UNIVERSITY OF NAIROBI

SCHOOL OF MATHEMATICS

Modeling of Malaria Prevalence in Kisumu County Kenya Using Logistic Regression

By

Geoffrey Munene Wachira

This research project is submitted to the School of Mathematics of the University of Nairobi in partial fulfillment of the requirement for the degree of Masters of Science in Biometry.

DECLARATION

This research project is my original work and has not been published for the award of any university degree

Geoffrey Munene Wachira   Reg No: I56/80474/2012

Signature: ____________________   Date: ______/______/__2014__

This research project is submitted as partial fulfillment for the degree of Masters of Science in Biometry with my approval as the university supervisor

Dr J. M. Ndiritu

Signature: ______________________________   Date: ____/____/2014

DEDICATION

I dedicate this project to my mum Virginia Wanjiru Mwai who nurtured me with the importance of education.

ACKNOWLDGEMENT

I acknowledge and thank all those who contributed to success of my research project and subsequent preparation of this report.

I would also like to thank my supervisor Dr. Ndiritu for willfully guiding me all the way in this research. Indeed his unreserved support and invaluable guidance is an inspiration.

My appreciation also goes to my family for giving me company and words of encouragement during the challenging times of this academic work and the entire research period. I am grateful for the moral support.

I am particularly grateful to all my fellow colleagues in the M.SC (Biometry) and M.SC (Social Statistics) class. Our endurance and support for each other has not been in vain.

In a special regard, I extend my appreciation to the University of Nairobi and particularly the School of Mathematics for shaping up my statistical skills. I am now more enlightened in this field.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

DDT	Dichlorodiphenyltrichloroethane
IRS	Indoor Residual Spraying
KHS	Kenya Health Service
LLINs	Long-Lasting Insecticide Nets
MOH	Ministry Of Health
MSAT	Mass screening and treatment
OLS	Ordinary Least Squares
RBM	Malaria Partnership program
RR	Rate Ratios
TB	Tuberculosis
UN	United Nations
UNICEF	United Nations Children's Fund
WHO	World Health Organization

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

Malaria has been a long life-threatening parasitic disease transmitted by female anopheles mosquitoes. This has contributed to child morbidity in the world. It threatens 2.4 billion people, or about 40% of the world's population living in the world's poorest countries, and more than one million deaths are attributable to the disease annually (WHO, 2000). According to WHO/UNICEF (2005), the disease is a major public health problem in Africa with over 200 million clinical episodes.

In semi-arid and highland regions of Africa, malaria is epidemic and causes deaths annually (Worall et al, 2004). However, the risks of morbidity and mortality associated with malaria, particularly in semi-arid and highland regions, vary spatially and temporally (Snow and Marsh, 2002). Most malaria infections, particularly in sub-Saharan Africa, are caused by Plasmodium falciparum. Malaria presents a major socio-economic challenge to African countries, considering that it is the most affected region. This challenge cannot go unnoticed given that good health is not only a basic human need, but also a fundamental human right and a prerequisite for economic growth (UN, 2003).

Malaria is caused by a parasite that is transmitted from one person to another through the bite of the Anopheline mosquito (a female Anopheles mosquito). When the mosquito bites an infected person, it ingests microscopic malaria parasites found in the person's blood. The malaria parasite must grow in the mosquito for a week or more before the infection is passed to another person. Thereafter, if the mosquito bites another person, the parasites pass from the mosquito's mouth to the person's blood. They feed on the blood cells, multiply inside the liver thereby destroying the red blood cells. This causes a cut off in blood circulation, which could lead to premature death, (WHO, 2000). Symptoms of malaria include fever, shivering, pains in the joint, vomiting, anemia, hemoglobinuria, retinal damage, and convulsions. The classic symptom of malaria is cyclical occurrence of sudden coldness followed by rigor then fever and sweating lasting four to six hours. This occurs every two days in plasmodium vivax

(*P. vivax)* and plasmodium Ovale (*P. ovale*) infections, while every three days for plasmodium malariae (P. *malariae*) (Nyika, 2009).

Malaria can be prevented by the use of mosquito coils and repellants, spraying the insides of houses (where most *Anopheles* species feed and rest) with insecticides (indoor residual spraying, IRS) and by sleeping under bed nets that have been treated with long-lasting insecticides (long-lasting insecticide nets, LLINs). Mass screening and treatment (MSAT) with effective anti-malarial drugs can also reduce malaria transmission (Griffin et al., 2010).

However, the levels of malaria risk and transmission intensity exhibit significant spatial and temporal variability related to variations in amount of rainfall, temperature, altitude, topography, and human settlement pattern (Abeku et al, 2003). The malaria situation in Kenya is typical of Sub-Saharan Africa making its transmission in Kisumu an all- year - round affair and seasonal variation (Afari *et al*. 1995). It is the major cause of morbidity and mortality, directly contributing to poverty, low productivity, and reduced school attendance in Kisumu region.

The Ministry of Health (MOH, 2009) records show that between 3-3.5 million cases of malaria are reported each year, over 900,000 of which are children under five years. Malaria is said to account for 61% of under-five hospital admissions and 8% of admissions of pregnant women. The country can be stratified into three malaria epidemiologic zones: the northern savanna; the tropical rainforest; and the coastal savanna and mangrove swamps (PMI, 2009).

Climatic factors, particularly rainfall, temperature and relative humidity have a strong influence on the biology of mosquitoes. In malaria endemic countries, climate factors reportedly contribute to the increased number of mosquitoes and thus make transmission favorable. Once adult mosquitoes have emerged, the ambient temperature, humidity, and rains will determine their chances of survival. Warmer ambient temperatures shorten the duration of the extrinsic cycle, thus increasing the chances of transmission (Jackson, 2010). Statistical assessment of malaria prevalence can be used to investigate associations between these climatic variables and the distribution of the different species responsible for malaria transmission (Sweeney, 1997).

Topography or elevation variables have variability, which when analyzed statistically could be used to determine the potential distribution of the vector and malaria risk areas. This study will model the malaria risks with topographic and climatic covariates, as well as the forest cover in the area to explore the variability of the disease using these factors.

1.2 Statement of the problem

Malaria continues to be an economic burden and a great threat globally and almost impossible to eradicate for the past six decades.

Malaria is by far the leading cause of death in Kisumu County. The disease is responsible for 25% of deaths of children below 5 years, and also claims the lives of many pregnant women (RoK, 2013). Malaria and mosquito control challenges operate at a wide spatial scale.

Prevalence of Malaria is also known at a limited number of specific sample locations. The pattern and variation of risk cannot be accounted for by only the known covariates. Data points of measured malaria prevalence are not evenly or randomly spread across the area to be mapped and must be interpolated spatially. Elevations have mostly influenced the rate of mosquito replication at different spatial scales. Logistic regression approach would be used to explore the effect of topography and the climatic conditions on the malaria prevalence in the study area. This study sought to do a statistical modeling of malaria prevalence in Kisumu County Kenya.

1.3 Objective of the study

1.3.1 General Objective

The general objective of the study was to establish a Logistic Regression model to the prevalence of malaria in Kisumu County Kenya

1.3.2 Specific Objectives

The study was guided by the following specific objective

- To establish the effects variation in climatic condition including temperature, amount of rainfall on prevalence of malaria in Kisumu County.

- To explore the effect of topography, land use and forest cover on prevalence of malaria in Kisumu County.

1.4 Research Questions

The study sought to answer the following research questions

i. What are the effects of variations in climatic condition on prevalence of malaria in Kisumu County?
ii. What are the effects of topography, land use and forest cover on prevalence of malaria in Kisumu County?

1.5 Justification of the Study

Malaria has become a great concern globally and has impacted negatively on the economies of developing nations. Health workers generally are unable to identify high or risk areas in the areas they operate so as to tailor interventions and do effective health monitoring. Research conducted so far by medical and climate professionals have either lacked knowledge or showed less concern for the variation of the climate conditions that accompany the transmission of malaria. The geographical distribution of any major disease forms an important basis for locating appropriate interventions for its control and a means to monitoring their effectiveness. It also provides a possibility for identifying ecological factors with which the disease may be associated.

The link between climate and medical data has not been well defined, and health information systems have been weak due to the lack of case detection, irregularity in reporting, under reporting and poor coordination (Dziedzom, 2009). There is a need for a risk map to draw attention to hot spots and areas where intervention measures can be tailored to improve the monitoring of the occurrences, distribution and control of malaria in different geographical areas and time periods. Statistical modeling using Logistic Regression is also necessary to correlate the factors that are associated with these variations in rainfall, land use and temporal heterogeneity of malaria transmission at the different geographic locations.

1.6 Summary

The paper is generally structured under five chapters as outlined below

- Chapter Two is the literature review of previous work done on the disease and how temperature, amount of rainfall, topography and land use and forest cover has been used in the disease application.
- Chapter Three outlines Material and Methods used.
- Chapter Four discusses the results which are the outcome of materials applied in the study.
- Chapter Five deals with Conclusion and Recommendation of the study.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Malaria is caused by a parasite called Plasmodium, which is transmitted through the bites of infected female anopheles mosquitoes. In the human body, the parasites multiply in the liver, and then infect red blood cells. There are four of this different species causing the human malaria disease: Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale and Plasmodium malaria. (WHO, 2010)

2.2 Malaria Situation

Malaria is a vector-borne disease that is widespread in the tropical and subtropical areas of the world. This has become a serious challenge for most developing countries where between 300 and 500 million people are infected annually. There have been several efforts by government and other development partners in the health sector to eradicate malaria in the country, but its prevalence rate is still on the increase. This has prompted the question why the malaria cases are still on the increase despite these efforts.

The country is still at the control stage of Malaria programme, which is the first step in the fight against the disease. The second stage is the elimination of the disease, which must be supported by a functioning health system, while the third stage is the eradication, which is usually global and undertaken when vaccines exists (GNA, 2010).

Developing countries are still having a chunk of their national budget being used for malaria eradication. Studies according to (GNA, 2010) have established the fact that malaria affects mostly the poor impacting negatively on their socio-economic lifestyle. The burden of malaria is therefore greatest among the world’s poorest countries (Worrall, 2003). The disease in Kenya has impoverished most of the poor communities. The adult population is mostly affected reducing working hours and lowering productivity and economic livelihood impacting substantially on the national gross domestic product, (Asante and Asenso-Okyere, 2003).

2.3 Malaria in Kenya

In Kenya, malaria is the number one cause of morbidity accounting for 40-60% of outpatient reports (Asante and Asenso -Okyere, 2003). Reported malaria cases represent only a small proportion of the actual number of episodes as majority of people with symptomatic infections are treated at home and are, therefore, not reported (World Malaria Report, 2005). In a recent study in Kenya, it was found out that the cost of malaria care is just 1% of the income of the rich households, and 34% of the income of the poor. Usually malaria attacks are associated with poor social, economic and environmental conditions. The main victims are the poor who are often forced to live on marginal lands. Malaria endemic communities are therefore caught in a vicious circle of disease and poverty, (Akazili, 2002). In Kenya the vulnerability of the disease is about 50% making Kenya to spend considerable amount of money on mosquito control products such as fly proof nets and mosquito repellent (KHS, 2004).

Kenya has stepped up progress ever since malaria became endemic in the country. Through various donor agencies and seminars, there has been a concerted effort by the government and the Kenya Health Service (KHS) to make the country a malaria free state. Since the 1950 -1960’s: the World Health Organization (WHO) has supported indoor residual spraying using Dichlorodiphenyltrichloroethane (DDT) in many households. Chloroquine drug was used extensively and intensified at all health facilities in Kenya between the 1970’s and 1980’s.

There are large among-site variations in the abundance and temporal dynamics of malaria vector populations indicating that the risk of parasite transmission differs among sites, (Munyekenye, *et al*, 2005). Even in one topographic area, mosquito vectors and malaria infections may not be distributed homogeneously, and some households within the same area have a higher malaria incidence than others, (Njagi *et al.*, 2004).

Many factors may be responsible for this spatial heterogeneity of malaria vectors and transmission intensity such as land use and land cover changes, topography, house building materials, and design and the level of household protection measures against mosquitoes. In most cases, it is difficult to identify the factor that contributes most to these variations.

2.4 Malaria in Kisumu

In Kisumu, malaria is predominantly a rural disease, and the main malaria vectors are Anopheles gambiae sensu stricto, An. Arabiensis, and An. Funestus, (Zhou et al, 2006). Anopheles gambiae generally increases in density after the start of the long rains, while An. funestus density is seen to vary in direct proportion to the proximity of permanent breeding grounds rather than rainfall (Njagi *et al.,* 2004). In the adult stage, these anopheline species share many of the same habitats. In the Usambara Mountains, Tanzania, and in Kisumu Kenya, Githeko et al. (2003) reported that altitude plays an important role in determining malaria infection due to its effect on temperature. Temperature decreases with increasing altitude, and at lower altitudes, the high temperature levels accelerate the sporogonic cycle of malaria parasites in the presence of vectors and the breeding habitats. Land use such as deforestation and swamp reclamation by eliminating shade modifies the local climate and microclimate, and in the presence of stagnant water, new habitats for malaria vectors are formed. Consequently, the new habitats provide new breeding grounds leading to increased vector densities and subsequently an increase in malaria transmission. Over the past four decades, deforestation and swamp cultivation have widely occurred in Kisumu, and these are now thought to be a major contributing factor to the abundance of breeding habits and the survival of malaria vectors. The ever-increasing human population and the need for food security place large pressure on land and threaten the survival of undisturbed natural forests and swamps.

2.5 Environmental Risk Factors of Malaria

2.5.1 Rainfall

Malaria is greatly influenced by rainfall in the tropics. It creates an opportunity for *anopheles* mosquitoes to lay eggs, which can reach adulthood within nine to twelve (9-12) days, necessary for the mosquito life cycle. Rainfall is one of the climatic variables that aid in the multiplication of mosquito breeding places and increases humidity, which improves mosquito survival rates. The rainy season is a fertile period for the breeding sites, which are numerous. These species have the highest population density during the rainy season and these accounts for the high incidence of malaria at this period of the year (Reid, 2000). Studies have established complex relationship between malaria and rainfall because water is very vital for larval development. A prolonged dry season can decrease mosquito numbers by reducing breeding sites and also minimizes malaria incidences.

2.5.2 Topography

Topography generally has a great influence on mosquito replication and thus affects the rate of malaria cases. Higher topographies results in cooler temperatures, which limits the reproduction rate of the parasite. Entomologic studies in eight villages to investigate the patterns of malaria transmission in different ecologic zones in Eritrea showed a positive relationship between the malaria cases and topography. Mosquito collections conducted for 24 months showed that the biting rates in the higher elevations as a result of the lower temperatures were twice as high as the lowlands (Shillu et al., 2003). The complexity of topography and landscape in the highlands contributes to the spatial heterogeneity of vector abundance and malaria transmission intensity. It has implications for the survival of the vector for different altitudes (Minakawa et al., 2002).

Githeko et al. (2003) investigated whether the risk of infection with malaria parasites was related to topography in the Usambara Mountains, Tanzania. Clinical surveys were carried out in seven villages, situated at altitudes from 300m to 1650m. Each village was mapped and incorporated into a Digital Terrain Model. Univariate analysis showed that the risk declined with increasing topography. This was attributed to the fact that such elevations washed away water when it rained therefore decreasing potential of water to stagnate, creating breeding sites for mosquitoes.

Lindsay et al (2000), discussing the effect of the 1997-98 El Niño on highland malaria in Tanzania, discovered quite an opposing results of the malaria incidence with the associated highlands. The study showed that the level of malaria infection was rather following this event than in the previous year, suggesting that heavy rainfall may have washed away mosquito breeding sites

Cohen et al (2008) in their study of topography-derived wetness indices and household-level malaria risk in two communities in the western Kenyan highlands and tried to show the effect of topography on the malaria incidence. They found that the transmission of *Plasmodium falciparum* generally decreases with increasing topography.

2.5.3 Temperature

Malaria incidence is closely linked with temperature. It affects malaria transmission in several ways among which we can account for two reasons: either the minimum temperature is so low that it prevents parasite and vector development or the temperature is too high resulting in increased mortality of the vector. A minimum temperature of 16 degrees Celsius restricts parasite development and also prevents the development of the vector in its aquatic stages. At 17 degrees Celsius parasites develop, but not rapidly enough to cause an epidemic (Lindsay and Martens, 1998).

Temperature also plays a fundamental role in the rate of multiplication of the parasite in mosquitoes and directly influences the mosquito development, gonotrophic cycle and longevity, as well as the duration of the extrinsic cycle of the Plasmodium parasite. In warmer temperatures, the mosquitoes develop more rapidly, accelerating the mosquito life cycle and replicating the parasite growth rapidly (WHO/AFRO (2001).

The optimum temperature for the malaria parasite extrinsic incubation period is about 20º-27ºC, while the maximum temperature for both vectors and parasites is 40°C (MARA/ARMA, 1998). Malaria transmission in areas colder than 20°C can still occur because Anophelines often live in houses, which tend to be warmer than external temperatures. Larval development of the mosquito also depends on temperature. Higher temperatures increase the number of blood meals taken and the number of times eggs are laid by the mosquitoes (Martens et al, 1995).

Broker et al. (2002), studied to see the spatial distributions of Helminth (one type of parasites) in Cameroon. They collected epidemiological and population data. Land surface temperature was derived from NOAA-AVHRR. They used a Logistic regression model to identify significant environmental variables, which affect the transmission of infection. The variables used in the regression analysis were mean, minimum and maximum land surface temperature; total annual rainfall and altitude. The result revealed that maximum temperature was an important variable in determining Helminth distribution. At higher temperatures, it is realized that female adult mosquitoes feed more frequently and digest blood more rapidly.

2.5.4 Land Use and Forest cover

Land use and land cover changes have a significant influence on malaria transmission intensity. It affects the spatial and temporal variations in the distribution of anopheline larval habitats. In a study investigated by Yeshiwondim et al (2009), the spatial and temporal variations in the distribution of anopheline larval habitats and land use changes in western Kenya highlands over a 4-year period showed that Anopheles gambiae complex larvae were mainly confined to valley bottoms during both the dry and wet seasons. They were also located in man-made habitats where riparian forests and natural swamps had been cleared. The association between land cover type and occurrence of anopheline larvae was statistically significant.

Forest cover may double the high rate of malaria in some of the areas recording high malaria cases. The disease incidence is very high in the forest and forest fringes as compared to plains or urban areas (Sharma, 1991). Mosquitoes in the forested area according to the study were seen to live longer, than those in the deforested area in both dry and rainy seasons in the highlands. Forested areas have high humid conditions which favour the ecological reproduction.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines the model in statistical applications of malaria data to be applied. Wide availability of advances in technology have also allowed for the collection of vast quantities of data with geo-referenced sample locations.While choosing model,greater attention was given to one that incorporates information that influences the response variable despite the fact that not everything associated with the response is known.

3.2 Study Areas

The study was carried out in Kisumu County in two highland villages, Lunyerere and Fort Ternan, and the lowland periurban Nyalenda, a suburb of Kisumu city. The area is characterized by broad U-shaped valleys that exhibit different topography. Majority of the valley bottoms in this area were previously forested covered with natural swamps that were fed by water through underground seepage. However, in recent times, the land has been cleared to farmland, where the community members practice small-scale food crop farming.

3.3 Data Collection

The demographic data of the areas involved in the study was obtained using the 2009 Kenya population Census Data. The populations from 2007 to 2013 used in the study was extrapolated using the 1999 and 2009 growth rates computed with the help of SPSS software. The meteorological data comprising the rainfall and temperature data was obtained from Nyanza Regional Meteorological office, Kisumu. The monthly rainfall readings were obtained, while the temperature readings were in degrees Celsius. The malaria cases data was obtained from the Kisumu General hospital. The malaria incidence per hundred (100) people of the population was then calculated i.e. Prevalence= {(number of cases/population)*100} for the communities under study. The malaria rates from 2009-2013 will be computed as well as the total rates of the years under study. The SPSS statistical software was used for the graph of relations between malaria prevalence and temperature/rainfall. It was also used in the linear regression analysis of the buffered distances from the rivers/forest /elevations and malaria rates. The present study hypothesizes that the risk of malaria infection has a dynamic relationship with elevation. The distance in kilometers from the valley bottoms was used as the measure of elevation/topography and was put in a scale of 1-4; where 1 represents a location that is less than 1km from the valley bottom, 2 represents a location that is 1-2kms from the valley bottom, 3 represents a location that is 3-4kms from the valley bottom and 4 represents a location that is more than 4kms from the valley bottom. The extent of land use and forest cover was also scaled from 1-4; where 1 represents plain lands, 2 represents moderately grassed land, 3 represents land with grass and vegetation covers while 4 represents a land locations with thick vegetation covers. Here, the study adopts a Logistic regression model for the effects of temperature, amount of rainfall, elevation and land and forest cover on the risk of malaria infection.

3.4 Statistical Analysis of Malaria Prevalence

3.4.1 Logistic Regression analysis

The relationship between malaria parasite prevalence and each individual potential explanatory variable was first investigated by inspection of scatter-plots and by single variable regression analysis. Since parasite prevalence data are binomial fractions, a logistic regression model for grouped (blocked) data will be used as is standard practice for the analysis of such data (Hosmer and Lemshow, 1989). Predictions of prevalence made from the logistic model will always fall within the interval of probabilities of 0 to 1. Larger surveys are implicitly accorded more weight than the smaller ones. The SAS system will be used for the analysis. Each of the explanatory variables was adjusted for all of the others by performing multiple regressions in the usual way. Non-linearity in the relationship between parasite prevalence and a predictor variable was explored by adding polynomial terms and then grouping the values of continuous variables into categorical ones.

A logistic regression model will be used to assess the effects of variations in; temperatures (T°C), topography, land use and forest cover and amount of rainfall (Pmm) of the previous months on clinical malaria rates among study participants. The interaction terms of the variables will be included in the model to control for the high correlation of the three individual variables.

Since temperature and rainfall are continuous variables and their relationships with clinical malaria might not be linear, multivariate procedures will be used to determine the best-fitting relationship they had with clinical malaria.

In the problems associated with topographical prediction, it was noted that degree of correlation among observations depends on their relative locations. This is due to similarities in topographical attributes and the dispersions associated with the forest cover and land use. The products of this is the autocorrelation and the models seeks also to respond to this topographical dependence while retaining the climatic condition (temperature and amount of rainfall) response which is catered for by the logistic (Logit) function.

When modelling in logistic regression, a formula is required to convert back and forth from the logistic equation to the OLS- type equation. The Logistic regression model has the form;

$$\mathrm{logit(Y) = naturallog(odds)=ln}\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \tag{3.1}$$

Taking the antilog of Equation 3.1 on both sides, one derives an equation to predict the probability of the occurence of the outcome of interest as follows:

$$\pi = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x}} \tag{3.2}$$

where π is the probability of outcome of interest or “event” such as a patient tests positive for malaria prevalence. β_0 is the Y intercept, $\beta_0, \beta_1 \ldots \beta_p$ are the regression coefficients, $x_1, x_2 \ldots x_p$ are the explanatory variables and $e = 2.71828$ is the base of the system of natural logarithms.

This model will be used to predict malaria prevalence from the climatic conditions (rainfall and temparature changes), topography, and land use and forest cover in the region of kisumu.

Note that linear regression would not work well here since it could produce probabilities less than zero or greater than one.

Equation (3.2) describes a family of sigmoidal curves and will be used to give the probability of malaria prevalence.

Probability that the patient has no malaria is thus given by, $1-\pi(x)$

$$=\frac{1+e^{\beta_0+\beta_1 x_1+...+\beta_p x_p}-e^{\beta_0+\beta_1 x_1+...+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+...+\beta_p x_p}} \tag{3.3}$$

$$1/(1+e^{\beta_0+\beta_1 x_1+...+\beta_p x_p}) \tag{3.4}$$

Equation (3.4) gives the odds of malaria prevalence.

The log odds of malaria prevalence thus equal,

$$\ln(\pi(x)/1-\pi(x))=\beta_0+\beta_1 x_1+...+\beta_p x_p \tag{3.5}$$

Equation (3.5) can as also be represented as;

$$\ln(\frac{\pi(x)}{1-\pi(x)})=e^{\beta_0+\beta_1 x_1+...+\beta_p x_p} \tag{3.6}$$

The regression coefficients are typically estimated by the maximum likelihood (ML) method, which is preferred over the weighted least squares approach.

3.5 Maximum Likelihood (ML) method in Logistic Regression

The ML method is designed to maximise the likelihood of reproducing the data given the parameter estimates. Data are entered into the analysis of 0 or 1coding for the dichotomous outcome, continuous values for continuous predictors, and dummy coding (e.g., 0or 1) for categorical predictors.

The null hypothesis underlying the overall model states that all βs equal zero. A rejection of this null hypothesis implies that at least one β does not equal zero in the population, which means that the logistic regression equation predicts the probability of the outcome better than the mean of the dependent variable Y. The interpretation of results is rendered using the odds ratio for both categorical and continuous predictors.

The goal of logistic regression is to estimate the p+1unknown parameters β in the logistic regression. This is done with maximum likelihood estimation which entails finding the set of parameters for which the probability of the observed data is greatest. The maximum likelihood equation is derived from the probability distribution of the dependent variable.

Since each y_i represents a binomial count in the i^{th} population, the joint probability function of Y is:

$$f(y|\beta) = \prod_{i=1}^{N} \frac{n_i!}{y_i!(\mathrm{n}_i - \mathrm{y}_i)!} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i} \tag{3.7}$$

The maximum likelihood estimates are the values for β that maximize the likelihood function in Eq.3.7. The critical points of a function (maxima and minima) occur when the first derivative equals 0. If the second derivative evaluated at that point is less than zero, then the critical point is a maximum.

After rearranging the terms, the equation to be maximized can be written as;

$$\prod_{i=1}^{N} \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i} (1-\pi_i)^{n_i} \tag{3.8}$$

After taking *e* to both sides of equation (3.9) we have

$$\left(\frac{\pi_i}{1-\pi_i}\right) = e^{\sum_{k=0}^{k} x_{ik}\beta_k} \tag{3.9}$$

Which after solving for π_i, Equation (3.10) becomes

$$\pi_i = \left(\frac{e^{\sum_{k=0}^{k} x_{ik}\beta_k}}{1 + e^{\sum_{k=0}^{k} x_{ik}\beta_k}}\right) \tag{3.10}$$

Substituting Eq. (3.9) for the first term and Eq. (3.10) for the second term, Eq.3 become,

$$\prod_{i=1}^{N} \left(e^{\sum_{k=0}^{K} x_{ik}\beta_k}\right)^{y_i} \left(1 - \frac{e^{\sum_{k=0}^{K} x_{ik}\beta_k}}{1 + e^{\sum_{k=0}^{K} x_{ik}\beta_k}}\right)^{n_i} \tag{3.11}$$

By simplifying the first product, equation (3.11) can be written as;

$$\prod_{i=1}^{N} \left(e^{y_i \sum_{k=0}^{K} x_{ik}\beta_k}\right)\left(1 + e^{\sum_{k=0}^{K} x_{ik}\beta_k}\right)^{-n_i} \tag{3.12}$$

This is the kernel of the likelihood function to maximise. This can be simplified by taking its logs so as to differentiate. Thus, taking the natural log of Eq.(3.12) yields the log likelihood function:

$$l(\beta) = \sum_{i=1}^{N} y_i (\sum_{k=0}^{K} x_{ik} \beta_k) - n_i . \log(1 + e^{\sum_{k=0}^{K} x_{ik}\beta_k}) \tag{3.13}$$

To find the critical points of the log likelihood function, set the first derivative with respect to each β equal to zero. In differentiating Eq.(3.13), we get,

$$\frac{\partial}{\partial \beta_k} \sum_{k=o}^{K} x_{ik} \beta_k = x_{ik} \tag{3.14}$$

Since the other terms in the summation do not depend on β_k and can thus be treated as constants. In differentiating the second half of Eq. 3.13 with respect to β_k we get,

$$\begin{aligned}
\frac{\partial l(\beta)}{\partial \beta_{k\partial}} &= \sum_{i=1}^{N} y_i x_{ik} - n_i . \frac{1}{1 + e^{\sum_{k=0}^{K} x_{ik}\beta_k}} . \frac{\partial}{\partial \beta_K} (1 + e^{\sum_{k=0}^{K} x_{ik}\beta_k}) \\
&= \sum_{i=1}^{N} y_i x_{ik} - n_i . \frac{1}{1 + e^{\sum_{k=0}^{K} x_{ik}\beta_k}} . e^{\sum_{k=0}^{K} x_{ik}\beta_k} . \frac{\partial}{\partial \beta_k} \sum_{k=o}^{K} x_{ik}\beta_k \\
&= \sum_{i=1}^{N} y_i x_{ik} - n_i . \frac{1}{1 + e^{\sum_{k=0}^{K} x_{ik}\beta_k}} . e^{\sum_{k=0}^{K} x_{ik}\beta_k} . x_{ik} \\
&= \sum_{i=1}^{N} y_i x_{ik} - n_i \pi_i x_{ik}
\end{aligned} \tag{3.15}$$

The maximum likelihood estimates for β can be found by setting each of the K+1 equations in the Eq. (3.14) to zero and solving for each β_K .

Each such solution, if any exists specifies a critical point either a maximum or a minimum.

3.6 Fitting data to the Logistic Regression model

In some instances the model may not reach convergence. When a model does not converge this indicates that the coefficients are not meaningful because the iterative process was unable to find appropriate solutions. A failure to converge may occur for a number of reasons: having a large proportion of predictors to cases, multicollinearity, sparseness, or complete separation.

- Having a large proportion of variables to cases results in an overly conservative Wald statistic and can lead to nonconvergence.
- Multicollinearity refers to unacceptably high correlations between predictors. As multicollinearity increases, coefficients remain unbiased but standard errors increase and the likelihood of model convergence decreases. To detect multicollinearity amongst the predictors, one can conduct a linear regression analysis with the predictors of interest for the sole purpose of examining the tolerance statistic used to assess whether multicollinearity is unacceptably high.
- Sparseness in the data refers to having a large proportion of empty cells (cells with zero counts)
- Another challenge that may lead to a lack of convergence is complete separation, which refers to the instance in which the predictors perfectly predict the criterion.

3.7 Evaluating goodness of fit

3.7.1 Deviance and likelihood ratio tests

In logistic regression analysis, deviance is used in lieu of sum of squares calculations. This is same as the sum of squares calculations in linear regression and is a measure of the lack of fit to the data in a logistic regression model. Deviance is calculated by comparing a given model with the saturated model – a model with a theoretically perfect fit. This computation is called the likelihood-ratio test:

$$D = -2\ln\ L_1 \qquad (3.6.1)$$

Where;

$$L_1 = \frac{\text{likelihood of the fitted model}}{\text{likelihood of the saturated model}}$$

D represents the deviance

ln represents the natural logarithm

Two measures of deviance are particularly important in logistic regression: null deviance and model deviance. The null deviance represents the difference between a model with only the intercept (which means "no predictors") and the saturated model. And, the model deviance represents the difference between a model with at least one predictor and the saturated model. In this respect, the null model provides a baseline upon which to compare predictor models. Given that deviance is a measure of the difference between a given model and the saturated model, smaller values indicate better fit. Therefore, to assess the contribution of a predictor or set of predictors, one can subtract the model deviance from the null deviance and assess the difference on a χ^2_{s-p}, chi-square distribution with degree of freedom equal to the difference in the number of parameters estimated.

Let,

$$D_{null} = -2\ln\ L_2 \tag{3.6.2}$$

Where, $L_2 = \dfrac{\text{likelihood of the null model}}{\text{likelihood of the saturated model}}$ and,

$$D_{fitted} = -2\ln\ L_1 \tag{3.6.3}$$

Then,

$$D_{fitted} - D_{null} = -2l\ L_1 - (-2\ln\ L_2)$$

$$= -2\ L_1 - L_2$$

$$= -2\ln(\frac{L_1}{L_2})$$

$$= -2\ln\frac{\text{likelihood of the fitted model}}{\text{likelihood of the null model}} \tag{3.6.4}$$

If the model deviance is significantly smaller than the null deviance then one can conclude that the predictor or set of predictors significantly improved model fit.

3.7.2 Pseudo-R^{2}s

It represents the proportional reduction in the deviance wherein the deviance is treated as a measure of variation analogous but not identical to the variance in linear regression analysis. One limitation of the likelihood ratio R^2 is that it is not monotonically related to the odds ratio, meaning that it does not necessarily increase as the odds ratio increases and does not necessarily decrease as the odds ratio decreases.

The Pseudo-R^{2}s is represented as;

$$R_L^2 = \frac{D_{null} - D_{model}}{D_{null}}$$

3.7.3 Hosmer–Lemeshow test

The Hosmer–Lemeshow test uses a test statistic that asymptotically follows a χ^2 distribution to assess whether or not the observed event rates match expected event rates in subgroups of the model population.

3.8 Data analysis

The data will be coded and then a data base will be developed in statistical software. Due to the type of analysis required R-Software and Microsoft Excel will be used. A form of matrix will be developed where temperature and amount of rainfall were each given a variable name. The data base was first established in excel sheet and saved as CSV comma delimited. Then R Command were used to export the data from excel to R-Software where modeling and analysis was done.

The data obtained was first arranged in logical order followed by drawing tables and graphs. Descriptive analysis was used and also an aspect of correlation to show if there is any association between various climatic condition factors. The data was fitted to give a logistic model with various aspects relating the malaria prevalence to the climatic conditions.

CHAPTER FOUR: RESULTS AND DISCUSSION

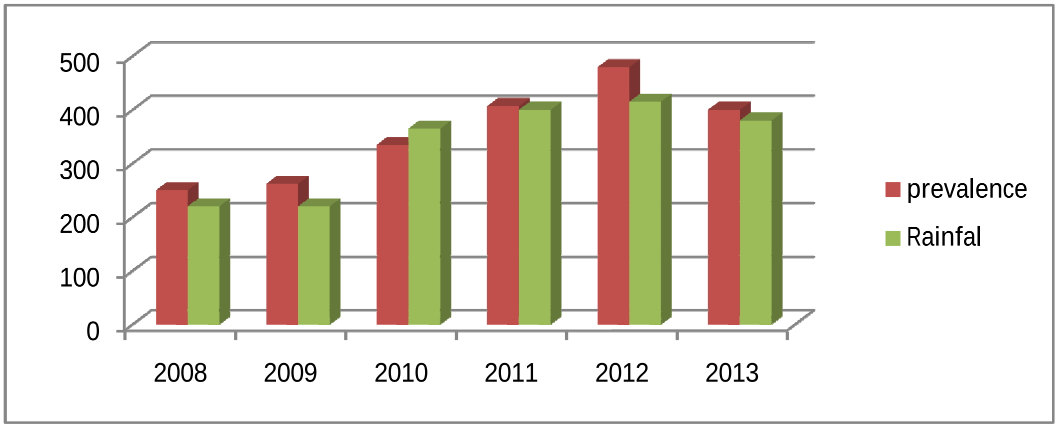
4.1 Results and Discussion

This chapter presents the data analysis, presentation and interpretation of the study, the study analyzed to establish Logistic Regression model of malaria prevalence in Kisumu County Kenya. One of the objectives of this study was to create a risk map using geostatistical approach. The risk map created in such instances highlights areas of high risk that needs to be identified so as to tailor major interventions and monitoring activities.

There was generally a fair increase in the malaria risk from the year 2009 to 2013. Malaria rose averagely by 16% from 2007 to 2009 and by 20% during the period 2009 to 2012. There were many areas reporting high risk and low risk respectively in the recurring years. Other high risk areas saw some gradual reduction especially in the other years. There was a steady rise in the general prevalence of the disease but a reduction in the key towns that were reporting in the previous years. A great improvement concerning risk was seen in 2011. There was an average risk occurrence in those areas except for areas around the district capital that was still reporting very high cases. The cumulative malaria risk was consistent with the general steady rise in the incidences of the years under study.

4.1.1 Malaria and Rainfall

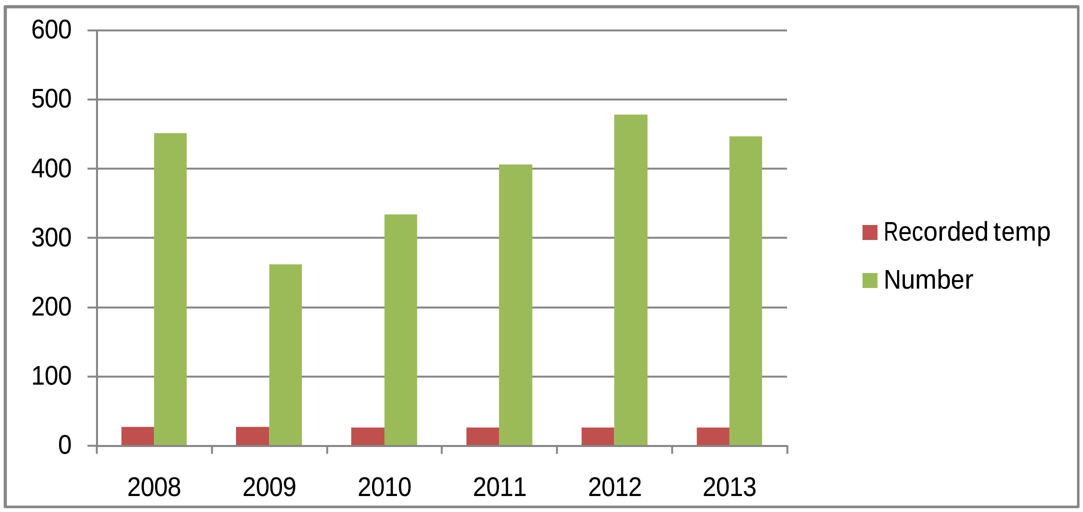
Figure 4.1: Graph of Prevalence against Rainfall



From the finding on figure shown above, the study found that the increase in rainfall resulted in an increase in the malaria prevalence. The study area according to the meteorological office, generally record one of the highest amount of rainfall in the region. Rainfall therefore increases parasitic density soon after the start of the rainy season because the rains provided good breeding sites for the mosquito vectors. As the vector population increased, transmission of infection was enhanced thus the increase in the vector density. Moreover, rainfall may have increased the humidity which could have improved mosquito survival rates.

4.1.2 Malaria and Temperature

Figure 4.2: Graph of malaria prevalence against Temperature



From the finding shown in figure 4.2 above, the study revealed that temperature did not have any effect on malaria prevalence as the figure did not show any pattern. The relationship between temperature and malaria prevalence in Figure 4.2 above showed no trend. There may have been more overriding factors other than temperature such as agricultural practice, human migration, population density, poverty and access to health services resulting in this high prevalence at low temperatures. Temperature could not be justified in this study to be accounting for the lower prevalence that escalates malaria at higher altitudes because of the fairly stable temperature all-round the region.

4.2 Results of the Statistical Model

Table 4.2: Regression Results of Covariates

Covariate	P-Value	R-Square	T-Value
Rainfall	0.04	0.36	1.5
Temperature	0.32	0.16	0.26
Forest cover	0.47	0.10	0.64
Topography	0.03	0.14	0.42

From the finding on the results shown in table 4.1 above, it shows increase in rainfall was significant to the malaria prevalence (P-value 0.04< 0.05)

This confirms that rainfall increases parasitic density soon after the start of the season. Rains provided good breeding sites for the mosquito vectors as well (Patz, 2001).

Temperature did not affect the prevalence of malaria significantly (P-value 0.32>0.05). Elevation factors, Land use and forest covers did not show significant effect of malaria prevalence as well.

4.2.1 Result of the Logistic Regression Model

$$\log\left(\frac{p}{1-p}\right) = -1.04 - 0.23tem + 0.62rai + 0.27top + 0.42\,for$$

The fitted Logistic Regression model is,

- temp=Temperature
- rain= Amount of rainfall
- top= Scale of topography of the region
- for= scale of forest cover and land use

For every unit change in temperature, the log odds of malaria prevalence decreased by 0.23. This can be interpreted to mean that temperature changes have no effect to rates of malaria prevalence in Kisumu County.

Rainfall in the region increases the log odds of malaria prevalence by 0.62

For every unit change in topography, the log odds of malaria prevalence increases by 0.27

For every unit change in forest cover and land use, the log odds of malaria prevalence increases by 0.42

4.2.2 Malaria and Forest Cover

From the finding on the results shown in table 4.2 above, the results indicates that areas nearer the forest were reporting higher cases though statistically weak (p-value 0.47<0.05)

4.2.3 Malaria and Topography

The model shows that higher elevations in some instances resulted in fewer cases which follow the normal trend and higher altitudes have less favourable ecological factors unlike low temperatures that can trigger malaria rise. However some higher elevations recorded higher cases consistent with the malaria incidence. Higher elevation in general has long been recognized to be associated with malaria due to its association with cooler temperatures that slows the development of anopheline vectors and the Plasmodium parasites they transmit. Most of the hills in the study area have no settlements as they normally even have forests in those areas. Settlements around higher elevations however showed varied malaria cases as confirmed by the regression models clearly showing that disease incidence is not homogeneous. In this case malaria risk displayed an alternating result with elevation.

The other areas on lower elevations may be closely related to the streams/swamps and forest impact. This may have been as a result of the terrain being suitable for water accumulation on valley bottoms. In those areas water is not washed away when it rains as in the mountainous or higher altitudes. The fact that the elevation varied in some instances with the disease prevalence may have been as a result of the fact that the elevation differences was generally insignificant on the district spatial scale as compared to generally known above 2000m altitudes.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

5.1 Summary of Findings

The disease incidence was seen to be very high in the forest and forest fringes as compared to plains or non-forested areas within 1km-3km.These areas as a result of their high humid conditions with corresponding higher rains enhances the ecology of the vector to thrive thereby increasing the disease prevalence. Mosquitoes in forested areas according to Afrane et al. (2006) live longer than those in the deforested area in both dry and rainy seasons in the highlands. The forest habitats coupled with its humid conditions therefore support rapid multiplication of the vector. This may have resulted in the high cases in the forested areas. Most of the settlements are not within the forest except for the few (within the forest) that are not only influenced by the forest conditions but also on valley bottoms where the streams and rivers pass.

These environments often have areas where water could stagnate. Beyond 4km the disease prevalence could be attributable to other factors such as pits, untidy surroundings marshy areas and the non-usage of mosquito treated nets and repellants. Higher elevation in general has long been recognized to be associated with malaria due to its association with cooler temperatures that slows the development of anopheline vectors and the Plasmodium parasites they transmit. Most of the hills in the study area have no settlements as they normally even have forests in those areas. The other areas on lower elevations may be closely related to the streams/swamps and forest impact. This may have been as a result of the terrain being suitable for water accumulation on valley bottoms. In those areas water is not washed away when it rains as in the mountainous or higher altitudes.

The fact that the elevation varied in some instances with the disease prevalence may have been as a result of the fact that the elevation differences was generally insignificant on the district spatial scale as compared to generally known above 2000m altitudes. It can therefore be deduced that different factors (rivers/streams, forest, elevation, temperature and rainfall) affect the disease risk differently.

The study revealed that increase in rainfall resulted in an increase in the malaria prevalence. The study area according to the meteorological office, generally record one of the highest amount of rainfall in the region. Rainfall therefore increases parasitic density soon after the start of the rainy season because the rains provided good breeding sites for the mosquito vectors. As the vector population increased, transmission of infection was enhanced thus the increase in the vector density. Moreover, rainfall may have increased the humidity which could have improved mosquito survival rates.

The relationship between temperature and malaria prevalence in Kisumu County showed no trend. There may have been more overriding factors other than temperature such as agricultural practice, human migration, population density, poverty and access to health services resulting in this high prevalence at low temperatures. Temperature could not be justified in this study to be accounting for the lower prevalence that escalates malaria at higher altitudes because of the fairly stable temperature all-round the region.

These results highlights the fact that malaria occurrence within Kisumu County was not imported and that the local differences in topographic variables, rainfall, temperature and forest may be the reason accounting for the small spatial dependencies in the malaria transmission

5.2 Conclusion

The predictive results of the Logistic Regression Model to the disease prevalence gives credence to the fact that the covariates used which were forest and land cover, temperature, rainfall and elevation had different and independent influence on the malaria prevalence.

There was a significant varied effect of elevation with the disease prevalence. This varied statistical relationship may have resulted from the small spatial scale of the district with elevation differences less than 50m. There was a general trend of high disease incidence between 1-3 km from the forest edge and different factors beyond 4km. The annualized rainfall pattern showed a relationship with the disease prevalence. With the high levels of rains increasing the disease occurrence as it served as effective breeding grounds for the mosquitoes to thrive.

5.3 Recommendations

The study results suggest that there are 'malaria hot-spots' in the study area. The government and other Health related Non-Governmental Organizations should consider these results when planning malaria control measures. In particular, malaria Risk maps should be updated on a regular basis as new data emanates and a concerted effort targeted towards areas nearer the forest zones.

Efficient data gathering systems should be employed to obtain data in most of the areas to improve prediction of the risk in the area. If the malaria data is graded by month, there can be a seasonal evaluation of the disease burden.

There should be a further study using remote sensing technologies for a change detection to further explore the effect of land use and forest cover on the disease.

Since there was a trend with elevations but not too significant, it would be important to do further studies on a larger spatial scale in the region or national level. This would even bring to light the rate at which the different malaria species can thrive with changing altitudes.

REFERENCES

Abeku TA, de Vlas SJ, Borsboom GJJM, Tadege A, Gebreyesus Y, Gebreyohannes H, Alamirew D, Seifu A, Nagelkerke NJD, Habbema JDF., 2003. Effects of meteorological factors on epidemic malaria in Ethiopia: a statistical modeling approach based on theoretical reasoning. *Parasitology, 128:585-593*.

Afari EA, Appawu M, Dunyo S, Baffoe-Wilmot A & Nkrumah FK.,1995. Malaria infection, morbidity and transmission in two ecological zones in Southern Ghana.*African Journal of Health Sciences* 2, 312-315.

Akazili, T. 2002. *Geographical information systems*. Viakt IT, Arendal, Norway. WHO (World Health Organization). 1993. *WHO study group on the implementation of the global plan of action for malaria control*. WHO, Geneva, Switzerland. Technical report series, 839.

Asante FA & Asenso-Okyere K., 2003.Economic burden of malaria in Ghana.Technical report submitted to the World Health Organization (WHO), Africa Regional Office (AFRO) University of Ghana, Legon

Asnakew K Yeshiwondim , Sucharita Gopal, Afework T Hailemariam, Dereje O Dengela, Hrishikesh P Patel. Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia: International Journal of Health Geographics 2009, 8:5

Brooker, S. Clarke, J. K. Njagi et al.(2004), “Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya,” Tropical Medicine and International Health, vol. 9, no. 7, pp. 757–766, 2004.

Dziedzom AP., 2009. Identification of malaria hot spots for focused intervention in tribal state of India: a GIS based approach, *Int J Health Geographics*. 2009; 8: 30.

Githeko AK, John MA, Peter KO, Francis KA, Bryson AN, John IG and Guiyun Y., 2006. Topography and malaria transmission heterogeneity in western Kenya highlands: prospects for focal vector control *Malaria Journal*, 5:107.

Griffin JA, Graczyk TK, Geller N, Vittor AY., 2010. Effects of environment change on emerging parasitic diseases. *Int J Parasitol*; 30: 1395-405.

Jackson, AK. & Yan, G., 2010. Effects of microclimatic changes caused by land use and land cover on duration of gonotrophic cycles of Anopheles gambiae (Diptera: Culicidae) in western Kenya highlands. *Journal of Medical Entomology,* 42(6): 974–980.

Lindsay SW, Bodker R, Malima HA, Kisinzia W., 2000. Effect of 1997–98 El Niño on highland malaria in Tanzania. *Lancet*, 355, 989–990.

Minakawa N, Sonye G, Mogi M, Githeko A, Yan G., 2002. The effects of climatic factors on the distribution and abundance of malaria vectors in Kenya. *J Med Entomol 39:* 833–841.

Munga, N. Minakawa, G. Zhou et al.(2006), “Association between land cover and habitat productivity of malaria vectors in western Kenyan highlands,” American Journal of Tropical Medicine and Hygiene, vol. 74, no. 1, pp. 69–75, 2006.

Munyekenye, A. K. Githeko, G. Zhou, E. Mushinzimana, N. Minakawa, and G. Yan, (2005)“Plasmodium falciparum spatial analysis, western Kenya highlands,” Emerging Infectious Diseases, vol. 11, no. 10, pp. 1571–1577.

Nyika A, Kilama W, Chilengi R, Tangwa G, Tindana P, Ndebele P, Ikingura J.,2009. Composition, training needs and independence of ethics review committees across Africa: are the gate-keepers rising to the emerging challenges? J Med Ethics; 35(3): 189-193.

Snow, J and Marsh R., 2002. Averting a Malaria Disaster in Africa –Where does the Buck-up Stop? in the Bulletin of the World Health Organisation, pp.381-4.

Sweeney, J., 1997. Costs to households of seeking malaria care in the Kassena-Nankana District of Northern Ghana. In: Third MIM Pan-African Conference on Malaria, Arusha, Tanzania, 17-22 November 2002. Bethesda, MD, Multilateral Initiative on Malaria

WHO The World Health Report 2000. Health Systems: Improving performance. Geneva: World Health Organisation.

Worall (2004). Progress in Rolling Back Malaria in the African Region. Malaria, Liaison Bulletin of the Malaria Programme