MULTILEVEL ANALYSIS APPLIED TO BINARY DATA: MALARIA PREVALENCE IN SAURI MILLENIUM VILLAGE, KENYA

BY

LELERAI L. ELIUD

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN APPLIED STATISTICS

School of Mathematics, Statistics and Actuarial Science

Maseno University

ABSTRACT

Millennium Villages Project is an initiative that is meant to demonstrate that the millennium development goals could be achieved in an integrated approach in putting a combination of interventions in place. Among these interventions are the health interventions of reducing the prevalence of common diseases. Malaria is one of these common diseases. In 2005 Sauri Millennium Village was started in western Kenya which was then followed by 13 other villages across Africa. A baseline study was done in 2005 to measure the bench marks of the millennium development goals indicators in the village. As part of these surveys, blood data was collected to estimate the baseline prevalence of malaria in the Sauri Millennium village. This data was linked to socioeconomic data to study factors affecting malaria prevalence. Malaria affects individuals who are clustered in households and villages. In addition to individual effects, households and villages have characteristics that influence malaria prevalence. The individual characteristics under study were age and gender. The household characteristics were income and the education status of the household the individual belongs. The village level factors were the counts of water bodies and the area covered by woods of the villages. Logistic regression models were applied to understand the determinants of malaria. Considering the multilevel structure of data, the analysis goes beyond the single-level modelling and explores the value of multilevel modelling in understanding the malaria risk factors. The analysis showed that malaria prevalence among the population at baseline was about 50% and was similar for males and females. The results also showed that malaria prevalence decreases with age. Income and education status of the households were also found to have an effect on malaria prevalence. The utility of the multilevel techniques in answering the research questions clearly demonstrated the value of statistical techniques in understanding factors affecting health outcomes. The recognition of complex structures of data in statistical modelling processes, yield reliable results that help health strategists make informed decisions in taming malaria.

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND INFORMATION

Malaria is the leading cause of morbidity and mortality in Kenya, and over 70% of Kenyans are at risk of malaria infection. Nationwide, malaria is responsible for approximately 30% of outpatient visits and 19% of hospital admissions [18]. In the Lake Victoria region, malaria accounts for a reported 30–50% of out-patient care and 20% of bed occupancy rates in public hospitals. The region has high perennial transmission of malaria and deemed holoendemic, which is characterized by parasite rates consistently over 75% in infants between 0-11 months and a significant immune response among all age groups, especially adults [14] .

Malaria affects individuals who belong to some environment. This could be a household or a geographic location where the individual come from. When individual level factors are addressed, environmental factors will still contribute to persistence of malaria prevalence. In this context, this thesis models malaria prevalence with respect to individual, household and village level risk factors. To understand the determinants of Malaria logistic regression models are applied considering that the response variable is binary. The data used in the study has individual, household and village level factors. Considering the clustered data structure, this study introduces multilevel logistic regression modelling and explores it's value in understanding determinants of Malaria using a clustered data. Single-level logistic regression models which are more often used in understanding malaria determinants will be fitted. Malaria determinants will be studied using both modelling methods. A comparison will made between the two models to study the efficiency of both methods in understanding the determinants of Malaria.

Overall, the purpose of the study is: *first*, to model malaria prevalence using single-level .*Second*, modelling malaria prevalence using multilevel models. *Third*, is to make interpretation of both modelling results to understand the determinants of malaria. *Fourth*, is to make comparison of single and multilevel models to assess their efficiency in understanding the determinants of malaria.

The thesis is divided in five sections. After the introduction the *second section* Discusses literature of the previous study done with respect to current work of the thesis The *third section* describes the research methods used in the study: it presents the data collection background,

single and multilevel modelling in STATA. The *fourth section outlines* empirical results and discussions both in exploratory data analysis and modelling process. In *the fifth section* the study work done will be summed up and concluded. *Sections six and seven* have references and appendices on analysis results respectively.

1.2 MALARIA RISK FACTORS AND THE RESPONSE VARIABLE

Malaria is influenced by a number of factors as shown in various pieces of literature. Various methodologies have been used to understand these factors. These study seeks to contribute to various effort made in understanding these factors using the multilevel analysis methodologies. These factors could be characteristics of individual who is infected by malaria or characteristics of the environment where the individual belongs. Environment can be defined in various ways. A household where an individual belongs can be defined as an environment and a bigger geographical area could as well be an environment. This study seeks to understand determinants of malaria specifically considering characteristics of the individuals, household and the village the individual belongs. The factors considered are defined at three levels: individual (age and gender), household (household income status, household education status, and village (area covered by woods and water bodies' count).The presence of malaria for individual is defined as a non-zero *Plasmodium falciparum* count from examination of a blood smear. The response variable is a binary variable that takes 1 for the presence of malaria parasite and zero for the absence of malaria parasite.

1.3 STATEMENT OF THE PROBLEM

In Barsauri, where the Millennium Village was established, malaria is the most prevalence disease among the common diseases according Millennium research project documents. Malaria transmission is believed to be associated with a number of risk factors that are both at individual and environmental level. The multilevel nature of Malaria risk factors (Individuals, Households and Villages) brings modeling complexities that need to be considered in an effort to curb malaria but only single level models have been used in the previous studies. There is therefore need to use multilevel modeling technique to solve this statistical challenge.

1.4 OBJECTIVE OF THE STUDY

The following are the objectives of the study;

1. To model malaria prevalence using single-level models
2. To model malaria prevalence using multilevel models
3. To use the interpretation of the single-level and multilevel models to study the effects of individual, household and village-level risk factors to malaria prevalence
4. To compare the efficiency of single-level and multilevel models in studying the effects of risk factors to malaria prevalence

1.5 SIGNIFICANCE OF THE STUDY

To achieve better health outcomes in war against malaria, research findings are highly relied upon during public health decision-making processes. This study will highly contribute to the generation of reliable information that is helpful to public health decision makers by its virtue of exploring suitable statistical modelling methods that would effectively give an understanding of malaria risk factors. Recognizing multilevel nature of malaria risk factors adds so much value to this effort. This will in the long run influence the designs of interventions put in place to combat malaria and other health outcomes.

CHAPTER 2: LITERATURE REVIEW

In 2005, Columbia University in collaboration with Millennium Promise started Millennium Villages in 14 African countries. The origins of this villages lie in the Millennium Summit held in New York City in the year 2000, at which 147 presidents, prime ministers, and monarchs pledged to meet eight general development targets—the Millennium Development Goals (MDGs)—such as the halving of global poverty and achieving universal primary school completion by 2015.These villages were to demonstrate how the Eight Millennium Development Goals are achieved in an integrated approach. This community driven project in the villages was guided by the recommendations of the UN Millennium Project on the types of interventions in agriculture, nutrition, health, education, energy, water, communications, and the environment. In Kenya Sauri Millennium Village-the first millennium village- was established. Considering the high prevalence of Malaria in Lake Victoria region as documented in several pieces of literature, reduction of malaria was given high priority by the project.

In its effort to curb malaria the project conducted a baseline studies to get the bench marks of the prevalence as well as to identify factors influencing malaria prevalence. This section discusses previous research done on malaria prevalence and the risk factors associated with it. It looks at various statistical methodologies applied in identifying and understanding malaria risk factors, particularly multilevel models. Single level models as applied by various researchers, will also be reviewed and results compared with multilevel models on suitability of each of the models in analyzing multilevel data. Finally, the section will narrow down to share the statistical methods applied in Sauri Millennium Village to study malaria risk factors and seek to identify any lessons which could add value to this study with respect to the stated study problem and objectives.

2.1 SAURI MILLENNIUM VILLAGE

With reference to the Millennium Villages Project's Sauri baseline report [7], The Village is located in Gem district,western highlands of Kenya (34.53°E; 0.01°N), approximately 30 kilometers north of Lake Victoria. The average temperature for the region is 24°C, annually ranging from 18 – 27°C. The village receives approximately 1,800 millimeters of annual

rainfall, which is split between a longer rainy season from March to June (~720 mm) and a shorter rainy season from September to December (~630 mm).

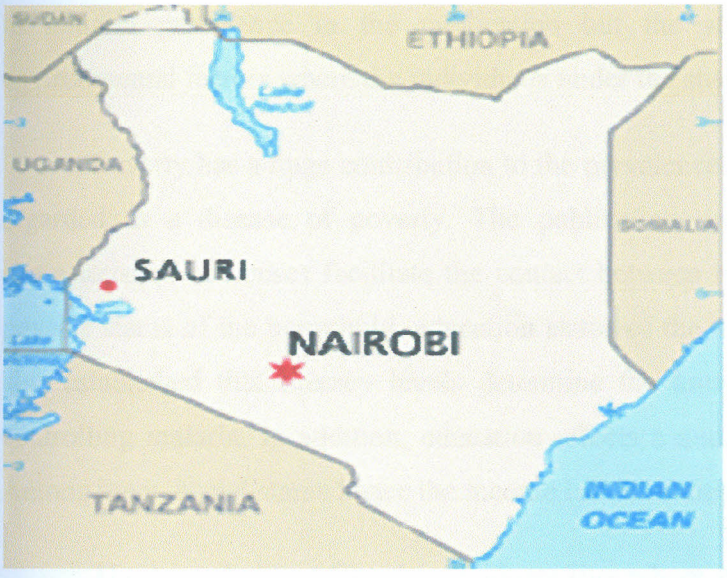

Figure 1: Sauri Millennium village (source: Millennium Promise website)

2.2 MALARIA PREVALENCE AND THE RISK FACTORS

Malaria prevalence is influenced by a number of factors defined in different levels as mentioned by [3]. These factors range from individual characteristics, household characteristics to geographic region characteristics where the individual resides. A number of studies have established the influence of various factors to malaria prevalence without recognizing the implication of the multilevel structure of data on the inferences being drawn on the determinants of malaria prevalence. It was mentioned in Lopez *et al* [9] that in the processes of gathering information for the purposes of decision making in health, it is important that the disease burden is described including the associated risk factors.

The risk factors considered in the study are at various levels. Gender and age are individual level factors, income levels and education status of households are household level factors where the individual belongs then wood area and water bodies are village level factors where the household belongs. As much previous literature have studied either of the considered factors with respect to malaria prevalence, none of the studies have investigated the influence of these factors to malaria prevalence simultaneously considering their levels in the data hierarchy.

In his publication of blood group and demographic characteristics on malaria infection Akanbi *et al* [1] considered age and gender in the study. Age and gender have been stated to have an effect to malaria prevalence in the publication but the study did not consider the effect of environmental factors where the individuals under the study belong.

Poverty has a huge contribution to the prevalence of malaria. In Sintasath [13] Malaria is regarded as a disease of poverty. The publication went further to point out that specific characteristics of houses facilitate the contact between humans and mosquitoes. In addition to poverty status of the household, education status of the household is a factor under the study. It also established that literacy levels determine the knowledge of the household members in controlling malaria. In addition, education affects a number of other factors in the household. Among them, social status hence the income levels of the household [3].

The households infected individuals live where they are surrounded by trees and other vegetation and also the water points in form of springs and rivers. The closer a household would live in a forested area the higher the risk of household members being infected with malaria [6]. Mosquitoes are known to breed in stagnant waters. This implies that households which are closer to these waters are at higher risk of getting malaria.

2.3 MULTILEVELMODELING

Multilevel modelling is an attempt to define statistical models which make sense of data with grouped or clustered observations, where specifically the pattern of clustering is known. Given the clustering in the data where individuals are clustered in households and households clustered in villages, it is important to allow for dependence or correlations among the responses observed for individuals belong to the same cluster [11].

Much research involves multilevel data structures but despite this common phenomenon in both behavioural and social sciences, past studies have often failed to address them adequately in data analysis [12]. More often, existing research has centered on the effects of individual-level risk factors while giving little attention to multilevel nature of higher levels risk factors effects as described by [9]

Raudenbush and Bryk [12] pointed out that this neglect has led to limitations that have generated concerns about aggregation bias, misestimate precision, among others. Gelman & Hill[4] also cautioned that failure to recognize the grouping structure of data in modelling leads to inaccurate inferences. This is due to the correlation within groups, and the misleading bias obtained when this is ignored.

The usefulness of multilevel modelling in analysis of the datasets that include risk factors in more than one level was pointed out by Krieger and Waterman [8].Moreover, the publication highlighted that this approach is yet to become a tool for public health monitoring and surveillance because of its complexities regardless of its reliability and use in the research. This was also supported by Mauny *et al* [10] who went further to mention that there is an increasing realization that environmental factors could have significant effects on health.

Multilevel modelling allows for several levels in the modelling process. In their teaching paper Trammer and Trammer [15] state that Multilevel modelling techniques allow us to assess variation in a dependent variable at several levels simultaneously: for example, we can assess how much a health measure like blood pressure varies between areas and how much it varies between individuals within the areas, or we can assess how much examination scores vary between schools compared with the extent of variation in examination scores for pupils within schools; similarly we could compare variations in unemployment or limiting long term illness at the individual and area levels. In the study conducted by Mauny *et al*[10], individual-level and village-level factors were addressed. The need to add the household-level as an intermediary level of the modelling process should be addressed.

Various studies have revealed the effects of factors under study to malaria prevalence. It would be of value to get more understanding of these factors as determinants of malaria prevalence by including each of the three levels and their respective factors in the modelling process. The clustering of individuals within the households and households within the villages implies that there is more correlation of the individuals from the same cluster than among individuals of different clusters. In understanding these malaria risk factors, this has to be taken to consideration for us to get statistically reliable results which will improve the accuracy of analysis results inferences [11]. Spatial analytic methods would have of been of value but they

will reduce the scope of investigation given that the individual characteristics are not taken to account [10].

CHAPTER 3: RESEARCH METHODOLOGY

3.1 STUDY SETTING

Sauri Millennium Village is comprised of 11 sub-villages with an estimated population of 967 households and 5,500 people. The figure below shows Sauri Millennium Sub-villages, distributions of homesteads and a few geographical features.

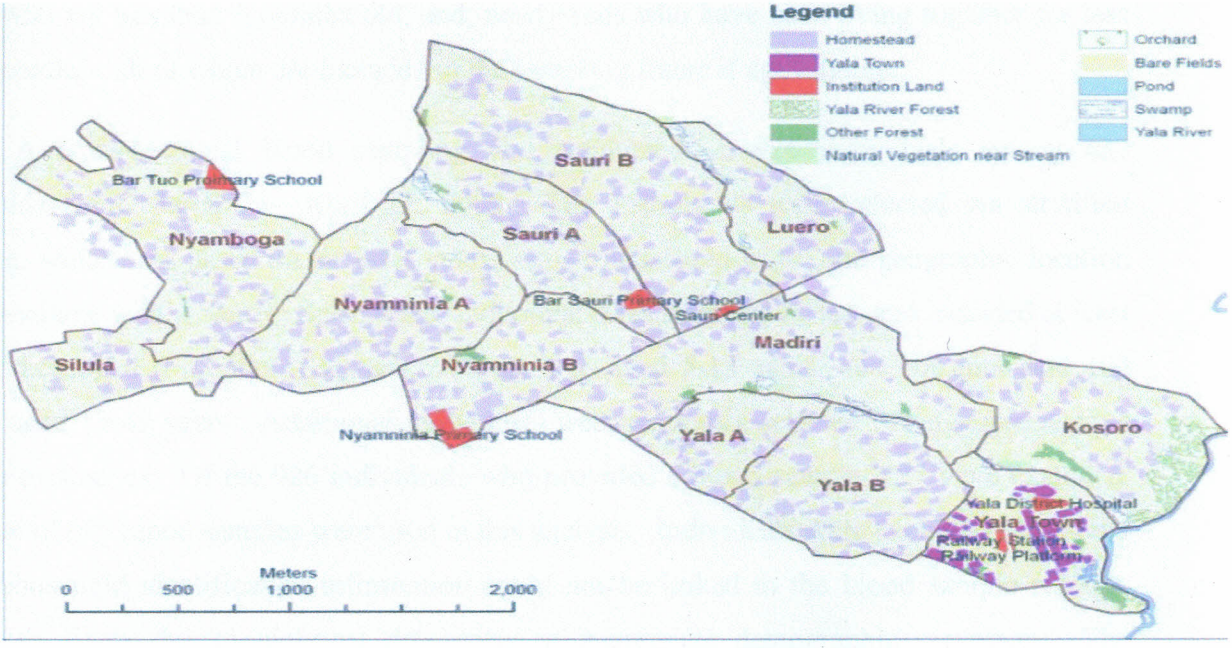

Figure 2: Map showing the location of Sauri sub-villages

3.2 DATA COLLECTION

The data set used in this analysis has been taken from Sauri Millennium Village Baseline Survey conducted in 2005.

3.2.1 HOUSEHOLD SELECTION

Within each MVP village, a detailed household mapping was conducted prior to the initiation of interventions. This process included a household and population census, Global Positioning System (GPS) readings, and the generation of a household wealth score. Following

this process, proportional sampling was used to randomly select 300 geographic and wealth-stratified households to undergo detailed periodic assessments.

3.2.2 INDIVIDUAL SELECTION

Within each participating household, individuals are recruited for study inclusion based on the results of preliminary demographic assessment. Household members are defined as those who have lived in the household for at least 3 of the past 12 months, and who 'normally eat from the same pot.' The only exceptions are for persons who are the main provider for the household; infants who are less than 3 months old, and; newlyweds who have been living together for less than 3 months – all of whom are included in the sampling frame if age eligible.

A cross-sectional blood sampling was conducted on 926 individuals among 432 households from March 7 – April 22, 2005. The households were selected via stratified sampling, which was based on an asset index, sex of household head, and geographic location (i.e. sub-village within Sauri). Selection of individuals was stratified by age and included at least 200 children aged 6 months – 5 years, 100 children aged 5-13 years, and 100 men and 100 women aged 13-49 years. Additional individuals were randomly selected to safeguard against possible exclusions. Of the 926 individuals who provided a blood sample, 227 were excluded. So a total of 699 blood samples were used in this analysis. Individuals were excluded if personal and/or household identification information could not be linked to the blood sample records, which was likely due to incorrect data entry or incomplete demographic surveying. The distribution of excluded individuals is assumed to be a random occurrence. Informed consent was obtained from each participant after a trained nurse or community health worker explained the purpose of the study as well as the procedures, risks and benefits to participants.

The demographic survey collected basic resident information, such as age, sex, literacy, education, and orphan status. Household rates of education, literacy, and orphan burden were summations from the individual demographic surveys. The socioeconomic survey conducted on each household generated values for income per capita, which is a composition of household-level agricultural, non-agricultural, and remittance income less related expenses, such as hired labor, rent, and cost of inputs, divided by the number of household members. Income is expressed as dollars per day based on 1993 purchasing power parity (PPP).

3.2.3 RISK FACTORS AND RESPONSE VARIABLE

The factors considered are defined at three levels: individual (age and gender), household (household income status, household education status, and village (area covered by woods and water bodies' count).The presence of malaria for individual is defined as a non-zero *Plasmodium falciparum* count from examination of a blood smear. The response variable is a binary variable that takes 1 for the presence of malaria parasite and zero for the absence of malaria parasite.

3.3 THE MODELING PROCESS

3.3.1 MODELLING BINARY OUTCOME

The standard way of modelling binary outcomes is by using the logistic regression. Given that the response variable takes responses 1 or 0, the transformation is done so that a model is derived that will observe the key assumptions of a regression model. The expectation of a binary response (0 or 1) is the probability that the response is 1 (Rambe-Hesketh & Skrondal, 2008).

Consider the variable representation of a binary response model.

$$y_i^* = \beta_0 + \beta_1 x_i + \epsilon_i \qquad (1)$$

Where y_i^* is a continuous variable underlying the observed binary response y_i(of having malaria or not having malaria parasite in blood) for individual i, x_i is a malaria risk factor for individual i, ϵ_i is residual with mean zero and variance σ_ϵ^2. β_0 Is a constant and β_1 is coefficient of the malaria risk factor.

Given that y_i^* is unobserved this implies that:

$$y_i = \begin{cases} 1 \; if \; y_i^* \geq 0 \\ 0 \; if \; y_i^* < 0 \end{cases}$$

Based on the above equation the probability that y_i=1 can be denoted as π_i and the probability that y_i=0 can be denoted as $1 - \pi_i$. this implies that:

$$\pi_i = \Pr(y_i^* \geq 0) = \Pr(\beta_0 + \beta_1 x_i + \epsilon_i \geq 0)$$

In the model, x_i is a fixed variable whereas ϵ_i is a random variable. This implies that:

$$\pi_i = \Pr(\epsilon_i \geq -(\beta_0 + \beta_1 x_i))$$

If ε_i has a symmetric distribution the above equation can be written as:

$\pi_i = \Pr(\epsilon_i \leq (\beta_0 + \beta_1 x_i))$Which is cumulative distribution function (CDF) evaluated at $\beta_0 + \beta_1 x_i$ the exact form of which will depend on distribution assumed by ϵ_i.

If $\epsilon_i$is assumed to follow a standard logistic distribution, then:

$\Pr(\epsilon_i \leq \beta_0 + \beta_1 x_i) = \mathrm{F}(\beta_0 + \beta_1 x_i)$ Where $\mathrm{F}(\beta_0 + \beta_1 x_i)$ is the CDF of standard logistic distribution which is also know us the logit link function. For the logit link, the model is written as

$$\Pr(y_i = 1/x_i) = \mathrm{F}(\beta_0 + \beta_1 x_i) = \frac{exp(\beta_0+\beta_1 x_i)}{\{1+\exp(\beta_0+\beta_1 x_i)\}} \quad (2)$$

This could be written as

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i \quad (3)$$

The logit link is more appealing because it produces a linear model for the log of the odds [11].

3.3.2 SINGLE LEVEL MODELLING MALARIA PREVALENCE

In single modelling level it is assumed that the observations are not clustered and that the residual variance does not vary with observations of the explanatory variables. Considering this assumption, the single level logistic regress model will is written as

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 \text{gender}_i + \beta_2 \text{age5}_i + \beta_3 \text{age17}_i + \beta_4 \text{age60}_i \quad (4)$$
$$+ \beta_5 \text{age\_above60}_i + \beta_6 \text{income}_i + \beta_7 \text{education}_i + \beta_8 \text{income} * \text{education}_i + \beta_9 \text{woodarea}_i$$
$$+ \beta_{10} \text{waterbodiescount}_i$$

Considering the assumption of non-clustering in single level logistic regression the residual is assumed to have a logistic cumulative density function with mean zero and variance ($\pi^2/3$ =3.29).

3.3.3 MULTILEVEL MODELLING

A three-level logit model is specified for the malaria prevalence with individual i nested in household j which are nested in village k.

The Model:

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \beta_{000} + \beta_1 \text{gender}_{ijk} + \beta_2 \text{age5}_{ijk} + \beta_3 \text{age17}_{ijk} + \beta_4 \text{age60}_{ijk}$$
$$+ \beta_5 \text{age\_above60}_{ijk} + \beta_6 \text{income}_{jk} + \beta_7 \text{education}_{jk} + \beta_8 \text{income} * \text{education}_{jk} + \beta_9 \text{wooda}$$
$$+ \beta_{10} \text{waterbodiescount}_k + \zeta_{jk} + \zeta_k$$

$\zeta_{jk} \sim N(0, \sigma_h^2)$ is a random intercept varying over the households(Level 2), and $\zeta_k \sim N(0, \sigma_v^2)$ is random intercept varying over villages(Level 3).

3.3.4 Intraclass Correlation

Intraclass correlations for different levels are as follows. The village k level intraclass correlations is given as $\rho(\text{village}) = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_h^2 + 3.29}$

Intraclass correlation for the same household j and the same village k is given as

$$\rho(\text{household, village}) = \frac{\sigma_v^2 + \sigma_h^2}{\sigma_v^2 + \sigma_h^2 + 3.29}$$

3.4 FITTING BINARY OUTCOMES IN STATA

There is a number of statistical software which could have been used during data analysis. In this study STATA version 10 was used to analyze data. This is due to the availability of the software and the fact that it is credible software that would give analysis results as desired by the thesis work.

3.4.1 ADAPTIVE QUARDRATURE PARAMETER ESTIMATION METHOD

The marginal likelihood is the joint probability of all observed responses given the observed covariates. This can be easily evaluated and maximized in linear mixed models unlike in generalized linear mixed models. This is because in generalized linear models, the marginal likelihood does not have a closed form and must be evaluated by approximate methods hence the usefulness of the adaptive quardrature [11]. During the modelling process, adaptive quardrature was used to estimate parameters for each of the models.

3.4.2 SINGLE-LEVEL LOGISTIC REGRESSION

`logit` syntax was used to estimate single-level logistic regression models. In `logit` syntax, clustering of variables for different levels in the model is not considered. The syntax follows the regression assumption of having a constant variance of residuals for each risk factor observation.

```
logit response var risk factor1 risk factor2 riskfactor3....
```

3.4.3 MULTILEVEL LOGISTIC REGRESSION MODELS

`Xtmelogit` syntax was used to estimate logistic regression multilevel models. In `xtmelogit`, the clustering variables for different levels in the model are given starting from the top level and then going down the levels. In this study the highest level is the village followed by the household.

```
xtmelogit response var risk factor1 risk factor2.... || village: || hhid:, intpoints(1)
```

Where `intpoints(1)` refers to the integration points during the parameter estimation process.

CHAPTER 4: ANALYSIS RESULTS AND DISCUSSIONS

4.1 EXPLORATORY ANALYSIS AND DATA DESCRIPTION

4.1.1 VARIABLE DESCRIPTION

Variable	Variable Label	Variable values and labels	Comment
malpreval	presence of malaria parasite in the blood of the respondent	1=parasite present 0=no parasite	
age of the respondent in years			
age5	1st knot of the cubic spline		age included in the model using the cubic spline with four knots-see Table 8
age17	2nd knot of the cubic spline		
age60	3rd knot of the cubic spline		
age_above60	4th knot of the cubic spline		
age5yrs	age group for under 5s	1=all under 5s, 0=other	age included in the model in form of age groups to make a comparison of estimated odds of getting malaria among different age groups-see Table 9
age5_17yrs	age group of those between 5 and 17 years	1= between 5 and 17 years, 0=other	
age17_60yrs	age group of those between 17 and 60 years	1=between 17 and 60 years, 0=other	
age60yrs	age group of those 60 years and above	1=all 60 years and above, 0=other	
gender	gender of the respondent	1=females, 0=males	
income1day	income status of the household in terms of a dollar day	1=Below a dollar, 0=above a dollar	
hhseceduc	education status of the household	1=Household having no one with atleast a secondary level education, 0=Household having atleast one person with secondary level education	
incomeXeduc	interaction of income and education		
woodarea	area covered by woodlot		
wterbodiescount	number of ponds and springs in the area		

4.1.2 MALARIA AND GENDER

Table 1: Malaria prevalence among males and females in the population

Gender	Malaria Prevalence	SE	CI	n
Male	144(53%)	0.03	47%,58%	272
Female	205(48%)	0.02	44%,53%	427
difference	5%	0.04	-3%,12%	
Total	350(50%)			699

Table 1 shows the malaria prevalence among males and females. 50% of the respondents have malaria parasite present in their blood. Among the males the prevalence is 53% and among the females it is 48%. This does not seem different. The confidence intervals of the difference between males and females malaria prevalence indicates that there is no significance difference between the two as it contains 0. This implies that gender does not have any influence on malaria i.e. whether one is a female or male in area one still has approximately the same chance of getting malaria. Figure 1 below checks malaria prevalence in each of the age groups for both males and females.

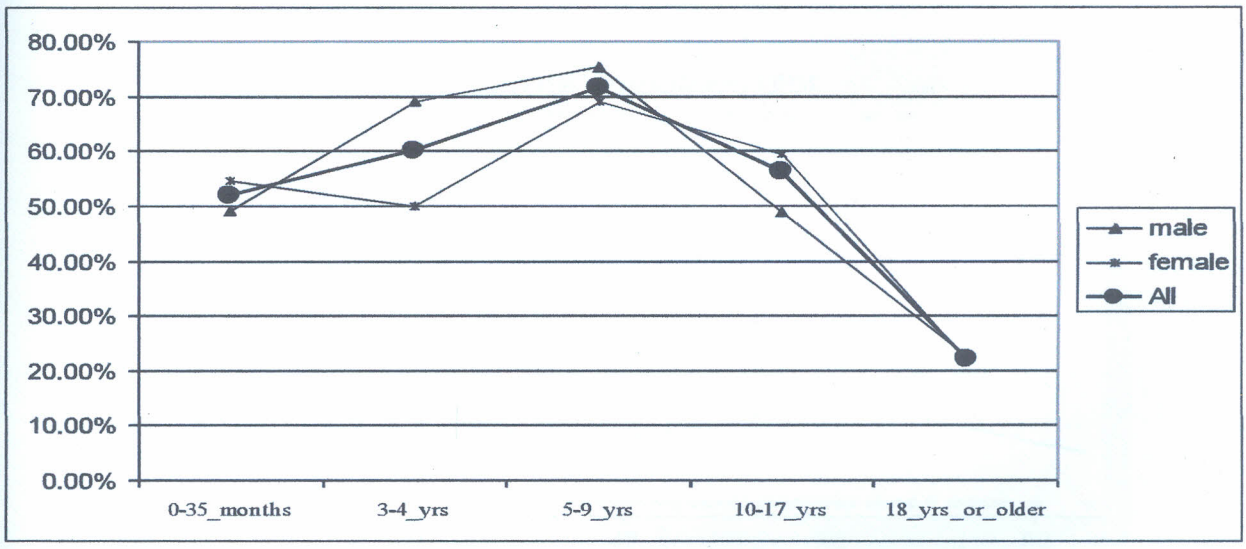

Figure 3: Malaria prevalence among different age groups and gender

In each of the age groups there does not seem to be any much difference in malaria prevalence between males and females except for ages 3-4 years. The results still reflect the results of the overall prevalence between males and females shown in table 1. Further tests during modelling process will verify if indeed gender has influence on malaria as all the risk factors will be included in the models unlike at this stage where age and gender are the only considered living other risk factors which could as well explain variation of malaria prevalence. At this exploratory stage, figure 3 reveals that age has a non-linear relationship with malaria and that younger ages have more malaria compared to older ages. This revelation will be considered during the modeling process.

4.1.3 MALARIA AND AGE

Figure 1 also shows malaria prevalence is high at younger ages compared to when one is 18 years and above. At younger ages the prevalence range from 52% to 71.6 % where as it is only 22% at ages 18 years and above. There seem to be quite some variation of the malaria prevalence between different age groups. To get a clear picture on the relationship of malaria prevalence and age, a cubic spline with four knots was fitted and the predicted prevalence was plotted versus the age. A cubic spline was necessary considering figure 1 that showed that the relationship of age and malaria prevalence is not linear

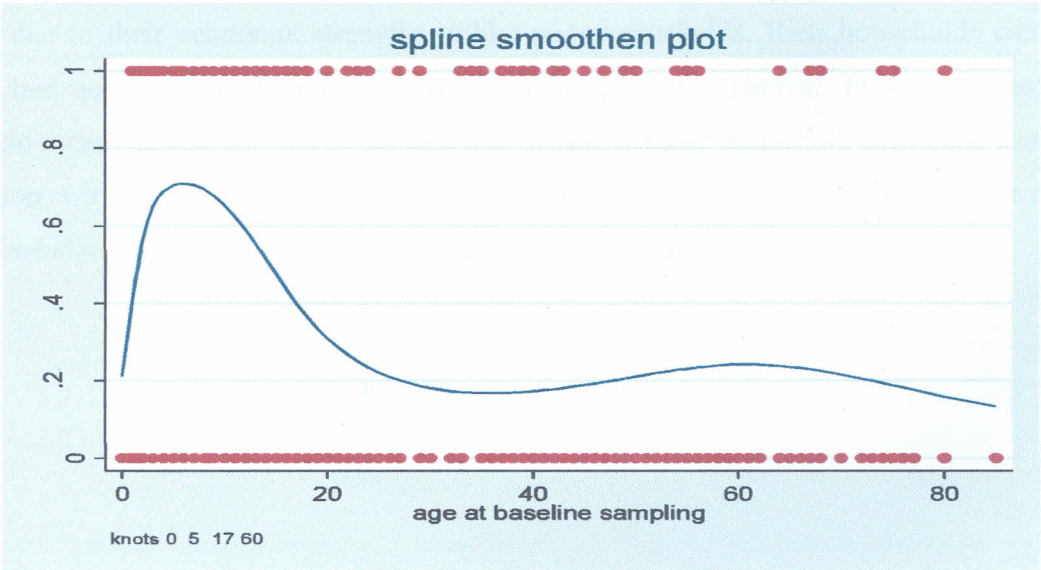

Figure 4: Estimated malaria prevalence with respect to age

The cubic spline plot in figure 4 shows a clear effect of age on malaria prevalence. Malaria prevalence goes up as one grows from 0 to 10 years then comes down at later ages.

4.1.4 Malaria and Wealth

Table 2: Among prevalence among those below and above a dollar a day

Poverty status	Malaria prevalence	SE	CI	n
Above poverty line	64(40%)	0.04	32%,47%	161
Below poverty line	284(53%)	0.02	49%,57%	538
difference	-13%	0.04	-22%,-4%	
				699

Table 2 shows individuals from households which are below the poverty line having higher malaria prevalence of 53% compared to those from households above poverty line with malaria prevalence of 40%. The difference of the malaria prevalence between the poor and the wealthy is significantly large. The significance of the difference is clearly given by the confidence interval of difference which does not contain 0. This reveals that individuals living in poor households are more susceptible to malaria compared to individuals living in rich households. There is a possibility that households which are rich have means of preventing malaria due to their economic strengths unlike poor households. Rich households can afford malaria bed nets that would prevent transmission of malaria parasite by mosquitoes to the household members and also nutritive food that would enhance household members' immunity. In addition, with adequate financial resources one would access better education where one can gain knowledge and skills on the best practices to control malaria in their households.

4.1.5 MALARIA AND EDUCATION

Table 3: Malaria prevalence among individuals in households with and without a member with secondary education

Education status	Malaria prevalence	SE	CI	n
No household member with secondary education	214(54%)	0.025	49%,59%	395
Household member has secondary education	134(44%)	0.028	38%,49%	304
difference	10.1%	0.038	3%,18%	
				699

As shown in table 3, household members from households with no member who has a secondary level education have more malaria prevalence of 54% compared to those with a member who has secondary level education with 44%. This reveals members from less educated household are more susceptible to malaria than the more educated households.

4.1.6 HOUSEHOLD INCOME AND EDUCATION INTERACTION

Considering the results in Table 2 and 3, Could there be any interaction between income and education status with respect to malaria infection?

Table 4: Income and education interaction with respect to Malaria prevalence

	Poverty status		
Education status	Below poverty line	Above poverty line	Total
No household member with secondary education	178(55%)	36(49%)	214(54%)
Household member has secondary education	106(49%)	28(32%)	134(44%)

In Table 4 Looking at the individuals belonging to households below poverty line, Malaria is lesser among those from households with at least one member of household having secondary education compared(49%) to those who have none(55%). The difference is even much higher for individuals belong to households above poverty line (32% to 49%). The Table clearly

shows that lesser poverty and higher education contributes reduced malaria (32%) compared to lower education and higher poverty (55%).This implies that there is some interaction between education and poverty. The influence diagram below further illustrates this relationship.

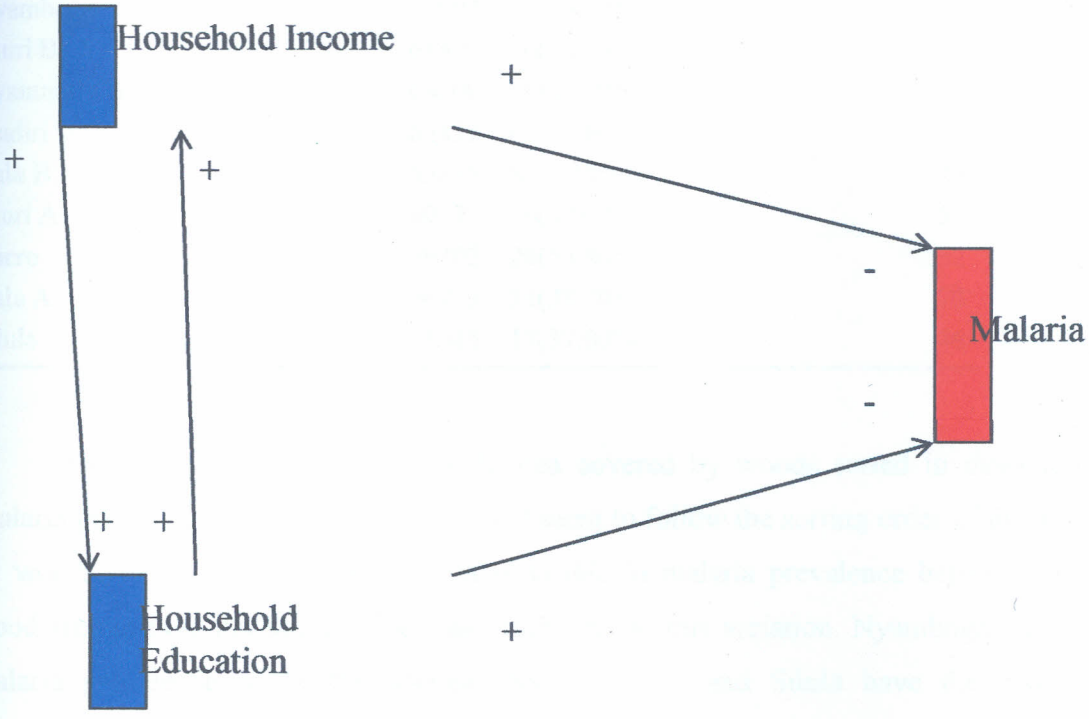

Figure 5: Household education and income malaria influence diagram

Figure 3 depicts the relation between household education, household income and malaria. The figure shows both income and education have influence to malaria. Higher education in the households (+), leads to less malaria among the population (-). On the other hand, higher household incomes (+) lead to less malaria (-) among the population. Education is also related to income. Higher income contributes to better education since the household can afford to educate its members. Also higher education leads household members to get better jobs or doing better businesses because of knowledge and skills earned through education that will give them better incomes.

4.1.7 MALARIA AND THE AREA COVERED BY WOODS

Table 5: Malaria prevalence among individuals from different villages with the wood areas

Village name	Area covered by woods(M^2)	Malaria prevalence	Respondent count
Nyamninia A	84857	37(35.60%)	104
Kosoro	80879	29(50.00%)	58
Nyamboga	74692	45(66.20%)	68
Sauri B	68470	36(56.30%)	64
Nyamninia B	64800	25(47.20%)	53
Madiri	62852	62(59.60%)	104
Yala B	53645	8(53.30%)	15
Sauri A	40223	30(52.60%)	57
Luero	29302	28(53.80%)	52
Yala A	26713	31(39.70%)	78
Silula	25645	17(37.00%)	46

Table 5 shows the villages with area covered by woods sorted in descending order. Malaria prevalence for each village does not seem to follow the sorting order of the area covered by woods. As much as some variation is visible in malaria prevalence between the villages, wood area does to not seem to have any influence to this variation. Nyamboga has the highest malaria prevalence of 66.2% whereas Nyamninia A and Silula have the lowest malaria prevalence of 37%. Nyamninia A with an area of 84856.9 meters square covered by woods has the least malaria prevalence compared to villages with smaller areas covered by woods.

4.1.8 MALARIA AND THE WATER BODIES COUNT

Table 6: Malaria prevalence among individuals from different villages with the wood areas

Village name	Water bodies count	Malaria prevalence	Respondent count
Luero	5	28(53.80%)	52
Yala A	4	31(39.70%)	78
Silula	3	17(37.00%)	46
Nyamninia A	3	37(35.60%)	104
Nyamboga	3	45(66.20%)	68
Madiri	3	62(59.60%)	104
Nyamninia B	2	25(47.20%)	53
Kosoro	2	29(50.00%)	58
Sauri A	2	30(52.60%)	57
Sauri B	1	36(56.30%)	64
Yala B	1	8(53.30%)	15

Table 6 shows the water bodies count sorted in descending order. Just like Table 5, Malaria prevalence does not follow this order clearly showing that the water bodies count has no any influence to the prevalence of malaria in the villages.

4.2 THE MODELLING PROCESS

In this section, single and multilevel models are fitted. The modelling process begins with a null model then risk factors are added one at a time. The modelling section will just give the final fitted models as appendices will clearly show the step by step process of risk factors inclusion. In addition, odds ratios are computed to get more understanding of malaria risk factors.

4.2.1 SINGLE LEVEL MODELLING

Appendix 1 shows step-by-step addition of risk factors as the single-level model is fitted. Considering the non-linear relationship between age and malaria prevalence as indicated in figure 1 and 2, a cubic spline with four knots was used to fit the age in the model.

The following are the results of the final fitted model.

Table 7: Single-Level Logistic regression model results

```
logit malpreval gender    age5   age17  age60   age_above60 income1day hhseceduc
incomeXeduc woodarea wterbodiescount

Logistic regression                               Number of obs   =        699
                                                  LR chi2(10)     =     128.52
                                                  Prob > chi2     =     0.0000
Log likelihood =  -420.2428                       Pseudo R2       =     0.1326

------------------------------------------------------------------------------
   malpreval |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |   -.184517   .1781686    -1.04   0.300    -.5337211    .1646871
        age5 |    .428296   .0797406     5.37   0.000     .2720073    .5845848
       age17 |  -.1599404   .0266254    -6.01   0.000    -.2121253   -.1077555
       age60 |  -.0260685   .0105082    -2.48   0.013    -.0466642   -.0054727
 age_above60 |   .0453821   .0448453     1.01   0.312    -.0425131    .1332772
  incomeiday |   .7336893    .289322     2.54   0.011     .1666286     1.30075
   hhseceduc |   .7866672   .3596812     2.19   0.029      .081705    1.491629
 incomeXeduc |  -.5465518   .4055789    -1.35   0.178    -1.341472    .2483683
    woodarea |   2.06e-06   4.41e-06     0.47   0.641    -6.59e-06    .0000107
wterbodies~t |  -.0598767   .0916662    -0.65   0.514    -.2395391    .1197857
       _cons |  -1.431728   .6003195    -2.38   0.017    -2.608333   -.2551239
------------------------------------------------------------------------------
```

The fitted single-level model:

$$\log\left(\frac{\Pi_i}{1-\Pi_i}\right) = -1.43 - 0.18\text{gender}_i + 0.43\text{age5}_i - 0.16\text{age17}_i - 0.03\beta 4\text{age60}i$$

$$+ 0.05\text{age\_above60}_i + 0.736\text{income}_i + 0.79\text{education}_i - 0.55\text{income} * \text{education}_i + 2.06e-06\ \text{woodarea}_i$$

$$- 0.06\text{waterbodiescount}_i$$

Table 7 shows that age have significant effect on Malaria prevalence except for ages 60 years and above. Income and education also seem to have a significant effect on malaria prevalence. Gender, wood area cover and water bodies count do not have a significant effect on Malaria. It is assumed that clustering between observations from households and villages does not exist hence variation between households and between villages is not accounted for. This has implications on estimated parameters standard errors computation which key in understanding the determinants of malaria. In addition, a research like to what extent does Malaria Prevalence vary among households or villages cannot be answered by the model. The next section of modeling process where a logistic multilevel model is fitted will address some of these concerns.

4.2.2 MULTILEVEL MODELLING

Appendix 2 shows step-by-step addition of risk factors as multilevel model is fitted. Considering the non-linear relationship between age and malaria prevalence as indicated in figure 1 and 2, a cubic spline with four knots was used to fit the age in the model. The following are the results of the final fitted model. Unlike in previous Single-level Model, the fitted multilevel models accounts for between villages and between households variation.

Table 8: Multilevel Logistic regression model results

```
Mixed-effects logistic regression                Number of obs      =       699

------------------------------------------------------------------------------
                |   No. of       Observations per Group       Integration
 Group Variable |   Groups    Minimum    Average    Maximum      Points
----------------+-------------------------------------------------------------
        village |       11         15       63.5        104           7
           hhid |      338          1        2.1          8           7
------------------------------------------------------------------------------

                                                 Wald chi2(10)      =     83.98
Log likelihood = -413.83507                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
   malpreval |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |  -.2046431   .1942729    -1.05   0.292    -.5854111    .1761248
        age5 |   .4782208   .0923859     5.18   0.000     .2971478    .6592938
       age17 |  -.1746259   .0301261    -5.80   0.000    -.2336719   -.1155798
       age60 |  -.0285253   .0114033    -2.50   0.012    -.0508755   -.0061752
 age_above60 |   .0496992   .0486965     1.02   0.307    -.0457441    .1451426
  incomelday |    .835472   .3328741     2.51   0.012     .1830507    1.487893
   hhseceduc |    .954531   .4190899     2.28   0.023     .1331299    1.775932
 incomeXeduc |  -.5812646   .4668491    -1.25   0.213    -1.496272    .3337428
    woodarea |  4.50e-06   8.22e-06     0.55   0.584    -.0000116    .0000206
wterbodies~t |  -.0273539   .1599311    -0.17   0.864     -.340813    .2861052
       _cons |  -1.927023   .9211515    -2.09   0.036    -3.732446   -.1215991
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.      [95% Conf. Interval]
-----------------------------+------------------------------------------------
village: Identity            |
                  var(_cons) |   .1491694    .104356      .0378613     .587711
-----------------------------+------------------------------------------------
hhid: Identity               |
                  var(_cons) |    .345217   .2966682      .0640615    1.860317
------------------------------------------------------------------------------
LR test vs. logistic regression:     chi2(2) =    12.82   Prob > chi2 = 0.0016
```

The fitted multilevel model:

$$\log\left(\frac{\Pi_{ijk}}{1-\Pi_{ijk}}\right) = -1.92 - 0.25\text{gender}_{ijk} + 0.48\text{age5}_{ijk} - 0.18\text{age17}_{ijk} - 0.03\text{age60}_{ijk}$$

$$+ 0.05\text{age\_above60}_{ijk} + 0.84\text{income}_{jk} + 0.96\text{education}_{jk} - 0.58\text{income} * \text{education}_{jk} + 4.50e-06\text{wooda}$$

$$- 0.03\text{waterbodiescount}_{k}$$

As shown in appendix 2 there are no relatively major changes in the village level variance as cofactors are introduced compared to household level variance. Table 8 (With the final inclusion of all the risk factors) shows age, income and education have significant effects to malaria prevalence. Gender, wood area cover and water bodies' count don't seem to have any significant effect on malaria prevalence.

The LR test Versus logistic regression shows that the multilevel logistic regression model is a better fit compared to Single-level logistic regression (p-value=0.0016).

4.2.3 ESTIMATED INTRACLASS CORRELATION

In table 8, between household variance (σ_h^2) is estimated to be 0.34 and between villages variance (σ_v^2) is 0.15.

The Intraclass correlation coefficient for the village k level is given as

ρ(village) =0.15/(0.15+0.34+π^2/3) which is estimated as 4%. This implies that 4% of residual variance in having malaria is attributed to unobserved village characteristics.

Intraclass correlation for the same household j and the same village k is given as

ρ(household, village) =(0.15+0.34)/(0.15+0.34+π^2/3). This is estimated to be 13%. This implies that 13% of residual variance in having malaria is attributed to unobserved household and village characteristics.

The intraclass correlations coefficient shows the usefulness of multilevel models in accounting for village and household level variations.

4.2.4 FURTHER UNDERSTANDING OF MALARIA RISK FACTOR USING ODDS RATIOS

Appendix 3 gives detail results of odds ratios computation. Table 9 shows that a child who is 5 years and below has 6 times odds of getting malaria compared to an individual who is 60 years and above. The same applies to individual who is age between 5-17 years. An individual who is aged between 17-60 years has 1 times odds of getting malaria compared to individual who is aged 60 years and above which is significantly smaller compared to younger ages.

Table 9: Odds ratios in the fitted logistic multilevel model

```
Mixed-effects logistic regression               Number of obs      =       699

------------------------------------------------------------------------------
                |   No. of       Observations per Group       Integration
 Group Variable |   Groups    Minimum    Average    Maximum      Points
----------------+-------------------------------------------------------------
village |       11          15       63.5         104           7
hhid    |      338           1        2.1           8           7
------------------------------------------------------------------------------

                                                Wald chi2(9)       =     71.91
Log likelihood = -428.85055                     Prob> chi2         =    0.0000

------------------------------------------------------------------------------
malpreval | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
gender       |   .7841968   .1475457    -1.29   0.196      .5423411    1.133907
     age5yrs |   5.726957   2.978726     3.36   0.001      2.066302    15.87282
  age5_17yrs |   6.805599   3.622574     3.60   0.000      2.397605    19.31769
 age17_60yrs |   1.083766   .5876544     0.15   0.882      .3744424    3.136792
  income1day |   2.028571   .6504782     2.21   0.027      1.082045    3.803077
hhseceduc    |   2.337456   .9438929     2.10   0.035      1.059298    5.157852
incomeXeduc  |   .6453886   .2906789    -0.97   0.331      .2669596     1.56026
woodarea     |   1.000003   7.77e-06     0.40   0.688      .9999879    1.000018
wterbodies~t |   .9960441   .1510212    -0.03   0.979      .7399776    1.340721
------------------------------------------------------------------------------
```

Females have approximately equal odds of getting malaria with their male counterparts (0.78). Showing that whichever gender one belongs does not matter for one to be infected with Malaria parasite. Considering the interaction term "income*educ" for income and education, there is need to be very careful when interpreting any of the terms involved in the interaction [17] (income and education). For example, in the above model "income1day" cannot be interpreted as the overall comparison of household below poverty line to households above poverty line , because this term is part of an interaction. It is the effect of poverty when the

"other" terms in the interaction term are at the reference values (i.e. hhseceduc = 0). Similarly, the "hhseceduc" cannot be interpreted as the overall comparison of "less educated" to "more educated households". It is the effect of "hhseceduc" when "other" terms in the interaction term is at the reference value (i.e. Income1day = 0).

`Lincom and adjust` stata commands were used to assess the interaction between income and education after model the in table 8 was fitted. The results are shown in Table 10 and 11.

Table 10: The odds computed from the adjust command

income less than $1 per day	secondary educ level memmber present: sec level	no sec level
no	.358695	.967689
yes	.88951	1.31545

Table 10 clearly indicates increase in odds of getting malaria when an individual is from a poorly educated households (.36 to .96) and (.89 to 1.32) in both poor and rich households respectively. The same trend is shown when looking at increase in odds of getting malaria by individuals from rich to poor households both in well and poorly educated households (.36 to .89) and (.97 to 1.31) respectively. This reveals a strong interaction between education and income. When both poverty and less education are in place the odds is the highest (1.32).

Table 11: Assessing the income effects

```
. xi:lincom income1day + 0* incomeXeduc, or

( 1)  [eq1]income1day = 0

------------------------------------------------------------------------------
malpreval | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   2.028571   .6504782     2.21   0.027     1.082045    3.803077
------------------------------------------------------------------------------
```

Table 12-assessing the education effects

```
. xi:lincomhhseceduc + 0*incomeXeduc, or

( 1)  [eq1]hhseceduc = 0

------------------------------------------------------------------------------
malpreval | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
        (1) |   2.337456   .9438929     2.10   0.035     1.059298    5.157852
------------------------------------------------------------------------------
```

Table 11 show an individual from a poor household is 2 times likely to get malaria even if the individual belongs to a well-educated household revealing the effect of poverty and almost equally in table 12, an individual from a poorly educated household is 2 times likely to get malaria even if an individual is a from a rich household showing the effect of education as much some is from a rich household. These further reveal the interaction between income and education confirming illustrations of the influence diagram in figure 3.

4.2.5 COMPARISION OF SINGLE AND MULTILEVEL MODELS

This section compares the Single-level logistic regression and multilevel logistic regression models as applied in modelling malaria prevalence.

4.2.5.1 MALARIA DETERMINANTS

Both models showed that age, income and education have a significant effect on Malaria as given in Table 7 and 8. However, multilevel models provided a more accurate and comprehensive description of relationships in the clustered malaria data as compared the single conventional logistics models by; correcting underestimated standard errors, by estimating components of variance at several levels, and by estimating cluster-specific intercepts. Investigation of between households and between villages level effects variation

Single-level modelling assumed that there is no clustering of observations. As a result, it was not possible to explain the extent to which malaria prevalence varied between households and between villages. Multilevel modelling has allowed for the assessment of variation of malaria prevalence between households and between villages. Between household variance is estimated to be 0.34 and between villages variance is 0.15 as shown in Table 8 which are led to computation of 13% and 4% of intraclass correlation coefficients.

4.2.5.2 CORRECT INFERENCES

Table 13: Coefficients and standards errors for Single and Multilevel Logistic regression

	Multilevel Model	Single-level Model
Parameter	coefficients(standard errors)	coefficients(standard errors)
Fixed		
Intercept	-1.927023(.9211515)	-1.431728(.6003195)
Individual		
age5	.4782208(.0923859)	.428296(.0797406)
age17	-.1746259(.0301261)	-.1599404 (.0266254)
age60	-.0285253(.0114033)	-.0260685(.0105082)
age_above60	.0496992(.0486965)	.0453821(.0448453)
Gender	-.2046431(.1942729)	-.184517(.1781686)
Household		
Income	.835472(.3328741)	.7336893(.289322)
Education	.954531(.4190899)	.7866672(.3596812)
Income*Education	-.4379026(.4503936)	-.5465518(.4055789)
Village		
Woodarea	4.50e-06(8.22e-06)	2.06e-06(4.41e-06)
Water	-.0273539(.1599311)	-.0598767(.0916662)
Random		
village	.1491694(.104356)	
Household	.345217 (.2966682)	

The interpretation of coefficients for Single-Level and Multilevel Model differ. In Single-level Model coefficients have population averaged interpretation whereas in Multilevel coefficients have cluster specific interpretation. This has implications on inferences made in determining which risk factors indeed have a significant effect to Malaria prevalence. In Table 13, the coefficients and standard errors of risk factors in Multilevel Model are larger compared to coefficients in Single-Level Model. In the multilevel model for instance, income, education, and income-education interaction coefficients standard errors are 0.333,0.419 and 0.450 respectively. In the Single-level model income, education, and income-education interaction coefficients standard errors are 0.289, 0.360 and 0.406 respectively. Considering all observations to be independent, single-level modelling assumes more Information in the data than there actually is. Consequently, standard errors based on an independence assumption are underestimated hence in accurate considering that the assumption of independent residuals is invalid since observations from the same cluster are correlated. In addition, multilevel modelling allows one to have correct estimates of higher level effects for an unbalance in the lower level effects. The data used

in the study is unbalanced application of multilevel modelling As shown in table 14, estimates of the odds changed when multilevel model was fitted.

Table 14: Household level factor interaction odds estimates

	Single-Level model		Multilevel model	
	sec level	no sec level	sec level	no sec level
Rich	0.41	0.94	0.36	0.97
poor	0.93	1.23	0.89	1.32

In drawing the inference on the extent of household factors interaction, the multilevel estimates are much reliable considering that the multilevel Interpretation is Household clustering based and not the entire sample based.

CHAPTER 5: SUMMARY, CONCLUSION AND RECOMMENDATIONS

There are so many statistical approaches used in modelling malaria prevalence. The application of multilevel analysis of binary outcome in the study did generate some helpful inferences on determinants of malaria in a clustered data. The outcome of the study did reveal the value of Multilevel Models in understanding Malaria as compared to single-level modelling. This Chapter summarizes the thesis work drawing conclusion from the study findings and making recommendations based on the study outcome. The study is also not exempted from a few limitations which will be mentioned latter.

The findings in the study aided by multilevel analysis clearly strengthened the necessity of achieving MDGS by an integrated approach specifically in combating malaria, where in addition to health interventions, improvement of households' levels of education and income status is being initiated by the MVP project.

Among the six malaria determinants studied, the study revealed that age, education and income status clearly have influence in malaria prevalence among the residents of Sauri. The study reveals younger ages are more endangered by malaria compared to the older ages. It is therefore indeed worthwhile that Millennium Projects pays adequate attention to younger children as it seeks to achieve its goal of reducing malaria prevalence. In addition, education and income increment interventions are equally essential in controlling malaria prevalence. Having households with good income status and well educated individuals will lead to significant reduction of malaria prevalence among the household members in the future. Gender, water bodies count, and wood area in the villages did not show any significant influence to malaria prevalence.

One quick limitation that the study would point is that the geospatial clustering was not applied. This would have definitely added more value to the study findings. Further exploration of malaria prevalence variation among households and villages would be a great opportunity for further study.

The methodological strength of multilevel modelling in analysis of multilevel data came in handy when the study sought to understand the household level and village level influence on

malaria prevalence. Household level malaria determinants (income and education), were reliably understood by the application of multilevel modelling. In the same line, the design of Millennium Villages Project is such that an integrated approach is used to improve the livelihoods of the populations. In this approach, key interventions are put in place to achieve the Millennium Development Goals (MDGs). The findings in the study aided by multilevel analysis clearly strengthened the necessity of this approach specifically in combating malaria, where in addition to health interventions, improvement of households' levels of education and income status is being initiated by the project. The multilevel analysis enabled the study to explain all the sources of variation and compute the correct estimates of odds ratios and standard errors were essential for the conclusions drawn. On the other hand, some weaknesses were revealed of single-level models when they are applied in multilevel data. Failure to pay attention to the multilevel structure of data can easily lead to wrong inferences in understanding the factors influencing health outcomes as revealed by the study. Application of multilevel models in analysing the multilevel data has started a good number of years ago but they are rarely used in many parts of the world in understanding determinants of various health outcomes. The trend is even worse in Africa considering the inadequacy of data analysis software and skilled data analysts. Individuals affected by various fatal diseases which of more prevalent in Africa compared to any other continent belong to environments that have with them quite strong influences in these diseases. The study would highly recommend that for health policy makers to achieve the goals of disease eradication and reduction, multilevel modelling is fundamentally essential to understand determinants of various health outcomes.

REFERENCES

1.Akanbi, O., Badaki, J. A., Adreniran, Y., & Olotu, O. (2010). Effect of blood group and demographic 2.characteristics. *African Journal of Microbiology Research* , 877-880.

3.Boadu, K., & Trovato, F. (2006). Association of Social Class with Malaria Prevalence. *Canadian Studies in Population , Vol.33.2*, 271-299.

4.Diez-Roux, A. V. (1998). Bringing context back into Epidemiology:Variables and fallacies in Multilevel Analysis. *Am J Public Health* , 88:216-22.

5.Gelman, A., & Hill, J. (2008). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

6.Haque, U. (2011). Malaria Prevalence, Risk Factors and Spatial Distribution in Hilly Forest Area of Bangladesh. *PloS ONE* , e18908.

7.Karen, W. et al (2007). *Baseline Report MILLENNIUM RESEARCH VILLAGE SAURI, KENYA.* New York: Earth Institute at Columbia University.

8.Krieger, N., & Waterman, P. D. (2004). *MULTI-LEVEL MODELING:Modeling area-based socioeconomic measures: a multilevel perspective*. Retrieved from The Public Health Disparities: http://www.hsph.harvard.edu/thegeocodingproject/webpage/monograph/multilevel.htm

9.Lopez, A. D., Matters, C. D., Ezzati, M., Jamison, D. T., & Murray, C. J. (2006). Global and Regional Burden of Disease and Risk Factors:Systematic Analysis of Population Health Data. *The lancet , Vol.367*, 1747-57.

10.Mauny, F., Viel, J., Handschumacher, P., & Sellin, B. (2004). Multilevel Modeling and Malaria:A New Method for an Old Disease. *International Journal of Epidemiology , vol.367*, 33:1337-1344.

11.Rambe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata.* Texas: Stata Press.

12.Raudenbush, S., & Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* London: Sage Publications, Ltd.

13.Sintasath, D. M. (2005). Malaria Prevalence and Associated Risk Factors in Eritrea. *Am J. Trop. Med. Hyg.* , 682-687.

14.Snow, R. W. (1997). Relation between severe malaria morbidity in children and level of Plasmodium falciparum transmission in Africa. *349* (1650).

15.Trammer, M., & Trammer, E. (2012). *Carthie Mash Centre for Census and Survey Research(CCSR) Teaching web page.* Retrieved 02 24, 2014, from Carthie Mash Centre for Census and Survey Research(CCSR) website: www.ccsr.ac.uk/publications/teaching/multilevel-modelling.pdf

16. Goldstein, H.(2003). *Multilevel statistical models*. London: Arnold Publishers.

17. Williams, R. (2009). Using heterogenous choice models to compare logit and probit coefficients across groups. Sociological Methods & Research 37: 531–559.

18. Government of Kenya, Ministry of Health: National *Malaria Strategy 2001–2010. Division of Malaria Control*, April 2001.