

Weighted Multilevel Models for Malaria Indicator Survey

Bereket Tessema Zewude (✉ bereket.tessema2010@gmail.com)

2Department of Statistics, University of South Africa c/o Christiaan de Wet Road & Pioneer Avenue,
Private Bag X6, Florida 1710, Johannesburg, South Africa

Legesse Kassa Debusho

2Department of Statistics, University of South Africa c/o Christiaan de Wet Road & Pioneer Avenue,
Private Bag X6, Florida 1710, Johannesburg, South Africa

Research article

Keywords: Generalized linear mixed model, Odd ratios, Pseudo-maximum-likelihood, Scaling design weight, Weighted multilevel model Background Malaria

Posted Date: August 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-53309/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Weighted Multilevel Models for Malaria Indicator Survey

Bereket Tessema Zewude^{1,2*} and Legesse Kassa Debusho²
 Department of Statistics,

*Correspondence:

bereket.tessema2010@gmail.com

²Department of Statistics,

University of South Africa c/o
 Christiaan de Wet Road & Pioneer
 Avenue, Private Bag X6, Florida
 1710, Johannesburg, South Africa
 Full list of author information is
 available at the end of the article

Abstract

Background: The data generated from malaria indicator surveys have a multilevel structure where children or households are nested within cluster and this may result in dependent data. In a survey, clusters or units within clusters may be selected with unequal probabilities and often surveys are subjected to non-response. Modeling of these data must take into consideration these aspects as these could lead to incorrect inferences. The main purpose of this study is to assess the factors that affect child malaria diagnostic outcomes and to estimate the proportion of overall child-level variation in malaria outcome that is attributable to child and household-level predictors and the cluster-level predictors using various weighted multilevel models. In this study, cluster represents a "Kebele", which is the smallest administrative unit in regional states in Ethiopia.

Methods: A sample of 4,384 under five years age children were in this study from three regional states of Ethiopia. Various multilevel models with random cluster effects were used taking into account the survey design weights. These weights are scaled to address unequal probabilities of selection within clusters.

Results: The asymptotic chi-square mixture distribution test results suggested the need for the cluster-level random effects in all the models. The findings related to child / household- and cluster-level predictors are consistent with those available in existing literature. The overall variability has been partitioned into child/household-level and cluster-level variability and the results revealed that some of the differences between clusters in under five years age children malaria RDT outcomes explained by child/household factors.

Conclusions: The use of weighted multilevel model to examine within- and between cluster variability in malaria RDT outcomes for under five years age children and of the extent to which between clusters variability is explained by child / household and cluster-level factors. The approach used in this paper allows investigation of how individual- and cluster-level factors including the public health authorities' interventions plans related to health outcomes simultaneously. Such collective assessment approach may lead to more effective public health strategies and could have important policy implications for health promotion and for the reduction of health disparities.

Keywords: Generalized linear mixed model; Odd ratios; Pseudo-maximum-likelihood; Scaling design weight; Weighted multilevel model

Background

Malaria is a serious hemolytic vector borne disease caused by five *Plasmodium* parasite species that affect humans. These are *Plasmodium falciparum*, *Plasmodium vi-*

vax, *Plasmodium ovale*, *Plasmodium malariae* and *Plasmodium knowlesi*. The parasites are spread to people through the bites of infected female *Anopheles* mosquitoes. The first two species pose the greatest threat and they are the most common malaria pathogens which dominates in Africa [1]. For example, about 66.5%, 69.2% and 89.4% reported malaria cases in years 2016, 2017 and 2018 in Ethiopia were caused by *Plasmodium falciparum*, respectively and the remainder of the cases were caused by *Plasmodium vivax* [2].

According to the latest World Health Organization (WHO) Report 2019, an estimated 228 million cases of malaria occurred worldwide, compared with 251 million cases in 2010 and 231 million cases in 2017 [2]. An estimated 67% (272,000) malaria deaths in 2018 were children aged under 5 years. According to Centre for Disease Control (CDC) and Prevention, generally in Africa where there is high malaria transmission of young children who have not yet developed partial immunity to malaria and pregnant women, whose immunity is decreased by pregnancy, especially during the first and second pregnancies are the most vulnerable groups [3]. The estimated malaria cases in Ethiopia declined from about 7.7 to 2.4 millions and estimated deaths due to malaria declined from 14,424 to 4,757 between 2010 and 2018. The country has low malaria prevalence compared to most other malaria-endemic countries in Africa (e.g. Burkina Faso, Cameroon, the Democratic Republic of the Congo and Ghana). However, malaria is still a major public health problem in Ethiopia [4].

In most malaria indicator surveys (MIS), the sampling designs organize the study populations into clusters such as districts, villages or survey enumeration areas. Then select households or people within the clusters to collect data (e.g. see [5, 6]). The data generated from these surveys have a multilevel structure where children or households are nested within cluster and hence the surveys result in dependent data. Since children or households that live in the same clusters share similar environment and the public health administrative processes taken by the authorities to control malaria transmission in the areas, hence they are likely to resemble each other with respect to the effects from these. Thus, two randomly selected children from the same cluster, e.g. survey enumeration area, they may have similar malaria status (tested positive or negative) than the malaria status of two randomly selected children from different clusters, even when the measured child / household-specific characteristics of the selected children are identical to one another. Analysis of the malaria status data by conventional binary logistic regression model needs the outcomes to be independent across children given the predictor variables. However, if the observed outcomes are not independent of one another, failure to appropriately handle the intra-cluster correlations among the observations within clusters can lead to incorrect inferences on the effects of model covariates ([7, 8, 9]). The conventional analysis also prevents to study the association between specific cluster characteristics with the individual outcome [10] and ignores the hierarchical structure of the data.

In most malaria prevalence related studies, researchers apply either the generalized estimating equation (GEE, [11, 12]), for example see ([13, 14, 15]), by considering the intra-cluster correlation in order to improve estimates of the fixed effects and their standard errors; or the generalized linear mixed model by introducing random effects

to take into account the within correlation only (e.g. see [16, 17]). Rodriguez and Goldman [18], however emphasized that the estimation of similarity of observations within a cluster or group is not only improve these two estimates but also the estimation, in particular by introducing individual level or cluster-level factors as a set of control variables, may provide information how the cluster level factors influence the individual characteristics or vice versa. Multilevel regression models ([19, 20, 21]) can also handle correlated data. Multilevel models incorporate cluster-specific random effects that account for the dependency of the data by partitioning the total individual variance into variation due to the clusters and the individual-level variation that remains [22]. However, unlike multilevel models, GEE had been developed to handle correlated data without explicitly accounting for heterogeneity across clusters and it does not provide direct estimates of the variance structure, but treat these as nuisance parameters [23]. It does also not allow to examine influence of individual-level or cluster-level factors on between cluster variation, and sources of intra-cluster correlation [24].

In order to have proper plans in response to health risks, public health research becomes increasingly interested in investigating the reasons of variation between areas or area-level influences on health, i.e. contextual influences on health. Therefore, to analyze health related data, e.g. MIS data, one needs to adopt methods that allow investigations of within and between clusters variances ([24, 9]). The multilevel models take into account the clustered nature of the data, so they can correctly estimate standard errors which lead to more accurate inferential decisions [20] and they allow to investigate sources of variations within and across clusters. Furthermore, in complex survey, clusters or units within clusters or both may be selected with unequal weights or probabilities and often surveys are subjected to non-response. Modeling of data obtained from these surveys must take into consideration these aspects as these could lead to biased parameter estimates ([25, 8]).

The main objective of the current study is to assess the factors that affect the malaria rapid diagnostic test outcomes of under five years age children in three major regions of Ethiopia. In addition to assessing the factors that affect the malaria outcomes, we are also interested to estimate the proportion of overall child-level variation in malaria outcome that is attributable to child / household-level characteristics (or predictors) only, to the cluster-level characteristics (or predictors), and to both predictors simultaneously using various weighted multilevel models. In this study, "Kebele", which is the smallest administrative unit of Ethiopia, similar to a neighbourhood or a localized and delimited group of people was considered as a cluster.

The rest of the paper is organized as follows. The data, some basic review of statistical models for the analysis of multilevel data with binary outcomes and scaling methods of sampling weights are introduced in "Methodology" section. The results from applying these methods on the study data are discussed in "Results" section. Finally discussion, and conclusions and pointers for future study are given in "Discussion" and "Conclusion" sections respectively.

Methodology

Study data

The data for this study was obtained from the 2011 Ethiopia National Malaria Indicator Survey (MIS). The survey was conducted by the Ethiopian Health and Nutrition Institutes and its partners, the Ethiopian Ministry of Health in collaboration with the Central Statistics Agency (CSA), US President's Malaria Initiative (PMI), United Nations Children's Fund (UNICEF), Malaria Control and Evaluation Partnership in Africa (MACEPA/PATH), Malaria Consortium, The Carter Center (TCC), World Health Organization (WHO), and International Center for AIDS Care and Treatment Programs (ICAP).

The MIS was a large nationally representative survey designed to cover key malaria control interventions, treatment-seeking behavior, malaria prevalence; and also to assess anemia prevalence in children under 5 years of age, malaria knowledge among women, and indicators of socioeconomic status [5]. The survey consisted of a two-stage sample design. The first stage involved selecting clusters from a list of enumerations areas (EAs) covered in the 2007 Population Census, these areas made up the primary sampling units (PSUs). However, in the three regional states that were used in this study, namely Amhara, Oromia and Southern Nations, Nationalities and Peoples' (SNNP) the Kebeles were the clusters. A total of 332 Kebeles or clusters where 91 from Amhara, 153 from Oromia and 88 from SNNP regions were included in the analyses.

Outcome and explanatory variables

The response or outcome variable in this study was an indicator of whether a child under five years of age had been tested positive for malaria, coded as 1, or not, coded as 0. At the survey blood samples were taken from all children under five years of age in every sampled household per WHO guidelines after obtaining consent from residents and assistance of the parent/guardian of the child. Then malaria parasite testing was done using CareStartTM rapid diagnostic tests (RDT).

The predictor variables used to explain the malaria outcomes are defined at two levels, child / household characteristics or factors and Kebele or cluster characteristics or factors. As a child level predictor variables we used child age, gender and the anaemia category of a child which was a four-category variable with the categories no anaemia, mild, moderate and severe anaemia. For a household background variables, we considered those commonly investigated in the malaria literature (e.g. see [26, 16, 27, 17, 28, 29]) and these include number of household members, whether household had mosquito nests that can be used while sleeping or not which was a dichotomous variable with the categories yes or no; whether household sprayed interior walls of the dwelling against mosquitoes within the last 12 months prior to the survey or not which was a dichotomous variable with the categories yes or no; number of rooms in the dwelling, the dwelling has windows which was a dichotomous variable with categories yes and no; main source of drinking water which was a three-category variable with the categories unprotected, protected source and piped water; main material of the house wall which was a three-category variable with the categories no wall, wood and finished wall (e.g. bricks); main material of house floor

which was a three-category variable with the categories earth plastered by dung, rudimentary and finished floor (e.g. cemented); main material of the house roof which was a three-category variable with the categories natural (e.g. thatch / leaf), corrugated iron and wood; household toilet facility which was a three-category variable with the categories no facility, pit latrine and flush toilet; and household wealth index which was a five-category variable with the categories poorest, second, middle, fourth and richest. A cluster or Kebele explanatory variables are region which was a three-category variable with the categories Amhara, Oromiya and SNNP regional states; and median altitude in meters. For the analysis, the variable Kebele was used as a cluster, i.e. level 2, variable and the child / household predictor variables as level-1 characteristics. The MIS data include weight for sampling design at level-1.

Compositional and contextual effects

When individual factors, e.g. age of a child or household has mosquito nets or not, are characteristics of subjects who are more likely to be ill, variability of their distributions across areas will influence health outcomes in a given area, this is called a composition effect. The effects of variables defined at a group or cluster level on outcomes defined at an individual or child level after controlling for relevant individual level confounders are called contextual effects.

The median altitude of a Kebele may be a marker for Kebele-level factors potentially related to malaria infection, such as climate or environmental conditions, and these factors may affect everyone in the Kebele. The notions of compositional and contextual effects apply not only when the focus is upon context as geographical setting but also it can be applied when context is seen in terms of administrative setting [30]. In Ethiopia, the public service delivery is under the jurisdiction of the regional states. The regional health bureaus are responsible for administration of public health while the districts are responsible for planning and implementation of services. So, public health administrative processes that had been taken by district or regional authorities to control malaria transmission in the Kebeles might affected the positive malaria RDT outcomes of children living within a particular Kebele. Therefore, in this study we defined the child / household characteristics (predictors) as composition effects whereas as median altitude and region as contextual effects.

Statistical models for the analysis of multilevel data with binary

outcomes

Let y_{ij} denote the malaria outcome of the i th under five years age child in the j th Kebele or cluster identified by the CareStartTM rapid diagnostic tests (RDT) with probability π_{ij} , where $y_{ij} = 1$ denotes the child tested positive, while $y_{ij} = 0$ denotes the child tested negative for malaria. A two-level generalized linear mixed model (GLMM) or multilevel model with random effects for the outcome y_{ij} is given by

$$\eta_{ij} = g(\mu_{ij}) = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_j^{(2)}, \quad i = 1, \dots, n_j^{(1)}; \quad j = 1, \dots, n^{(2)}, \quad (1)$$

where $g(\cdot)$ is the link function, \mathbf{x}_{ij} and $\mathbf{z}_{ij}^{(2)}$ are vectors of explanatory variables or covariates, $\boldsymbol{\beta}$ is vector of fixed regression coefficients or parameters, $\mathbf{b}_j^{(2)}$ is vector

of random effects varying over Kebeles, $n_j^{(1)}$ and $n^{(2)}$ denote the number of level-1 units, i.e. number of children within the level-2 unit or Kebele j and the number of level-2 units or number of Kebeles, respectively. The subscripts (1) and (2) are used here and throughout the paper to represent level-1 units i and level-2 units j . It is assumed that \mathbf{b}_j is independently and normally distributed with zero mean vector and a $q \times q$ variance-covariance matrix \mathbf{G} , in short $\mathbf{b}_j^{(2)} \sim N(\mathbf{0}, \mathbf{G})$. The conditional expectation $\mu_{ij} = E(y_{ij} | \mathbf{x}_{ij}, \mathbf{b}_j^{(2)})$ is linked to the linear predictor η_{ij} via a link function $g(\cdot)$ and the conditional distribution of y_{ij} belongs to the exponential family. Note that throughout this paper we used two-level GLMM and multilevel model interchangeably because the study data has two levels.

As it was mentioned earlier, the MIS data has sampling weights for PSU. Using a pseudo-maximum-likelihood approach, we can extend the multilevel model framework to accommodate weights at different levels (e.g. see [8]). This approach is very useful in analyzing survey data that arise from multistage sampling. In these sampling designs, survey weights are often constructed to account for unequal sampling probabilities, non-response adjustments, and post-stratification. Suppose that the matrix \mathbf{G} depends on a vector of parameters $\boldsymbol{\theta}$. Specifically the parameter vector $\boldsymbol{\theta}$ describes the $\frac{q(q+1)}{2}$ distinct variance-covariance elements of the matrix \mathbf{G} . Let $w_j^{(2)}$ is the level-2 weight of the level-2 unit or Kebele j and w_{ij} is the level-1 conditional weight of the level-1 unit or child i within the level-2 unit or Kebele j . Then following the notation used in [8], the log-pseudo-likelihood for a multilevel model with random effects is expressed as

$$\log \left(p(y_{ij} | \mathbf{b}_j^{(2)}, \boldsymbol{\beta}, \boldsymbol{\theta}) \right) = \sum_{j=1}^{n^{(2)}} w_j^{n^{(2)}} \log \int \exp \left\{ \sum_{i=1}^{n_j^{(1)}} w_{ij} \log f(y_{ij} | \mathbf{b}_j^{(2)}, \boldsymbol{\beta}, \boldsymbol{\theta}) \right\} g(\mathbf{b}_j^{(2)} | \boldsymbol{\theta}) d\mathbf{b}_j^{(2)}, \quad (2)$$

where $\sum_{i=1}^{n_j^{(1)}} w_{ij} \log f(y_{ij} | \mathbf{b}_j^{(2)}, \boldsymbol{\beta}, \boldsymbol{\theta})$ is the log-likelihood contribution of level-1 units, which are conditional on the random effect $\mathbf{b}_j^{(2)}$ at level-2 and $g(\mathbf{b}_j^{(2)} | \boldsymbol{\theta})$ is the normal density of the random effect $\mathbf{b}_j^{(2)}$. The integrals in equation (2) can be approximated by the Adaptive quadrature (e.g. see [31, 32]).

To make inference about fixed effects and variance-covariance components, we can use the empirical (sandwich) variance estimators. Let $\boldsymbol{\alpha} = (\boldsymbol{\beta}', \boldsymbol{\theta}')$. For a 2-level model, Rabe-Hesketh and Skrondal [8] show that the gradient can be written as a weighted sum of the gradients of the top-level units and is given by

$$\sum_{i=1}^{n^{(2)}} w_i \frac{\partial \log(p(\mathbf{y}_i; \boldsymbol{\alpha}))}{\partial \boldsymbol{\alpha}} \equiv \sum_{i=1}^{n^{(2)}} \mathbf{s}_i(\boldsymbol{\alpha})$$

where $n^{(2)}$ is the number of level-2 units and $\mathbf{s}_i(\boldsymbol{\alpha})$ is the weighted score vector of the level-2 unit i . The estimator of the "meat" of the sandwich estimator can be written as

$$\mathbf{J} = \frac{n^{(2)}}{n^{(2)} - 1} \sum_{i=1}^{n^{(2)}} \mathbf{s}_i(\hat{\boldsymbol{\alpha}}) \mathbf{s}_i(\hat{\boldsymbol{\alpha}})'$$

So, the empirical estimator of the covariance matrix of $\hat{\alpha}$ is given as

$$\mathbf{H}(\hat{\alpha})^{-1} \mathbf{JH}(\hat{\alpha})^{-1}$$

where $\mathbf{H}(\hat{\alpha})$ is the second derivative matrix of the log pseudo-likelihood with respect to α and evaluated at $\hat{\alpha}$.

Scaling sampling weights

In complex surveys, clusters or units within clusters or both may be selected with unequal weights or probabilities and often surveys are subjected to non-response. Modeling of data obtained from such surveys must take into consideration these aspects as they can lead to biased parameter estimates ([25, 8]). When the sampling weights associated with level-1 units, they could lead to bias in variance components estimators if they are large [33]. Furthermore, the variance parameter estimators that are obtained by the pseudo-maximum-likelihood method can be biased when the sample size is small. To correct weights and reduce bias literature recommend scaling the weights before using them in the multilevel modelling (e.g. see [34, 25, 35, 8]). In this work, we have used the following two weight-scaling methods [34] in the fitted multilevel models for reducing possible biases of the variance parameter estimators. Let $n_j^{(1)}$ denote the number of level-1 units in the level-2 unit j and let $w_{i|j}$ denote the weight of the i th level-1 unit in level-2 unit j . The first method, called method 1, scales the weights so that the new weights sum to the effective cluster size [34],

$$\sum_{i=1}^{n_j^{(1)}} \lambda w_{i|j} = \frac{(\sum_{i=1}^{n_j^{(1)}} w_{i|j})^2}{\sum_{i=1}^{n_j^{(1)}} w_{i|j}^2},$$

where λ is the scale factor and it becomes

$$\lambda = \frac{\sum_{i=1}^{n_j^{(1)}} w_{i|j}}{\sum_{i=1}^{n_j^{(n)}} w_{i|j}^2}.$$

The second method, called method 2 [34], scales the weights so that the new weights sum to the cluster sample size. So that the scale factor is

$$\lambda = \frac{n_j^{(1)}}{\sum_{i=1}^{n_j^{(1)}} w_{i|j}}.$$

The multilevel models with random effects, i.e. two-level GLMMs, were fitted with the PROC GLIMMIX in SAS. Since PROC GLIMMIX in SAS uses the weights provided in the data set for analysis, to use the scaled weights the scaled weights should be provided in the data set. Note also that the level-1 and level-2 weights appear in separate places within the pseudo-maximum likelihood estimator function in equation (2), therefore one must take special care to include design weights at the right place in the PROC GLIMMIX.

Results

Child, household and Kebele characteristics

A total of 4,384 under five years age children were in this study, of which 49.82 per cent were male. The mean (SD) age of the children was 2.67 (0.02) years. The total number of members in each household ranged from 2 to 15, with a mean (SD) of 5.78 (1.99) members. Large per cent of the sampled households (47.77%) were in areas with median altitudes at least 2000 meters, 38.18 per cent were in areas with median altitude ranging between 1500 and 2000 meters, and 14.05 per cent were in areas with median altitudes less than 1500 meters. A total of 51.66 , 25.80 and 22.54 per cents of households were sampled from Oromiya, SNNP and Amhara regional states, respectively. The mean (SD) number of rooms in sampled households was 2 (0.95) and ranges between 1 and 9 rooms. About 61.84 per cent of children were lived in dwellings that had at least one window. Over half of the sampled households (53.38%) had no mosquito net. Information about indoor residual spraying of interior walls of the dwelling against mosquitoes within the last 12 months prior to the survey was also collected. Only 29.99 per cent of the sampled households had sprayed at least once within the last 12 months. About 58.6 per cent of the sampled household had unprotected source of drinking water whereas 18.1 and 23.3 per cents of them had protected and piped sources of drinking water, respectively. Almost 80 per cent of sampled households had either bamboo / wood with mud or stone with mud as main material of their dwellings wall, 12.93 per cent of the households had finished wall and 7.33 per cent of households' dwellings either had no wall or uncovered abode or bamboo / reed or carton as main material of their dwellings wall. The finished wall group included cement walls, walls made with stones and cement and bricks. Large per cent (44.68) of the sampled households had either thatch / leaf or rustic mat / plastic sheet as their main roof material, 35.18 per cent of them had either corrugated iron or calamine / cement fiber or cement / concrete and 20.14 per cent of the households had either sticks and mud or reed / bamboo or wood planks or wood or roofing shingles as their main roof material. The majority (93.17%) of the household had earth / sand or dung plastered floor, 4.82 per cent of the households had either wood planks or palm / bamboo as their main floor material and only 2.01 per cent of the households had a finished floor which were either ceramic tiles or cement or carpet. Out of the sampled households, 48.53 and 18.59 per cents of them had pit latrine and flush toilet facilities, respectively, whereas 32.88 of the households had no toilet facility. The wealth index used in this survey serves as an indicator of level of wealth of a household that is consistent with expenditure and income measures. Of the sampled households 22.48, 21.99, 18.86, 18.90 and 17.77 per cents of them were in the poorest, second poorest, middle, fourth and the richest (or highest) wealth quintile, respectively. Overall, 2.15 of the children had severe anaemia, 13.98 per cent had moderate anaemia, 14.16 per cent had mild anaemia and 69.71 per cent had no anaemia.

Fitted models

We fitted three multilevel models with Kebele-specific random effects. The first was the null model, denoted by M_0 , which did not contain any child / household or

Kebele level characteristics. It had only Kebele-specific random effects b_j and fitted to verify if there is indeed variation between Kebeles in under five years age children malaria RDT outcomes, that is

$$\text{logit}(\pi_{ij}) = \beta_0 + b_j,$$

where β_0 is identical for all Kebeles and b_j quantify differences between what is measured on average in the study area and what is measured in each Kebele. It is assumed that Kebele-specific random effect $b_j \sim \mathcal{N}(0, \sigma_b^2)$. Note that $\mathbf{G} = \sigma_b^2 \mathbf{I}$. In the second model we included the child / household-level predictor variables in M_0 , called M_1 , i.e. it has a form

$$\text{logit}(\pi_{ij}) = \sum_{k=0}^K \beta_k x_{ijk} + b_j.$$

The coefficients β_k , $k = 1, \dots, K$, where K is the total number of coefficients which depends on number of categories of predictors in the model, are fixed effect parameters and b_j is as defined in model M_0 . The third model defined using both child / household-level and two Kebele / cluster-specific predictor variables, i.e. it was M_1 with two Kebele - specific predictor variables (region and median altitude), called M_2 and has a form

$$\text{logit}(\pi_{ij}) = \sum_{k=0}^K \beta_k x_{ijk} + \sum_{l=1}^L \alpha_l v_{jlk} + b_j,$$

where α_l , $l = 1, \dots, L$, where L is the total number of coefficients for Kebele level predictors and it depends on number of categories of predictors, are the fixed effects for Kebele-level predictors and v_{jlk} is the k th observed value of the l th predictor at the j th Kebele. Note that the regression coefficients or fixed effects β_k and α_l in the above models represent the study area average effects whereas the Kebele-level variance σ_b^2 provides an estimate of what could be explained by each Kebele-level.

As mentioned earlier, the MIS data used in this study include a single overall level-1 weighting variable that incorporates level-2 design issues and the weights account for unequal probability of selection given different population sizes within Kebeles. As the study data do not have specific weight for level-2, here we did not weight level-2 in the analyses. However, as Rabe-Hesketh and Skrondal [8] indicated scaling level-2 weights has little practical effect. The above multilevel models were fitted after scaling the survey weight using the two scaling methods (1 and 2). The weighted analyses results across the fixed and variances of the random effects were almost identical at two decimal places. Even though the results obtained slightly different between these methods, we had the same inferential decisions for child / household- and Kebele-level characteristics in each of the fitted model. As we are interested in to discuss the point estimates, such as intercepts and odds ratios we reported here the results from scaling Method 1 [9].

The asymptotic $0.5 \chi_0^2 + 0.5 \chi_1^2$ mixture distribution [36] test statistic for testing $H_0 : \sigma_b^2 = 0$ against $H_1 : \sigma_b^2 > 0$ in models M_0 , M_1 and M_2 takes the values

$RLRT = 136.91, 48.57$ and 29.83 , respectively with p -value < 0.0001 in all the three tests. The large value of the test statistic or a very small p -value strongly suggests a rejection of the null hypothesis $H_0 : \sigma_b^2 = 0$ that no Kebele-specific random effects should be included in the model. Therefore, these results imply the need for the Kebele-random effects in the model.

The Type III tests for the predictor variables or fixed effects in Table 1 show that eight of the 14 child / household characteristics, namely the age of the child, number of rooms in the dwelling, whether the dwelling had windows or not, whether the household had any mosquito nets that can be used while sleeping or not, whether the household sprayed the interior walls of the dwelling against mosquitoes within 12 months before the survey or not, type of main roof material, the kind of toilet facility the household where the child lived had and the anaemia status or category of the child were significantly associated with the child malaria RDT outcome at 5% level in model M_1 . These eight variables and the two Kebele characteristics, region and the median altitude were significantly associated with the child malaria RDT outcome at 5% level of significant in model M_2 , i.e. the model that including both child / household characteristics and the two Kebele characteristics.

Table 1 Type III tests for the Covariates or predictor variables for the binary GLMMs.

Covariates	Numerator DF	p -valu	
		Model M_1	Model M_2
Age	1	0.0006	0.0014
Gender, Female	1	0.4202	0.3509
Household members	1	0.4715	0.3958
Number of rooms	1	0.0207	0.0050
Windows exist, Yes	1	0.0227	0.0404
Mosquito nets, Yes	1	0.0003	0.0100
Sprayed within 12 months	1	0.0001	0.0022
Drinking water	2	0.6895	0.4617
Main wall material	2	0.4539	0.2105
Main roof material	2	0.0200	0.0397
Main floor material	2	0.5515	0.3746
Toilet facility	2	0.0024	0.0108
Household wealth index	4	0.2233	0.3548
Anaemia category	3	< 0.0001	< 0.0001
Region	2		0.0009
Median Altitude	1		0.0020

Fixed effects and odds ratios

The estimated intercept in M_0 was -5.485 , while the estimated variance of the random effects was 4.248 with a standard error of 0.906 . Thus, the probability for under five years age child who lived in any Kebele where the study conducted tested positive for malaria was 0.41% given that the random effect of the Kebele was equal to zero on the logit scale. Thus, taking the inverse logit transformation of the interval for $\hat{\beta}_0$, for 95% Kebeles, the Kebele-specific probability under five years age child who lived in the areas had tested positive for malaria would lie in the interval

(0.0001, 0.2357). Since $\exp(\hat{\beta}_0 + b_j)/(1 + \exp(\hat{\beta}_0 + b_j)) \neq \exp(\hat{\beta}_0)/(1 + \exp(\hat{\beta}_0))$ for $b_j \neq 0$, the average Kebele predicted probability of under five years age child tested positive for malaria may differ from the average Kebele-specific probability of malaria infection. The regression coefficients associated with the child / household- and Kebele-level are the fixed effects (β_k and / or α_l) part of the fitted models M_1 and M_2 . The estimated values of these effects (with corresponding standard errors in brackets) and the p -values to test $\beta_k = 0$ are given in Table 2. The estimates in this table are in the log odds scale, however by taking exponential of the estimates, one can easily obtain estimates of the odds ratio (OR) of under five years age child tested positive for malaria relative to an appropriate reference group. These are given in Table 3 with their 95% confidence intervals (CI) in brackets. The p -values in Table 2 and the 95% confidence interval of odds ratios in Table 3 revealed similar conclusion drawn from Type III tests on the association of malaria RDT outcomes with child / household- and Kebele-level characteristics.

These results (model M_1) revealed that the odds of malaria increased as the age of a child increased (OR = 1.481 with 95% CI (1.185, 1.853)). The OR of positive malaria RDT outcome for female child relative to male child was 0.825 with 95% CI (0.516, 1.318). The 95% confidence interval contains 1 suggesting that the odds of positive malaria RDT outcomes were not significantly different between male and female groups, however female children were less likely infected by malaria in the study areas. The positive malaria RDT outcome slightly decreased as the number of household members increased but this decrease was statistically non-significant at 5% level (OR = 0.947 with 95% CI: 0.816, 1.099). Children lived in dwellings with windows were almost twice more likely to had malaria than those lived in dwellings without windows (OR = 2.042, 95% CI: 1.105, 3.773). However, relative to children in the households without mosquito net, the OR of positive malaria RDT outcomes was 3.132 (95% CI: 1.686, 5.820) for those in the households with mosquito nets.

Spraying of interior walls of the dwelling against mosquitoes within the last 12 months prior to the survey significantly reduced (by 50.5%) the odds of malaria infection relative to those children residing in dwellings that had not been sprayed during this period (95% CI: 0.345, 0.710). Compared to children in households that had unprotected source of drinking water, the ORs of positive malaria RDT outcomes for those in households that had protected source and piped water as the main sources of drinking water were 0.705 (95% CI: 0.316, 1.575) and 0.896 (95% CI: 0.396, 2.029), respectively. Compared to children in dwellings with bamboo / wood with mud or stone with mud as the main wall material, those in dwellings with no walls or walls constructed from cane / trucks / bamboo /reed, uncovered abode or carton, and those in dwellings with finished walls were associated with a higher risk of malaria, with the ORs of positive malaria RDT outcome 1.307 (95% CI: 0.502, 3.401). Whereas those children residing in dwellings with finished wall (cement walls, walls made with stones and cement and bricks) had less odds of positive malaria RDT outcome 0.624 (95% CI: 0.283, 1.373).

Relative to children residing in dwellings with corrugated iron as the main roof material, the OR of positive malaria RDT outcomes was about three times more (OR = 3.074 with 95% CI: 1.318, 7.169) for those residing in dwellings with natural roof (thatch / leaf and rustic mat / plastic sheet) and increased by 8.1% (OR =

Table 2 Estimated regression coefficients (SE) and variance components for the two-level GLMMs.

Predictor	Fitted models				
	Model M_0 Estimate (SE)	Model M_1 Estimate (SE)	p -value	Model M_2 Estimate (SE)	p -value
<i>Child related variables</i>					
Intercept	-5.485 (0.344)	-9.478 (0.8957)	< 0.0001	-8.060 (0.839)	< 0.0001
Age		0.393 (0.114)	0.0006	0.356 (0.112)	0.0014
Gender, Female		-0.193 (0.239)	0.4202	-0.219 (0.234)	0.3509
Household members		-0.055 (0.076)	0.4715	-0.065 (0.076)	0.3958
Number of rooms		0.277 (0.120)	0.0207	0.339 (0.121)	0.0050
Windows exist, Yes		0.714 (0.313)	0.0227	0.659 (0.321)	0.0404
Mosquito nest, Yes		1.142 (0.316)	0.0003	0.740 (0.287)	0.0100
Sprayed within 12 months		-0.703 (0.184)	0.0001	-0.756 (0.247)	0.0022
Drinking water (Ref Unprotected)					
Protected source		-0.349 (0.410)	0.3947	-0.494 (0.406)	0.2231
Piped water		-0.120 (0.417)	0.7924	-0.129 (0.409)	0.7535
Main wall material (Ref Wood)					
No wall		0.268 (0.488)	0.5835	0.534 (0.437)	0.2223
Finished wall, e.g. bricks		-0.472 (0.403)	0.2408	-0.473 (0.380)	0.2131
Main roof material (Ref Corrugated iron)					
Natural		1.123 (0.432)	0.0094	1.158 (0.459)	0.0117
Wood or other		0.078 (0.383)	0.8394	0.267 (0.373)	0.4751
Main floor material (Ref Finished)					
Rudimentary		0.338 (0.389)	0.3838	0.448 (0.373)	0.2294
Earth / dung plaster		0.483 (0.562)	0.3896	0.593 (0.548)	0.2797
Toilet facility (Ref No facility)					
Flush toilet		1.763 (0.533)	0.0009	1.453 (0.491)	0.0031
Pit latrine		0.879 (0.447)	0.0496	0.781 (0.404)	0.0533
Household wealth index (Ref Richest)					
Poorest		1.383 (0.712)	0.0521	1.154 (0.706)	0.1021
Second		1.281 (0.620)	0.0389	1.124 (0.617)	0.0685
Middle		1.170 (0.521)	0.0248	1.032 (0.518)	0.0464
Fourth		0.244 (0.323)	0.4491	0.244 (0.310)	0.4306
Anaemia category (Ref No anaemia)					
Mild		0.582 (0.348)	0.0942	0.466 (0.348)	0.1807
Moderate		1.745 (0.320)	< 0.0001	1.534 (0.330)	< 0.0001
Severe		2.722 (0.482)	< 0.0001	2.676 (0.502)	< 0.0001
Region (Ref Amhara)					
Oromiya				-1.625 (0.452)	0.0003
SNNP				-0.174 (0.409)	0.6715
Median Altitude				-0.002 (0.001)	0.0002
<i>Variance of random effects</i>					
σ_b^2	4.248(0.906)	2.131 (0.561)		1.860 (0.659)	

1.081 with 95% CI: 0.51, 2.317) for those residing in dwellings with wood / wood planks, reed / bamboo, sticks and mud, calamine/cement fiber and other as the

main roof material. Children residing in dwellings with earth/sand and/or dung plastered as the main floor material were most likely at the highest risk for malaria compared to those in dwellings with a finished floor surface (OR = 1.621 with 95% CI: 0.539, 4.873). Those who residing in dwellings with rudimentary floor (either wood planks or palm/bamboo) as the main floor material also had higher risk (OR = 1.403 with 95% CI: 0.655, 3.004) compared to those in dwellings with a finished floor surface. The risk of anaemia was found to be highly associated with malaria RDT outcome where the odds of malaria increased as the severity of anaemia of a child increased. The odds of positive malaria RDT outcome for under five years age children with mild, moderate and severe anaemia status were 1.790, 5.728 and 15.211 times more than that of the non-anaemic children, respectively. However, the increased for mild status was not statistically significant (p -value = 0.0942 and 95% CI for OR: 0.905, 3.538). Relative to those children with richest households, those with poorest, second, middle and fourth wealth index categories had 3.985, 3.600, 3.226 and 1.277 times higher odds of positive malaria RDT outcomes, respectively.

Compared to children who resided in the Amhara region, children who resided in Oromiya and SNNP regional states were less at risk with the odds ratios ranging 0.227 and 0.852, respectively (see model M_2 results in column 4 of Table 3). However, the difference was significant only between Amhara and Oromiya regions (p -value = 0.0013 and 95% CI for OR: 0.092, 0.559). The result in model M_2 revealed that the odds of malaria for a child decreased with an increase in median altitude (OR = 0.998 with 95% CI: 0.996, 1.000), specifically, the odds decreased by approximately 20% as the Kebele median altitude increased by 100 meters. Observe from both Table 2 and Table 3 that we had similar conclusions for child / household predictors in model M_2 as that of M_1 .

Between Kebeles variance

Considering the three models M_0 , M_1 and M_2 , the Kebele-level variance decreased as the child / household- and Kebele-level characteristics were introduced (see Table 2), where $\hat{\sigma}_b^2 = 4.248, 2.131$ and 1.860 for models M_0 , M_1 and M_2 , respectively. These suggest that when accounting for the child / household- and Kebele-level characteristics, the part of the variability which is relevant at the Kebele-level (level-2) becomes lower. These also mean that the Kebele-level variance quantifies part of the variability which is relevant at Kebele level but not explained, for example in models M_2 by Kebele-level characteristics or predictors introduced in the model [20]. Generally, if the Kebele-level characteristics are relevant, then they would be associated with the RDT outcome, which was the case here and they would also explain the Kebele-level variance from M_1 . The Kebele-level predictor variables explained about 12.72% of the variance of the random effects from model M_1 . The percentage of Kebele-level variance explained by the two Kebele-level characteristics, region and median altitude was 56.21% ($= (4.248 - 1.860) / 4.248 \times 100$). Hence, in the model M_2 , 43.79% of the Kebele-level variance remain unexplained, indicating that some unmeasured or unknown Kebele characteristics could be missing [20]. Furthermore, recall that the two sets of predictors, i.e. child / household and Kebele, appeared to influence significantly the likelihood of positive malaria RDT outcomes.

Table 3 Estimated odds ratios (95% confidence intervals) for the two-level GLMMs.

Covariates	Odds ratio (95% CI)	
	Model M_1	Model M_2
Age	1.481 (1.185, 1.853)	1.414 (1.131, 1.768)
Gender, Female	0.825 (0.516, 1.318)	0.807 (0.505, 1.287)
Household members	0.947 (0.816, 1.099)	0.946 (0.811, 1.103)
Number of rooms	1.319 (1.043, 1.668)	1.418 (1.107, 1.816)
Windows exist, Yes	2.042 (1.105, 3.773)	1.692 (0.866, 3.307)
Mosquito nets, Yes	3.132 (1.686, 5.820)	1.969 (1.114, 3.483)
Sprayed within 12 months, Yes	0.495 (0.345, 0.710)	0.470 (0.289, 0.762)
Drinking water (Ref Unprotected)		
Protected source	0.705 (0.316, 1.575)	0.652 (0.292, 1.457)
Piped water	0.896 (0.396, 2.029)	0.933 (0.413, 2.105)
Main wall material (Ref Wood)		
No wall	1.307 (0.502, 3.401)	2.013 (0.907, 4.465)
Finished wall, e.g. bricks	0.624 (0.283, 1.373)	0.765 (0.365, 1.600)
Main roof material (Ref Corrugated iron)		
Natural	3.074 (1.318, 7.169)	3.074 (1.295, 7.827)
Wood or other	1.081 (0.51, 2.317)	1.081 (0.629, 2.713)
Main floor material (Ref Finished)		
Rudimentary	1.403 (0.655, 3.004)	1.859 (0.885, 3.903)
Earth / dung plaster	1.621 (0.539, 4.873)	1.885 (0.631, 5.626)
Toilet facility (Ref No facility)		
Flush toilet	5.830 (2.052, 16.567)	4.787 (1.755, 13.054)
Pit latrine	2.407 (1.002, 5.784)	2.352 (1.046, 5.288)
Household wealth index (Ref Richest)		
Poorest	3.985 (0.988, 16.073)	3.690 (0.895, 15.202)
Second	3.600 (1.068, 12.139)	3.161 (0.885, 11.286)
Middle	3.222 (1.160, 8.949)	2.713 (0.926, 7.944)
Fourth	1.277 (0.678, 2.402)	1.468 (0.775, 2.780)
Anaemia category (Ref No anaemia)		
Mild	1.790 (0.905, 3.538)	1.563 (0.762, 3.205)
Moderate	5.728 (3.058, 10.726)	5.226 (2.726, 10.020)
Severe	15.211 (5.918, 39.093)	13.828 (5.038, 37.951)
Region (Ref Amhara)		
Oromiya		0.227 (0.092, 0.559)
SNNPR		0.852 (0.368, 1.974)
Median Altitude		0.998 (0.996, 1.000)

Discussion

Analysis results

This study focused on malaria indicator survey data analysis using the weighted multilevel models, specifically on under five years age children RDT outcomes in three largest regional states in Ethiopia. Beside identifying significant risk factors associated with malaria infection, the multilevel analyses allowed us to examine within- and between Kebele variability in malaria RDT outcomes for under five years age children in the study areas and of the extent to which between Kebeles

was explained by child / household and Kebele-level factors. In the analyses, the overall variability has been partitioned in child/household-level and a Kebele-level variability and the results revealed that after controlling for child/household factors, Kebele-level variability reduced which shows the presence of composition effects of these factors, that is some of the differences between Kebele in under five years age children malaria RDT outcomes explained by child/household factors, see model M_1 result in Table 2. Part of the remaining Kebele-level variability was then partially explained and reduced when taking into account Kebele factors, see model M_2 result in Table 2.

The results of this study clarified the association between the malaria RDT outcomes of children under five years of age and some of the characteristics of children that their vulnerability to malaria infection increases significantly with increase in age and with the severity of their anaemia status but child's gender has no association with the outcome. The result on age agrees with the findings of ([37, 26, 38, 39, 40, 13, 17]), anaemia results do agree with recent WHO report (WHO2019) and the gender finding is similar to those of Baragetti, et al [41], Ayele, Zewotir, and Mwambi [26], Roberts and Matthews [13] and Ugwu and Zewotir [17]. In the study, it was observed that as the number of household members increased the risk of malaria infection in under five years age children decreased, however this association was statistically nonsignificant.

The household characteristics, specifically socio-economic factors such as number of rooms in the dwelling, whether the dwelling had windows or not and the household had mosquito nets that can be used while sleeping were significantly positively associated with child's malaria infection. The positive association of positive malaria RDT outcomes with the number of rooms and the dwelling had windows could be more rooms and windows created more entry points for the mosquito if these windows were not closed properly or during peak biting times. The positive associate of having mosquito nets with positive malaria RDT outcomes could be due to either households did not treat the nets with insecticide or households did give priority to adults to use the nets if they had no enough nets for everyone or inappropriate use of the nets or perhaps a child was exposed to mosquito bites during other times of the day or evening when the net was not in use. This also could be due to household did not get proper training on how to use the nets from local public health workers. However, incidence of a household had sprayed interior walls of the dwelling against mosquito in the past 12 months before the survey was significantly negatively associated with a child's malaria RDT outcome where the risk of malaria infection was substantially reduced in the event of household had sprayed interior walls of the dwelling against mosquitoes.

The results also showed that good quality source of drinking water was associated with a lower risk of malaria compared to unprotected source. The unprotected sources such as irrigation ditches / channels, or dam can create breeding sites for larvae. Compared to a child who was resided in dwelling with wooden wall (bamboo / wood with mud), those resided in dwellings with no wall or poor quality wall, e.g. made from uncovered abode or cane / trucks / bamboo /reed had high risk of malaria infection, whereas those resided in dwellings with finished wall surface of cement, stone with lime / cement, bricks or cement blocks had very low risk

of malaria infection. These results agree with the findings of Ghebreyesus et al. [42], Ayele, Zewotir, and Mwambi [26], Woyessa et al [43], Okebe et al [44] and Lwetoijera et al [45] where they found poor housing quality facilitate mosquito entry and hence exposed children for malaria infection. Reference to dwellings with finished floors (parquet or polished wood, ceramic tiles and cement), children resided in dwellings with natural floor (earth / sand plaster with dung) and rudimentary floors (wood planks and palm / bamboo) had higher risk of malaria infection and these results agree with Ayele, Zewotir, and Mwambi [26]. However, these results were statistically nonsignificant. The wealth status of the households had negative effect on the positive malaria RDT outcomes. That is, compared to the richest households, the study revealed that under five years age children from the poorest households were at highest risk of malaria infection followed by those children from households who had been in the second, middle and fourth wealth index categories.

The region that a child was resided significantly associated with the risk of malaria infection, where children resided in Oromiya and SNNP regional states had relatively lower risk compared to children in Amhara regional state. The result also showed that the Kebele median altitude negatively associated with the positive malaria RDT outcomes. That is children resided in lower altitude areas were at higher risk of malaria infection this was due to the fact that as altitude decreases environment becomes more favourable to malaria development.

Multilevel model

The simultaneously inclusion of characteristics of individuals living in different areas, i.e. composition factors and area-level characteristics, i.e. contextual factors in multilevel models with individuals as the units of analysis allowed to examine the Kebele effects after the child / household-level confounders have been controlled. This approach requires data sets including individuals nested within areas or the data sets should have hierarchical structure that allow to define composition factors and exposure group. One of the challenges, however is defining relevant geographical areas where the attention or research hypothesis is focussed on. These could be communities, neighbourhoods or areas, as Diez-Roux [46] suggested generally these refer person's immediate residential environments whose characteristics may be relevant to the specific health outcome being studied.

The effects of individual-level variables may differ by contextual characteristics or vice-versa. In other words, cross-level interactions involving individual- or contextual-level variables may occur. This can also easily be incorporated in the analysis and changes in the amount of variation due to the cross-level interactions can be determined. For example, studies show that the child anaemia prevalence depends on region (e.g. see [47]), therefore we have considered a model that include cross-level interaction, i.e. region by anaemia category interaction to see their joint effect. The inclusion of this term explained only 2% of the Kebele-level variability therefore the result is not reported in this paper. However, it should be noted that a cross-level interaction between a level-1 and a level-2 variable behaves like a level-1 variable [48]. That is, the cross-level interaction varies between children living in the same Kebele and so might explain part of the variation within Kebeles.

Malaria presence depends mainly on climatic factors such as temperature, humidity, and rainfall. Malaria is transmitted in tropical and subtropical areas, where *Anopheles* mosquitoes can survive and multiply, and where malaria parasites can complete their growth cycle in the mosquitoes. Human behavior factors, often dictated by social and economic reasons, can influence the risk of malaria for individuals and communities. For example, human activities can create breeding sites for larvae such as standing water in irrigation ditches and burrow pits; and population movement and migration [3]. In this study, however the contextual factors were defined at the Kebele level where the household resided. Therefore, relative to people activities and mobility, this level could hide part of the environment. As many dimensions and determinants of environment interrelated, interpretation of results concerning area factors is usually complex [46]. The differences between Kebeles could be due to bio-ecological or human factors. Further, the climate could affect the malaria cycle and vector development, but these were not introduced in the models because the MIS data do not contain these measurements. However, these factors are strongly correlated with altitude, which synthesizes climatic and vegetal conditions. Therefore, part of the altitude influence certainly reflects the influence of climatic factors on malaria.

Conclusion

In this paper, we demonstrate the use of weighted multilevel model to examine within- and between Kebele variability in malaria RDT outcomes for under five years age children and of the extent to which between Kebeles variability is explained by child / household and Kebele-level factors.

Malaria related infection, morbidity and mortality can be prevented and reduced through interventions such as vector control methods namely, long-lasting insecticidal nets and indoor residual spraying of households with insecticides and including case diagnostic testing and treatment. In Ethiopia the interventions are administered by the regional health bureaus which are responsible for administration of public health while the districts or Kebeles are responsible for planning and implementation of services. The Kebele-level factor, i.e. region, in the model could explain part of the variation between Kebeles that possibly caused by the interventions process across the administrative regions.

The Kebele-level variance can be interpreted as heterogeneity across Kebeles for the probability of under five years age child was tested positive for malaria. As this variance is not attributable to individual- and Kebele-level variables (where here individual level variables are child and household characteristics) that were in the model, the random effect could then be considered as a replacement of unmeasured Kebele-level factors that possibly influence the malaria RDT outcome. The human activities on the environment might contribute to between Kebele variability. As mentioned earlier, if data available on the activities, beside modeling the variability via random effect one can add cross-level interactions between individual- and Kebele-level characteristics and assess their significance too. Thus, the multilevel modelling allows investigation of how individual- and Kebele-level factors including the public health authorities' interventions plans related to health outcomes simultaneously. Such collective assessment approach may lead to more effective public

health strategies and could have important policy implications for health promotion and for the reduction of health disparities.

Declaration

Ethics approval and consent to participate

Ethical approval to conduct the study was obtained from the Health and Nutrition Research Institute and ethical committee of College of Science, Engineering and Technology, University of South Africa.

Consent for publication

Not applicable.

Availability of data

The data that support the findings of this study are available from Ethiopian Health and Nutrition Research Institute but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Ethiopian Health and Nutrition Research Institute.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

BTZ and LKD conceptualized the statistical method used in the study; BTZ performed the analyses; BTZ drafted the manuscript; LKD reviewed the findings of data analyses and the writing of the manuscript. All authors approved the final manuscript.

Acknowledgements

We thank the Ethiopian Health and Nutrition Research Institute for giving access to the data.

Author details

¹Department of Statistics, College of Natural and computational Science, Wolaita Sodo University, Wolaita Sodo, Ethiopia. ²Department of Statistics, University of South Africa c/o Christiaan de Wet Road & Pioneer Avenue, Private Bag X6, Florida 1710, Johannesburg, South Africa.

References

1. Westhuizen, J.H.V.-d., Ngarivhume, T., Klooster, C.A.V., Jong, J.M.D.: Medicinal plants used by traditional healers for the treatment of malaria in the chipinge district in zimbabwe. *Journal of Ethnopharmacology* **159**, 224–237 (2015)
2. World Health Organization: World Malaria Report 2019. Geneva (2019). World Health Organization
3. Malaria's Impact Worldwide. https://www.cdc.gov/malaria/malaria_worldwide/impact.html. Accessed on 7 June 2020.
4. Ethiopian Public Health Institute: Ethiopia National Malaria Indicator Survey 2015. Addis Ababa, Ethiopia (2016). Ethiopian Public Health Institute
5. Ethiopian Health and Nutrition Research Institute: Ethiopia National Malaria Indicator Survey 2011. Addis Ababa, Ethiopia (2012). Ethiopian Health and Nutrition Research Institute
6. NMC, P.: Nigeria mis final report (2015)
7. TLongford, N.: Model-based variance estimation in surveys with stratified clustered designs. *J R Stat Soc [Ser A]* **169**, 805–827 (2006)
8. Hesketh, R.S., Skrondal, A.: Multilevel modelling of complex survey data. *Aust. J Statist.* **38**, 333–352 (1996)
9. Carle, A.C.: Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Medical Research Methodology* **9**(49) (2009)

10. Merlo, J., Chaix, B., Yang, M., Lynch, J., Rastam, L.: A brief conceptual tutorial on multilevel analysis in social epidemiology: Interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiology & Community Health* **59**(12), 1022–1028 (2005)
11. Liang, K., Zeger, S.: Regression analysis for correlated data. *Annu Rev Public Health* (1993)
12. Burton, P., Gurrin, L., Sly, P.S.: Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. *Stat Med* **17**, 261–291 (1998)
13. Roberts, D., Matthews, G.: Risk factors of malaria in children under the age of five years old in uganda. *Malar J* (2016)
14. Ye, Y., Duah, D.: The president's malaria initiative contributed to reducing malaria burden in sub-saharan africa between 2004 and 2014: Evidence from generalized estimating equation analysis. *PLoS ONE* **14**(5) (2019). <https://doi.org/10.1371/journal.pone.0217103>
15. Shah, M.P., Steinhart, L.C., Mwandama, D., Mzilahowa, T., Gimnig, J.E., Bauleni, A., Wong, J., Wiegand, R., Mathanga, D.P., Lindblade, K.A.: The effectiveness of older insecticide-treated bed nets (itns) to prevent malaria infection in an area of moderate pyrethroid resistance: results from a cohort study in malawi. *Malar J* (2020)
16. Ayele, D., Zewotir, T., Mwambi, H.: The risk factor indicators of malaria in ethiopia. *Int J Med Med Sci* **5** (2013)
17. Ugwu, C.L.J., Zewotir, T.: Using mixed effects logistic regression models for complex survey data on malaria rapid diagnostic test results. *Malar J* **17**(453) (2018)
18. Rodriguez, G., Goldman, N.: An assessment of estimation procedures for multilevel models with binary responses. *J R Stat Soc Ser B (Statistical Methodology)* **158**, 73–78 (1995)
19. Goldstein, H., Brown, W., Rasbach, J.: Multilevel modelling of medical data. *Stat Med* **21**, 3291–3315 (2002)
20. Goldstein, H.: *Multilevel Statistical Models*. John Wiley & Sons Ltd., United Kingdom (2011)
21. Snijders, T., Bosker, R.: *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications, London (2012)
22. Austina, P.C., Merlod, J.: Intermediate and advanced topics in multilevel logistic regression analysis. *Stat Med* **36**, 3257–3277 (2017)
23. Omar, R.Z., Thompson, S.G.: Analysis of a cluster randomized trial with binary outcome data using a multilevel mode. *Stat Med* **19**, 2675–2688 (2000)
24. Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Rastam, L., Larsen, K.: A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology & Community Health* **60**(4), 290–297 (2006)
25. Grilli, L., Pratesi, M.: Weighted estimation in multilevel ordinal and binary models in the presence informative sampling designs. *Surv Methodol* **30**, 93–103 (2004)
26. Ayele, D., Zewotir, T., Mwambi, H.: Prevalence and risk factors of malaria in ethiopia. *Malar J* **11**(195) (2012)
27. Guerra, M., Bruno, D.S., Nicolas, N.M., Pedro, B., Ana, P.Z.: Malaria determining risk factors at the household level in two rural villages of mainland equatorial guinea. *Malar J* **17**(203) (2018)
28. Panda, B., Mohapatra, M.K., Paital, S., Kumbhakar, S., Dutta, A., S, S.K., et al.: Prevalence of afebrile malaria and development of risk-scores for gradation of villages: A study from a hot-spot in odisha. *PLoS ONE* **14**(9) (2019). <https://doi.org/10.1371/journal.pone.0221223>
29. Hasyim, H., Dale, P., Groneberg, D.A., Kuch, U., Müller, R.: Social determinants of malaria in an endemic area of indonesia. *Malar J* **18**(134) (2019)
30. Duncan, C., Jones, K., Moon, G.: Context, composition, and heterogeneity: using multilevel models in health research. *Soc Sci Med* **46**, 97–117 (2019)
31. Hesketh, R.S., Skrondal, A., Pickles, A.: Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata J* **2**, 1–21 (2002)

32. Hesketh, R.S., Skrondal, A.: *Multilevel and Longitudinal Modeling Using Stata, Volume II: Categorical Responses, Counts, and Survival*, 3rd Edition. Stata Press, Texas, USA (2012)
33. Lucas, J.P., Sébille, V., Tertre, A.L., Strat, Y.L., Bellanger, L.: Multilevel modelling of survey data: impact of the two-level weights used in the pseudolikelihood. *J Appl Stat* **41**(4), 716–732 (2014). DOI: 10.1080/02664763.2013.847404
34. Pfeiffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H., Rasbash, J.: Weighting for unequal selection probabilities in multilevel models. *J. R. Stat. Soc. Ser. B, Stat. Methodol* **60** (1998)
35. Asparouhov, T.: General multilevel modeling with sampling weights. *Communications in statistics, Theory and methods* **35**(3), 439–460 (2006)
36. Stram, D., WLee, J.: Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**(3), 1171–1177 (1994)
37. Clark, T.D., Greenhouse, B., Njama-Meya, D., Nzarubara, B., Maiteki-Sebuguzi, C., Staedke, S.G.: Factors determining the heterogeneity of malaria incidence in children in kampala, uganda. *J Infect Dis* **198**(156), 393–400 (2008)
38. BAdigun, A., Gajere, E.N., Oresanya, O., Vounatsou, P.: Malaria risk in nigeria: Bayesian geostatistical modelling of 2010 malaria indicator survey data. *Malar J* **14**(156) (2015)
39. Dawaki, S., Al-Mekhlafi, H.M., Ithoi, I., Ibrahim, J., Atroosh, W.M., Abdulsalam, A.M., et al: Is nigeria winning the battle against malaria? prevalence, risk factors and kap assessment among hausa communities in kano state. *Malar J* (2016)
40. MDogara, M., Ocheje, A.J.: Prevalence of malaria and risk factors among patients attending dutse general hospital, jigawa state, nigeria. *Int J Pub Environ Health* **11**, 270–277 (2016)
41. Baragetti, M., Fournet, F., Henry, M., Assi, S., Ouedraogo, H., Rogier, C.: Social and environmental malaria risk factors in urban areas of ouaga-dougou, burkina faso. *Malar J* **8**(13) (2009)
42. Ghebreyesus, T.A., Haile, M., Witten, K.H., et al.: Household risk factors for malaria among children in the ethiopian highlands. *Trans Roy Soc Trop Med Hyg* **94**, 17–21 (2000)
43. Woyessa, A., Deressa, W., Ali, A., Lindtjorn, B.: Malaria risk factors in butajira area, south-central ethiopia. *Malar J* **12**(273) (2013)
44. Okebe, J., Mwesigwa, J., Kama, E.L., Ceesay, S.J., Njie, F., et al.: A comparative case control study of the determinants of clinical malaria in the gambia. *Malar J* **13**(306) (2014)
45. Lwetoijera, D., Kiware, S., Mageni, Z., Dongus, S., Harris, C., et al.: A need for better housing to further reduce indoor malaria transmission in areas with high bed net coverage. *Parasit Vectors* **6**(57) (2013)
46. Diez-Roux, A.V.: Investigating neighborhood and area effect on health. *Am J Public Health* **91**, 1783–1789 (2001)
47. Ngwira, A., Kazembe, L.: Analysis of severity of childhood anemia in malawi: a bayesian ordered categories model. *Open Access Medical Statistics* **6**, 9–20 (2016). <https://doi.org/10.2147/OAMS.S95159>
48. Leyland, A.H., Groenewegen, P.P.: *Multilevel Modelling for Public Health and Health Services Research: Health in Context*. Springer, Switzerland (2020)