the data set according to its distribution. For instance, continuous variables are modelled using the linear regression and categorical variables using the logistic regression.

Imputation by FCS, as applied in SAS, is also an iterative process that starts by imputing every missing value with random draws from the distribution of the non-missing values. In this study, the FCS has been employed for imputation using two iterations, with a few imputations (nimpute = 4).

The FCS is carried out as follows:

1. Set up "*placeholders*" of one variable that suffers from missing values back to missing.
2. Set up a regression equation, according to the distribution of the variable, with the observed values as the dependent variables and the other variables ad the independent variables.
3. Replace the missing values from this variable with predictions from the regression equation.
4. Repeat the steps 1–3 for each variable that has missing values.

This is in turn forms the iterations process. At each iteration the imputed values are updated.

Multiple Imputations by Chained Equations (MICE) is a powerful and flexible method for handling missing data, more especially when the missingness is related to other variables in the dataset. It further improves the quality of statistical inferences by incorporating uncertainty and preserving relationships between variables, which can be particularly useful in complex datasets.

Dependent variable

The prevalence of malaria in children between 6–59 months old was detected using the results from rapid diagnostic tests (RDT). The RDT was performed by collecting a small blood sample and applying it to a test card. The RDT detects plasmodial antigens in the blood using monoclonal antibodies on a test strip. Therefore, the response variable used was binary, where the child tested either positive (had malaria) or negative (did not have malaria).

Independent variables

The independent variables considered in this study were rigorously selected based on different previous studies that were conducted by other authors [6, 14–16]. The independent variables included several socio-economic, demographic, and geographical factors. The socio-economic variables were: whether the household had electricity, toilet facility, type of place of residence; wealth index; mother's education level; the main material of the roof, floor and wall; if mosquito nets were used for children in the household [6, 14–16], these were collected at a household level. The demographic factors included age of the child, the gender, anaemia level, and whether a child had fever in the last two weeks. These were collected at an individual level. The geographic variables were region, and altitude.

Statistical analysis

The present study used bivariate procedures to show the association between childhood malaria and the selected independent variables. The analysis for the bivariate method used cross-tabulation technique with an application of SAS Enterprise Guide version 8.3. The Chi-squared test and p-value were used to determine whether the independent variables are significantly associated with childhood malaria or not. The variables from bivariate results with a p-value less than 5% level of significance were included in multivariate GLMM analysis.

Table 1 shows a comparison standard error from the complete case, imputed data with FCS using 2 iterations. The estimate (EST) shows some consistency for the predictors for the complete case and the imputed data.

Table 1 Estimated coefficients (EST) and standard errors (SE) for the predictors selected in the different analysis

Predictor	Reference category	Category	CC (N = 10081)		FSC1(N = 40980)		FSC2(N = 40980)	
			EST	SE	EST	SE	EST	SE
Altitude	> 1000 m	0–500 m	0,422	0.210	0.427	0.105	0.428	0.105
		501–1000 m	0.000	0.000	0.438	0.110	0.439	0.110
Anemia level	Not Anemic	Anemic	1.160	0.055	1.153	0.027	1.152	0.027
Age in Months	6–12	12–24	0.594	0.096	0.577	0.048	0.577	0.048
		25–36	0.967	0.097	0.956	0.048	0.955	0.048
		37–48	1.183	0.096	1.159	0.048	1.158	0.048
		49–59	1.455	0.097	1.433	0.048	1.432	0.048
Toilet Facility	Toilet with flush	Latrine	0.271	0.083	0.102	0.038	0.103	0.038
		No facility	0.000	0.000	0.246	0.041	0.253	0.041
Main wall material	Wood/Mud	Bricks	− 0.402	0.172	− 0.439	0.085	− 0.429	0.085
		Cement/B	− 0.321	0.072	− 0.328	0.036	− 0.331	0.036
		No Walls	− 0.581	0.136	− 0.554	0.067	− 0.571	0.067
		Other	− 0.640	0.203	− 0.653	0.100	− 0.655	0.099
Main roof material	Wood	Asbestos	0.000	0.000	0.239	0.117	0.274	0.116
		Metal/Zinc	0.000	0.000	0.150	0.049	0.151	0.049
		Other	0.000	0.000	− 0.288	0.115	− 0.272	0.114
		Thatch/Palm leaf	− 0.319	0.113	− 0.330	0.056	− 0.329	0.056
Slept under mosquito	I do not know	All	0.000	0.000	− 0.103	0.036	− 0.097	0.036
		No	0.000	0.000	− 0.081	0.036	− 0.077	0.036
Had Fever	No	Ye	0.474	0.048	0.477	0.024	0.477	0.024
Place of residence	Urban	Rural	0.251	0.060	0.263	0.030	0.263	0.030
Wealth Index	Richest	Middle	0.796	0.127	0.741	0.062	0.738	0.062
		Poorer	0.998	0.151	0.950	0.075	0.943	0.074
		Poorest	1.158	0.168	1.090	0.083	1.082	0.083
		Richer	0.736	0.104	0.693	0.051	0.693	0.051
Sex of child	Male	Female	0.000	0.000	− 0.060	0.023	− 0.061	0.023
Education Level	Secondary	No Education	0.486	0.071	0.486	0.035	0.487	0.035
		Primary	0.387	0.075	0.385	0.037	0.386	0.037
Can malaria be prevented?	No	Yes	0.000	0.000	− 0.144	0.047	− 0.145	0.047

However, there are predictors that are not significant when the complete case is used for modelling, most notably, the main roof material, whether the kids slept under a mosquito bed or not, the sex of the child, and whether the mother knows of any preventative ways for malaria. The number of imputations used was 4, which is why there are 40980 observations for the imputed data used. In comparing the standard errors from the table above for the cases, the standard errors found for the imputed data are lower than those of the complete case. This implies that the estimates obtained from the imputed data are more precise. In other words, there is less uncertainty around the estimates obtained using the imputed data.

Model formulation

Generalized linear mixed models (GLMM) are the extension of the generalized linear models (GLM) by the introduction of random effects into the linear predictor of the GLM [17]. Let N, m_i denote the number of clusters and the number of observations per cluster i, respectively; the model with q predictors can be expressed as follows:

$$X_i = g^{(-1)}(Y_i\beta + S_i b_i) + \varepsilon_i i = 1, ..., N$$

where, X_i is the $m_i \times 1$ response vector for the i^{th} cluster; $g^{-1}(.)$ is the inverse of a differentiable monotonic link function; Y_i is the $m_i \times$ q matrix of fixed covariates; β is a $q \times 1$ vector of a fixed-effects regression parameters; Z_i is a $m_i \times v$ design matrix of random effects, where v is a design parameter; b_i is a $v \times 1$ vector of a cluster-specific random effects$\varepsilon_i$is a $m_i \times 1$ error vector.

The parameters in GLMM can be estimated either by the standard maximum likelihood (ML) estimation, which in turn estimates the standard deviations of the

random effects with the assumption that the fixed effect estimates are precisely correct, or by the restricted maximum likelihood (REML) estimation, a variant that averages over some of the uncertainty in the fixed-effect parameters.

Suppose the normality assumption of $f(\boldsymbol{X}|\boldsymbol{\theta})$ is now relaxed. Assume that $\boldsymbol{X}$ and $\boldsymbol{\theta}$ are independent and $f(\boldsymbol{X}|\boldsymbol{\theta})$ is a member of the exponential family of distribution (19,20).

$$f(\boldsymbol{X}|\boldsymbol{\theta}) = exp\left\{\frac{x_i\theta_i - b(\theta_i)}{\varnothing} - c(x_i, \varnothing)\right\}$$

where $\varnothing$ is the scale parameter. Based on the model, the conditional x related to θ_i is given as follows:

$$E(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{\partial b(\theta_i)}{\partial \theta_i}$$

The model with both effects is given as follows:

$$g(\theta_i) = \boldsymbol{Y}_i\prime\beta + \boldsymbol{S}_i^{'}\boldsymbol{U}_i\prime$$

where, $\eta_i = g(\theta_i)$, g is the link function, and $\boldsymbol{U}_i$ a vector of random effects. In this study, malaria rapid diagnostic testing (RDT) results is either 0 (a negative RDT result) or 1 (a positive RDT result). Thus, the logistic regression where the $g(.)$ as the logit link, with $\boldsymbol{X}_i$ and $\boldsymbol{S}_i$ $(i = 1, \ldots, m)$ being p-dimension and q-dimension vectors of known covariate values, while $\boldsymbol{\beta}$ is a q-dimension vector of unknown fixed effects regression coefficient [18, 19].

Results

Countries comparison for malaria prevalence

Comparing some of the African countries with Nigeria to have a view of how Nigeria stands in terms of malaria prevalence. As shown in Table 2, the study is based on the MIS datasets obtained from countries that have been previously identified as having a high prevalence of malaria, namely – Mozambique, Tanzania, Angola, Burkina Faso, Uganda, Niger and Nigeria.

Exploratory data analysis

Table 3 shows the prevalence of malaria in children from the central region is 4.6%, compared to 30.5 and 5.0% for the north and south regions, respectively. The prevalence of malaria in children from an area where the altitude is between 0–500 m is 30.8%, compared to 9.2 and 0.1% for areas where the altitude is between 501–1000 m and more than 1000 m, respectively. The prevalence of malaria in children who are anaemic is 33.5%, compared to 6.6% for those who are not anaemic. The prevalence of malaria in children who are between 49–59 months old is 10.7%, compared to children who are between 37–48, 25–36, 12–24, and 6–12 months, with a prevalence of 9.37, 8.55, 7.24, and 2.14% respectively. Children from a household with a latrine toilet showed a high prevalence of malaria at 25.5%, compared to children from a household with no toilet facility and toilets with flush, at 9.4 and 5.3%, respectively.

Table 2 Selected countries by year of survey

Country	Year of Survey	Prevalence (%)
Tanzania	2021	7
Nigeria	2021	37
Niger	2021	28
Uganda	2018/2019	20
Burkina Faso	2017/2018	20
Mozambique	2018	40
Angola	2011	13

Children from a household with main floor material as earth/sand show a high prevalence of malaria at 26.8%, compared to children from a household with main floor material such as cement, ceramic tiles, other form of material, and wood; with a prevalence of 11.4, 1.4, 0.3, and 0.2%, respectively. Children who did not sleep under a mosquito bed net showed a high prevalence of malaria at 16.6%, followed by children who slept under a net at 15.5%, and some who slept under a mosquito net at 8.1%. Children who did not have a fever in the last two weeks showed a high prevalence of malaria at 21.1%, compared to children who had a fever at 19.0%, and those whose guardians did not know whether they had a fever or not were at 0.1%. Children from rural areas showed a high prevalence of malaria at 34.3% compared to those from urban areas with a 5.9% prevalence.

Children from households with the wealth index in the poorest class showed a high prevalence of malaria at 14.3%, followed by those from the poorer, middle, richer, and richest at 12.4, 6.8, 4.6, and 1.9%, respectively. Male children showed a high prevalence of malaria with 21.8%, compared to female children with 18.4%. Children from households without electricity showed a high prevalence of malaria at 30.9%, compared to those from households with electricity at 9.3%. Children whose mother has no education showed a high prevalence of malaria at 25.0%, compared to those children whose mother has primary, secondary, and higher education; at 7.2, 6.6, and 1.3%, respectively. Children from a household where the main wall material is wood/mud have a higher prevalence of malaria at 24.4%, compared to those from a household with

Table 3 Distribution of the frequency of the predictor variables

Variable	Category	Malaria (% negative)	Malaria(% positive)	p-value
Region	Central	1221 (10.6%)	528 (4.6%)	< 0.0001
	North	3705 (32.2%)	3509 (30.5%)	
	South	1959 (17.0%)	578 (5.0%)	
Altitude	0–500 m	5535 (48.1%)	3542 (30.8%)	< 0.0001
	501–1000 m	1309 (11.4%)	1063 (9.2%)	
	> 1000 m	41 (0.3%)	10 (0.1%)	
Anemia	Anemic	4047 (35.2%)	3855 (33.5%)	< 0.0001
	Not Anemic	2838 (24.7%)	760 (6.6%)	
	6–12	825 (7.95%)	222 (2.14%)	
Child's age in months	12–24	1589 (15.3%)	751 (7.24%)	< 0.0001
	25–36	1355 (13.06%)	888 (8.55%)	
	37–48	1384 (13.34%)	972 (9.37%)	
	49–59	1281 (12.35%)	1108 (10.68%)	
Toilet facility	Latrine	3126 (27.4%)	2916 (25.5%)	< 0.0001
	No Facility	1146 (10.0%)	1074 (9.4%)	
	Toilet with flush	2560 (22.4%)	608 (5.3%)	
Main Floor Material	Cement	2939 (25.8%)	1302 (11.4%)	< 0.0001
	Ceramic Tiles	958 (8.4%)	159 (1.4%)	
	Earth/Sand	2752 (24.1%)	3059 (26.8%)	
	Other	129 (1.1%)	38 (0.3%)	
	Wood	44 (0.4%)	26 (0.2%)	
Sleep under mosquito net	All children	2426 (21.1%)	1778 (15.5%)	< 0.0001
	No net	3479 (30.3%)	1905 (16.6%)	
	Some	965 (8.4%)	929 (8.1%)	
Child had fever in the last 2 weeks	Don't know	18 (0.2%)	8.10 (0.1%)	< 0.0001
	No	4546 (39.5%)	2426 (21.1%)	
	Yes	2321 (20.2%)	2181 (19.0%)	
Type of place of residence	Rural	4611 (40.1%)	3939 (34.3)	< 0.0001
	Urban	2273 (19.8%)	675 (5.9%)	
Wealth Index	Middle	1343 (11.7%)	780 (6.8%)	< 0.0001
	Poorer	1237 (10.8%)	1432 (12.4%)	
	Poorest	1281 (11.1%)	1647 (14.3%)	
	Richer	1332 (11.6%)	533 (4.6%)	
	Richest	1693 (14.7%)	223 (1.9%)	
Sex	Female	3401 (29.6%)	2111 (18.4%)	0.0001
	Male	3482 (30.3%)	2503 (21.8%)	
Electricity	No	3472 (30.4%)	3534 (30.9%)	< 0.0001
	Yes	3359 (29.4%)	1064 (9.3%)	
Mother's Education level	Higher	886 (7.7%)	154 (1.3%)	< 0.0001
	No Education	2582 (22.5%)	2879 (25.0%)	
	Primary	1070 (9.3%)	826 (7.2%)	
	Secondary	2347 (20.4%)	756 (6.6%)	
Main wall material	Bricks	102 (0.9%)	53 (0.5%)	< 0.0001
	Cement/Blocks	4092 (35.8%)	1631 (14.3%)	
	No walls	135 (1.2%)	112 (1.0%)	
	Other	39.4 (0.3%)	13 (0.1%)	
	Wood/Mud	2454 (21.5%)	2788 (24.4%)	

Table 3 (continued)

Variable	Category	Malaria (% negative)	Malaria(% positive)	p-value
Main roof material	Asbestos	178 (1.6%)	36 (0.3%)	< 0.0001
	Cement	101 (0.9%)	10 (0.1%)	
	Metal/Zinc	5307 (46.6%)	3489 (30.6%)	
	Other	71 (0.6%)	37 (0.3%)	
	Thatch/Palm leaf	430 (3.8%)	400 (3.5%)	
	Wood	718 (6.3%)	624 (5.5%)	

cement/blocks, no walls, bricks, and other materials, with 14.3, 1.0, 0.5, and 0.1%, respectively. Children from a household where the main roof material is metal/zinc showed a high prevalence of malaria at 30.6%, followed by wood, thatch/palm leaf, asbestos, other, and cement; at 5.5, 3.5, 0.3, 0.3, and 0.1%, respectively.

Model results

Table 4 shows that a child who is from a household that is at an altitude of between 0–500 m is about 1.53 times more likely to test positive from the RDT for malaria (OR = 1.534, p-value < 0.0001) compared to a child from a household that is at an altitude that is more

Table 4 Statistical analysis model on malaria determinants

Predictor	Reference category	Category	EST	SE	DF	t Value	p-value	Odd ratio (OR)
Altitude	> 1000 m	0–500 m	0,428	0.105	40939	4.08	< 0.0001	1.534
		501–1000 m	0.439	0.110	40939	3.98	< 0.0001	1.551
Anemia level	Not Anemic	Anemic	1.152	0.027	40939	42.07	< 0.0001	3.166
Age in Months	6–12	12–24	0.577	0.048	40939	12.13	< 0.0001	1.780
		25–36	0.955	0.048	40939	19.98	< 0.0001	2.598
		37–48	1.158	0.048	40939	24.34	< 0.0001	3.184
		49–59	1.432	0.048	40939	30.04	< 0.0001	4.188
Toilet Facility	Toilet with flush	Latrine	0.103	0.038	40939	2.74	0.0061	1.109
		No facility	0.253	0.041	40939	6.19	< 0.0001	1.287
Main wall material	Wood/Mud	Bricks	− 0.429	0.085	40939	− 5.03	< 0.0001	0.651
		Cement/B	− 0.331	0.036	40939	− 9.22	< 0.0001	0.718
		No Walls	− 0.571	0.067	40939	− 8.47	< 0.0001	0.565
		Other	− 0.655	0.099	40939	− 6.59	< 0.0001	0.519
Main roof material	Wood	Asbestos	0.274	0.116	40939	2.35	0.0186	1.315
		Metal/Zinc	0.151	0.049	40939	3.09	0.002	1.162
		Other	− 0.272	0.114	40939	− 2.38	0.0174	0.762
		Thatch/Palm leaf	− 0.329	0.056	40939	− 5.88	< 0.0001	0.720
Slept under mosquito	I do not know	All	− 0.097	0.036	40939	− 2.71	0.0067	0.907
		No	− 0.077	0.036	40939	− 2.17	0.0301	0.926
Had Fever	No	Ye	0.477	0.024	40939	20.05	< 0.0001	1.612
Place of residence	Urban	Rural	0.263	0.030	40939	8.85	< 0.0001	1.301
Wealth Index	Richest	Middle	0.738	0.062	40939	11.83	< 0.0001	2.091
		Poorer	0.943	0.074	40939	12.66	< 0.0001	2.569
		Poorest	1.082	0.083	40939	13.11	< 0.0001	2.950
		Richer	0.693	0.051	40939	13.59	< 0.0001	1.999
Sex of child	Male	Female	− 0.061	0.023	40939	− 2.64	0.0082	0.941
Education Level	Secondary	No education	0.487	0.035	40939	13.9	< 0.0001	1.627
		Primary	0.386	0.037	40939	10.42	< 0.0001	1.470
Can malaria be prevented?	No	Yes	− 0.145	0.047	40939	− 3.09	0.002	0.865

than 1000 m. A child who is from a household that is at an altitude of between 501–1000 m is about 1.55 times more likely to test positive from the RDT for malaria (OR = 1.55, p-value < 0.0001) compared to a child from a household that is at an altitude that is more than 1000 m. A child who is anaemic is about 3.12 times more likely to test positive from the RDT for malaria (OR = 3.166, p-value < 0.0001) compared to a child who is not anaemic. Children who are between 12–24 months old are about 1.78 times more likely to test positive from the RDT for malaria (OR = 1.780, p-value < 0.0001) compared to children who are between 6–12 months old. Children who are between 25–36 months old are about 2.60 times more likely to test positive from the RDT for malaria (OR = 2.598, p-value < 0.0001) compared to children who are between 6–12 months old. Children who are between 37–48 months old are about 3.18 times more likely to test positive from the RDT for malaria (OR = 3.184, p-value < 0.0001) compared to children who are between 6–12 months old. Children who are between 49–59 months old are about 4.19 times more likely to test positive from the RDT for malaria (OR = 4.188, p-value < 0.0001) compared to children who are between 6–12 months old.

Children who are from a household that has a latrine toilet facility were about 1.11 times more likely to test positive from the RDT for malaria (OR = 1.109, p-value = 0.0061) compared to children from a household where the facility has a toilet with a flush. Children from a household with no toilet facility were about 1.29 times more likely to test positive from the RDT for malaria (OR = 1.287, p-value < 0.0001) compared to children from a household where the facility was a toilet with a flush. Children from a household where the main wall material was bricks were about 0.65 times less likely to test positive from the RDT (OR = 0.651, p-value < 0.0001) compared to children from a household where the main wall material was wood/mud. Children from a household with the main wall material cement/Bricks were about 0.72 times less likely to test positive from the RDT (OR = 0.718, p-value < 0.0001) than children from a household with wood/mud. Children from a household with no walls were about 0.0.57 times less likely to test positive from RDT (OR = 0.565, p-value < 0.0001) than children with the main wall material, wood/mud. Children from a household where the main wall material was other were about 0.52 times less likely to test positive from RDT (OR = 0.519, p-value < 0.0001) compared to children from a household where the main wall material was wood/mud.

Children from a household with the main roof material asbestos were about 1.32 times more likely to test positive from RDT (OR = 1.315, p-value = 0.0186) than those with wood. Children from a household with the main roof material was metal/zinc were about 1.16 times more likely to test positive from RDT (OR = 1.162, p-value = 0.002) than children from a household with wood roof material. Children from a household where the main roof material was other were about 0.76 times less likely to test positive from RDT (OR = 0.762, p-value = 0.0176) compared to children from a household where the main roof material was wood. Children from a household with the main roof material were thatch/palm leaf and were about 0.72 times less likely to test positive from RDT (OR = 0.720, p-value < 0.0001) than children with wood roof material.

Children who are from a household where all of them slept under a mosquito net were about 0.91 times less likely to test positive for malaria using RDT (OR = 0.907, p-value = 0.0067) compared to children whose mothers are not aware whether the child slept under a net or not. Children who are from a household where none of them slept under a mosquito net were about 0.93 times less likely to test positive for malaria using RDT (OR = 0.926, p-value = 0.0301), compared to children whose mother is not aware whether the child slept under a net or not. Children who had a fever in the last two weeks were about 1.61 times more likely to test positive for malaria using RDT (OR = 1.612, p-value < 0.0001) than those who did not have a fever in the last two weeks. Children from households in rural areas are about 1.61 times more likely to test positive from the RDT (OR = 1.301, p-value < 0.0001) compared to children from households in urban areas.

Children from a household in the middle wealth index class were about 2.09 times more likely to test positive from RDT (OR = 2.091, p-value < 0.0001) than children from the richest class. Children from a household that falls in the poorer wealth index class were about 2.57 times more likely to test positive from RDT (OR = 2.569, p-value < 0.0001) than children from the richest class. Children from a household that falls in the poorest wealth index class were about 2.95 times more likely to test positive from RDT (OR = 2.950, p-value < 0.0001) than children from the richest class. Children from a household that falls in the richer wealth index class were about 2.00 times more likely to test positive from RDT (OR = 1.999, p-value < 0.0001) than children from the richest class.

Female children were about 0.94 times less likely to test positive from RDT (OR = 0.941, p-value = 0.0082) than male children. Children whose mothers did not have any education were about 1.62 times more likely to test positive from RDT (OR = 1.627, p-value < 0.0001) compared to children whose mothers had a secondary level of education. Children whose mothers had primary-level education were about 1.47 times more likely to test positive

from RDT (OR = 1.470, p-value < 0.0001) compared to children whose mothers had a secondary level of education. Children whose mother knows about some preventative matters for malaria were about 0.87 times less likely to test positive for malaria using RDT (OR = 0.865, p-value = 0.002) compared to children whose mothers do not know if any preventative matters can be taken to curb the spread of malaria.

Discussion

This study employed imputation to deal with some missing observations of some variables. Multiple imputation by chained equations was used, and the FCS with iterations of up to 2 were tested; when compared to the complete case, some variables were not significant in the CC. However, those variables were significant when using the imputed data. In comparing the standard errors for the cases, the standard errors found for the imputed data are lower than those of the complete case. This implies that the estimates obtained from the imputed data are more precise. In other words, there is less uncertainty around the estimates obtained using the imputed data [11, 13].

The results of the analysis showed that the following factors are associated with a positive malaria RDT result in children under the age of 5 years in Nigeria: altitude, anaemia level, age of the child in months, child had a fever in the last 2 weeks, toilet facility, main wall material, main roof material, slept under mosquito net, place of residence, household wealth index, sex of child, education level of the mother, and knowledge of preventative measures that can be taken to prevent malaria. The interactions between some of the selected factors were investigated. However, they were not significant in the current study. The key factors obtained from the study are divided into the following sections, which are socioeconomic, environmental, and health-related.

Health-related factors

The results showed that as children grow old, the chances of testing positive for malaria using the RDT increase. A possible explanation for the increase in chances of children testing positive as they grow old is that they are under minimal supervision from their parents, which can lead to them playing in areas that are mostly exposed to mosquitoes. The result is consistent with what Gaston and Ramroop [6] found. Based on the outcome of the results obtained from the model, children who had a fever in the past two weeks are more likely to test positive for malaria using RDT. The sex of the child also proved to be significant in the model. In the current study, male children are at a higher risk of testing positive for malaria compared to female children. The results also suggested that children who are anaemic are at a high risk of testing positive for malaria using RDT, as compared to those without anaemia. The result is consistent with that found by Sultana et al.[20], Gaston and Ramroop [6], and Oladeinde et al. [21].

Environmental factors

The study revealed that children from households at an altitude that is less than 1000 m are at a high risk of having a positive RDT result. This suggests that for future developments in Nigeria, citizens of Nigeria should consider building their homesteads in areas that are more than 1000 m above sea level [6, 22]. This can be achieved by looking into areas that are at an altitude of more than 1000 m for new developments and advising people not to build at places that are not regulated. The study revealed that children who slept under a mosquito net the night before the survey had a lower risk of having malaria. This might be because children who sleep under mosquito bed nets are being protected from mosquito bites.

Socioeconomic factors

Gaston and Ramroop [6] found that kids from households with better toilet facilities are less likely to test positive. The current study supports the findings that are prevalent in the above-mentioned study. In this study, it has been identified that children from households where the main toilet facility is a toilet with a flush stand have a better chance of testing negative for RDT than children from households with a latrine or no facility at all. This may be because mosquitoes are fond of damp areas, and therefore, with the latrine facilities, there will always be visibility of mosquitoes. When there is no facility, individuals are forced to use the bushes to relieve themselves, and mosquitoes are found in grasslands and forest areas. It could be worthwhile to invest in better facilities for households in Nigeria to try to curb the high prevalence of childhood malaria.

The main roof material also has an impact on the children's RDT result, and based on the outcome of the model used in this study, it shows that children from a household where the main roof material is metal/zinc have a greater possibility of testing positive for malaria than children from a household where the main roof material is wood. The study revealed that households from rural areas have a higher prevalence of malaria, and children from these areas are most likely to test positive for malaria using RDT. This may be explained by the fact that in rural areas, housing is poorer than in urban areas. Moreover, there are limited facilities such as health care, proper toilet facilities, availability of clean drinking water, which in turn might be fetched from rivers in the rural areas, and proper housing [6, 20, 23, 24]. Furthermore, when fetching water from the rivers for drinking

and other household uses, that might force them to pass through the forest and thus make them more susceptible to mosquito bites and therefore more chances of contracting malaria, which can lead to high transmission in the household [6, 20, 23, 24].

The results further revealed that the chances of children testing positive for malaria using RDT when they are from the poorer and poorest households are very high as compared to those who are from the richest households. The chances also decrease as the wealth index increases to the middle and richer. These results are consistent with previous studies [6, 25, 26]. The results from this study also showed that the risk of a child testing positive for malaria using RDT is lower when the parent has some form of formal education compared to a child whose mother has no formal education. These results were consistent with previous studies by Sultana et al. [20], Zgambo et al. [27], and Gaston and Ramroop [6].

Lastly, the results suggest that mosquito nets reduce the likelihood of a child testing positive for malaria using RDT; there could still be challenges in acquiring and effectively using them by the parents/caregivers. This, in turn, can be curbed by making it mandatory for each household to have a mosquito net and for the caregiver to undergo mandatory training to raise awareness about mosquito nets' use. Training should be conducted to make sure that caregivers from the most disadvantaged families know what some of the preventative measures are that are available for them to adopt in fighting the high prevalence of malaria.

Conclusion

The aim of this study was to determine the risk factors associated with positive malaria RDT results. The statistical model used to estimate the risk factors associated with a positive RDT result in children under 5 belongs to the family of generalized linear models (GLM). The generalized linear mixed models (GLMMs) effectively consider the fixed effects. Furthermore, the assumption of normality is relaxed [28], which is an advantage of using the method in a survey dataset. In this study, the missing observations were included; the multiple imputation using chained equations (MICE) was used to address the missingness of the observations [12, 29]. Thus, finally, modelling the complete dataset is advantageous since it leads to more accurate results.

The current findings revealed that there is an association in socio-economic, demographic, and geographical factors with malaria in children under 5 years in Nigeria. The following factors have been found to be significant for children to test positive for malaria using RDT: altitude, anaemia level, age in months, fever status in the past 2 weeks, toilet facility, main roof material, household wealth index, type of place of residence, sex of the child, and mother's education level. Policymakers can use these factors to identify hotspots of malaria and thus, in turn, be able to have preventative measures in place to curb the increase in malaria cases in children. The government should focus on disadvantaged households from rural areas with low altitude levels, especially those with poor toilet facilities. Additionally, children with anaemia should prioritize health care and support. The findings also reveal that mothers' education should be encouraged and supported.

The current study will help the government and policymakers to control and possibly eradicate malaria in children under 5 years of age in Nigeria. The primary focus should be on children who are anaemic, households located at an altitude less than 1000 m, children who had a fever in the last two weeks, proper toilet facilities, those from rural areas, and, lastly, male children. Policymakers can develop programs that educate community members, especially caregivers from rural areas, about proper sanitation. This can be achieved by radio, television, and billboard awareness. Furthermore, some form of funding can be set aside to improve infrastructure for those affected by the most risk factors and provide proper hospitals, housing, and toilet facilities. Educating caregivers about malaria and how it can be contracted, prevented, and treated, as well as when to seek medical care before it causes death, will also be effective. Furthermore, the model used in this study will help other researchers to compare results.

According to the NIMS 2021 report, it has been outlined that about 24% of mosquito nets were not used the night before the survey. The main reasons given for not using a mosquito net the night before the survey were that the net was not needed (24%), there were no mosquitoes (18%), it was too hot (16%), and others (12%). The proportion of respondents reporting that they did not use a net the night before the survey because it was not needed was higher in rural (26%) than urban (20%) areas. Considering the statistics outlined above, it is suggested that there is a lack of knowledge on why and when mosquito nets should be used. The data collection took place from October to December, which is deemed as summer in the African continent, and therefore there is a high prevalence of mosquitoes.

The current study's findings show that mosquito nets are vital to fight against the high prevalence of malaria in children under 5 years. Therefore, it is suggested that local governments and health organizations should engage community leaders from rural and urban areas to work together to develop strategies to combat malaria through education on why and when mosquito nets should be used. This can be achieved by rolling out

mosquito nets and giving proper training on how to use them for effective protection.

Future studies

Longitudinal studies can be used to address the problem of causality in future studies. Even though each category has significant risk factors: demographic, economic, and geographical. It might be worthwhile for future studies to investigate the interactions between some of the predictor variables to understand how they contribute to malaria prevalence. Other studies can also look into plans where there are going to be new developments and advice accordingly based on some of the risk factors that have been outlined in the current study to make sure that, for instance, the altitude level is high enough to avoid the high prevalence of malaria in children. Furthermore, the contributing factors to anaemia can be added to determine which ones add more significance to positive RDT results.

Limitations

This study used secondary cross-sectional data from NMIS, and this data may not be able to address the causality but rather the association. Therefore, a longitudinal study is suggested to address this problem.

Author contributions

Conceptualization, Talani Mhelembe, Shaun Ramroop and Faustin Habyarimana; Data curation, Talani Mhelembe; Formal analysis, Talani Mhelembe; Investigation, Talani Mhelembe; Methodology, Talani Mhelembe, Shaun Ramroop and Faustin Habyarimana; Project administration, Talani Mhelembe; Resources, Talani Mhelembe; Software, Talani Mhelembe; Supervision, Shaun Ramroop and Faustin Habyarimana; Validation, Talani Mhelembe; Visualization, Talani Mhelembe; Writing—original draft, Talani Mhelembe; Writing—review & editing, Shaun Ramroop and Faustin Habyarimana.

Funding

This research received no external funding.

Data availability

The data used in this study can be obtained from the Demographic Health and Surveys website.URL. https://dhsprogram.com/data/. The data used in this study can be obtained from the Demographic and Health Surveys (DHS): The DHS Program-Data.

Declarations

Ethics approval and consent to participate

The study used secondary data and therefore no separate ethical approval was required. However, access to the data was granted through an online request to the measure DHS program (http://www.dhsprogram.com). The data used in this study are publicly accessible and did not contain any personal identifiers.

Competing interests

The authors declare no competing interests.

Received: 13 September 2024 Accepted: 10 February 2025
Published online: 21 February 2025

References

1. Balarajan Y, Ramakrishnan U, Özaltin E, Shankar AH, Subramanian SV. Anaemia in low-income and middle-income countries. Lancet. 2011;378:2123–35.
2. Sumbele IUN, Samje M, Nkuo-Akenji T. A longitudinal study on anaemia in children with *Plasmodium falciparum* infection in the Mount Cameroon region: prevalence, risk factors and perceptions by caregivers. BMC Infect Dis. 2013;13:123.
3. WHO. Expert committee on malaria: twentieth report. Geneva: World Health Organization; 2000.
4. Wanzira H, Katamba H, Okullo AE, Agaba B, Kasule M, Rubahika D. Factors associated with malaria parasitaemia among children under 5 years in Uganda: a secondary data analysis of the 2014 Malaria Indicator Survey dataset. Malar J. 2017;16:191.
5. WHO. World malaria report 2022. Geneva: World Health Organization; 2022.
6. Gaston RT, Ramroop S. Prevalence of and factors associated with malaria in children under 5 years of age in Malawi, using malaria indicator survey data. Heliyon. 2020;6: e03946.
7. WHO. World malaria report 2015. Geneva: World Health Organization; 2015. P. 243. https://iris.who.int/handle/10665/200018. Accessed 21 Sep 2024.
8. Egbon O. Bayesian analysis of the joint risk of malaria and anemia among under-age five children in Nigeria. Research square. 2021. https://www.researchsquare.com/article/rs-148795/v1.
9. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. Am J Epidemiol. 2008;168:355–7.
10. Finkle WD, Greenland S, Miettinen OS, Ziel HK. Endometrial cancer risk after discontinuing use of unopposed conjugated estrogens (California, United States). Cancer Causes Control. 1995;6:99–102.
11. Hendry GM, Naidoo RN, Zewotir T, North D, Mentz G. Model development including interactions with multiple imputed data. BMC Med Res Methodol. 2014;14:136.
12. Rubin DB. The calculation of posterior distributions by data augmentation: comment: a noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. J Am Stat Assoc. 1987;82:543.
13. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. Am J Epidemiol. 2010;171:624–32.
14. Oguoma VM, Anyasodor AE, Adeleye AO, Eneanya OA, Mbanefo EC. Multilevel modelling of malaria risk in children under 5 years old in Nigeria. Research Square. 2020. https://www.researchsquare.com/article/rs-32524/v1.
15. Roberts D, Matthews G. Risk factors of malaria in children under the age of 5 years old in Uganda. Malar J. 2016;15:246.
16. Ugwu CLJ, Zewotir T. Spatial distribution and sociodemographic risk factors of malaria in Nigerian children less than 5 years old. Geospat Health. 2020. https://doi.org/10.4081/gh.2020.819.
17. Li P, Redden DT. Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. BMC Med Res Methodol. 2015;15:38.
18. McCullagh P. Generalized linear models. London: Routledge; 2019.
19. McCulloch CE, Searle SR. Generalized, linear, and mixed models. Hoboken: John Wiley & Sons; 2004.
20. Sultana M, Sheikh N, Mahumud RA, Jahir T, Islam Z, Sarker AR. Prevalence and associated determinants of malaria parasites among Kenyan children. Trop Med Health. 2017;45:25.
21. Oladeinde B, Omoregie R, Olley M, Anunibe J, Onifade A, Oladeinde O. Malaria and anemia among children in a low resource setting in Nigeria. Iran J Parasitol. 2012;7:31–7.

22. Chirombo J, Lowe R, Kazembe L. Using structured additive regression models to estimate risk factors of malaria: analysis of 2010 Malawi Malaria Indicator Survey Data. PLoS ONE. 2014;9: e101116.
23. Lwetoijera DW, Kiware SS, Mageni ZD, Dongus S, Harris C, Devine GJ, et al. A need for better housing to further reduce indoor malaria transmission in areas with high bed net coverage. Parasit Vectors. 2013;6:57.
24. Kazembe LN, Mathanga DP. Estimating risk factors of urban malaria in Blantyre, Malawi: a spatial regression analysis. Asian Pac J Trop Biomed. 2016;6:376–81.
25. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. Lancet Infect Dis. 2004;4:327–36.
26. Chitunhu S, Musenge E. Spatial and socio-economic effects on malaria morbidity in children under 5 years in Malawi in 2012. Spat Spatio Tempor Epidemiol. 2016;16:21–33.
27. Zgambo M, Mbakaya BC, Kalembo FW. Prevalence and factors associated with malaria parasitaemia in children under the age of 5 years in Malawi: a comparison study of the 2012 and 2014 Malaria Indicator Surveys (MISs). PLoS ONE. 2017;12(4): e0175537.
28. Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. J R Stat Soc Ser B Stat Methodol. 1999;61:381–400.
29. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc. 1996;91:473–89.

Publisher's Note