**Kwame Nkrumah University of Science and Technology**


# A LOGISTIC REGRESSION MODEL ON SOME MALARIA INTERVENTION STRATEGIES . A CASE STUDY OF OBUASI MUNICIPALITY.

BY

RICHARD BOAKYE YIADOM

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FUFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF M.SC INDUSTRIAL MATHEMATICS

October 13, 2015

# Declaration

I hereby declare that apart from certain documentary and other authorities which I have cited and duly acknowledged, this submission is the result of my own research and that it has neither in whole nor in part been presented for another degree elsewhere. I also declare that neither my supervisor nor any other person but the author alone is responsible for whatsoever errors and omissions that might appear in the work.

| | | |
|---|---|---|
| Richard Boakye Yiadom | ..... | ..... |
| Student | Signature | Date |

Certified by:

| | | |
|---|---|---|
| Nana Kena Frempong | ..... | ..... |
| Supervisor | Signature | Date |

Certified by:

| | | |
|---|---|---|
| Prof. S. Amponsah | ..... | ..... |
| Head of Department | Signature | Date |

# Dedication

I dedicate this work to my dear wife Fausty and children Terry, Florentyna and Elijah for their love, sacrifice, prayers, support and encouragement.

# Abstract

Some medical experts and researchers have in recent times expressed interest in Malaria due to its life threatening nature and debilitating effect especially on pregnant women and children. This study seeks to model the risk of reduction in malaria reporting by children under five years in the Obuasi Municipality given the types of interventions.

Data gathered from a baseline survey across 22 clusters on a maximum of 508 sampled children is being modeled by means of logistic regression. The data was restricted to a 5-year span $(2009-2014)$ because of the vested interest in the under 5- year group. A paired t-test which compares the p-value of 0.00 to a threshold p-value of 0.05 was used to establish whether there is a significant difference between the two sets of data on clinical visits in the past and in the current year. The same test was used to establish that there was a mean reduction of 1.36 (with 1.53 standard deviation) in malaria reporting. Our analysis also provides some parameter estimates for our model. These are tested using Wald statistics and form the basis of our model equation. Again since our analysis of parameter estimates reveals that the parameter 1.2330 corresponding to the use of both IRS and ITN interventions only, has the least p-value of $0.0052 < 0.05$, we conclude that children who experience both interventions are likely to risk reduction in malaria reporting and therefore contribute significantly to model.

# Acknowledgements

I am extremely indebted to our Supreme God whose immense love, protection and care can never be thrown into oblivion.

My profound gratitude goes to my supervisor, Nana Kena Frempong for his commitment, patience, direction and concern in seeing the study to its conclusion. May the Almighty God richly bless him. I also wish to express my warmest gratitude to Dr. Philomena Nyarko, (Government Statistician), my director Mr. Anthony Amuzu and Mr. J.O. Okantey both of Ghana Statistical Service for their cooperation, words of encouragement and advice especially when it got to the data collection stage.

Mr. Salifu Abdullai (Statistical Service) , Richmond Owusu ( Past National service personnel, Statistical service, Obuasi), Emmanuel Boateng (Ghana Health Service, Obuasi) and Bright Atiase (Malaria control Department, AngloGold Ashanti) need a pat on their shoulders. They were very instrumental in my write up. I am equally grateful to Charles Kwame Amponsah(KNUST), Kachu Sadat and Isaac Cromwell (National Service Personel, Statistical Service) who assisted in typing the original script. Finally to Adwoa, Gloria, Kate and Jackson (past National Service Personnel of Statistical Service, Obuasi), I say Ayekoo for your support during my data collection.

To God is Glory.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# INTRODUCTION

## 1.1 Overview

This chapter focuses on the profile of the study area, background of the study, the problem statement, objective of the study and the methodology employed. It also captures the justification for the write up based on the factors and conditions which motivated the study and not without the approval of the Thesis topic and validation of the instrument used in data collection and the testing of the results.

## 1.2 Profile of Obuasi Municipal

The Obuasi Municipal is one of the thirty (30) districts of Ashanti Region with its capital as Obuasi. The Municipality was created as part of the government's effort to further decentralize governance. It was carved out of the then Adansi West District on the strength of Executive Instruments (E. I.) 15 of December, 2003 and Legislative Instrument (L. I.) 1795 of 17 March .The Municipality is located between latitudes $5°35N$ and $5°65N$, and longitudes $6°35'W$ and $6°90'W$. It covers a total land area of 162.4 square km. It is located in the Southern part of the Ashanti Region. It is $64km$ from Kumasi, the regional capital.According to the 2010 Population and Housing Census Results Obuasi Municipal has a total Population of 168,641 of which 81,015 are male and 87,626 are female. Out of the total 143,644 persons are living in the urban localities with 24,997 persons in the rural localities. Rocks in the Municipality are mostly of Tarkwain (Precambrian) and Upper Birimian formation which are noted for their rich mineral bearing potentials. Areas around Birimian and Tarkwain zones known as reefs are noted for gold deposits.

The municipality has an undulating terrain with more of the hills rising above 500 meters above sea level. Obuasi Municipal experiences semi-equatorial climatic conditions with a double maxima rainfall regime. Mean annual rainfall ranges between 1250 mm and 1750 mm. Temperatures are uniformly high all year with the hottest month being March when temperature of $30\circ C$ is usually recorded. Mean Average annual temperature is $25.5\circ C$. Relative Humidity is quite high (75% - 80%) in the wet season. The vegetation is predominantly a degraded semi-deciduous forest. The forest consists of limited species of hard wood, which are harvested as timber. The Municipality is administratively composed of the Municipal Assembly headed by the Municipal Chief Executive, five Zonal Councils and two constituencies namely Obuasi East and Obuasi West and 38 Electoral areas. The Municipality has three traditional authorities namely Fomena, Akrokerri and New Edubeasi. Each of the traditional authorities has specific stool lands under its jurisdiction. There are three main religious groupings in the Municipality, namely Christianity, Islam and Traditional Religion. According to 2010 Population and Housing Census Report Christianity dominates with 81.7 percent followed by Islam with 13.3 percent and the Traditional Religion with 0.2 percent. The Municipality has institutions providing education from pre-school to secondary level. There are 140 Kindergarten,163 primary, 96 JHS, 5SHS and 2 VOC/TECH/COM Schools (GES,Obuasi 2012/2013 .Mining and its related activities is the mainstay of the Municipal economy. Some of the other major industrial activities in the municipality are forest/wood based and related industries, blacksmithing and metal based industries, construction and quarrying based industries, mining and allied industries and agro-based industries. Agro-based industries notably oil palm and palm kernel extraction and gari processing ventures can also be found in the municipality. A broad spectrum of economic and financial services exists in the municipality to facilitate business activities. There are seven (7) reputable financial institutions, six (6) insurance companies and a number of micro credit institutions are in the municipality.The

Municipality can be demarcated into 3 agro-ecological zones each of which specific agric-programmes could be prescribed.

The land is suited for the cultivation of economic tree crops namely cocoa, coffee, oil palm, citrus as well as staple foods such as plantain, banana, cassava, yams, vegetables, pineapple, cocoyam, maize, seed production and crop trials. Other crops cultivated in certain parts of the municipality are rice and sugar cane. There are twenty-two (22) health facilities in the Municipality which consist of seven (7) hospitals, two (2) health centres, eight (8) clinics, four (4) maternity homes and one (1) CHPS centre.

## 1.3 Background of the Study

Malaria continues to be one of the life threatening ailments among society and is known to be the leading cause of death in Africa. It is responsible for over 1 million deaths and approximately, 4 in 10 deaths among children globally (Helena 2011).

The epidemiology of malaria in children is difficult to assess as most of clinical symptoms are non-specific and most of the cases occur in settings where no routine testing is available.

Malaria remains a leading cause of ill health. More than 40% of the world's population (approximately 3 billion people) is exposed to malaria in 108 endemic countries. It caused between 655 000 and 1.240.0001 deaths in 2010. Approximately 81% of malaria cases and 91% of malaria deaths occur in the African Region, where it remains one of the commonest causes of death and serious morbidity, especially for children and pregnant women; approximately 86% of malaria deaths globally are of children under 5 years of age. In fact children are at highest risk for severe disease and death between six months and five years of age: during this period children are most vulnerable as they have lost maternal immunity and they haven't yet developed specific immunity to infection. However this does not mean that younger infants are exempt from the death toll, the contrary is true

given the fact that in addition to the well known inoculum through the blood meal of an infected female anopheles and through infusion of infected blood products, neonates and young infants might also be vertically infected by plasmodia crossing the placenta (Richard-Fabian and Elena, 2012).

The malady is still the leading cause of OPD attendance in The Obuasi Municipality. According to Obuasi Ghana Health Service, The disease accounts for approximately 51% Clinical cases. Again the disease accounts for nearly 80% of sub-Saharan's 350-500 clinical cases reported for less than 5 years children(Unicef, 2007). Myriad of preventive and antidote interventions continue to come up from health researchers and other related organizations all in an attempt to reduce malaria prevalence. These institutions include WHO, UNICEF, Medical Research Institutions and Universities. Children continue to be the most vulnerable group across the globe and continue to record high numbers in mortality.

Malaria is a difficult disease to control largely due to the highly adaptable nature of the vector and parasites involved. While effective tools have been and will continue to be developed to combat malaria, invariably, over time the parasites and mosquitoes will evolve means to circumvent those tools if used in isolation or used ineffectively. To achieve sustainable control over malaria, healthcare professionals will need a combination of new approaches and tools and research will play a critical role in development of those next-generation strategies. The disease has a significant impact on the health of infants, young children and pregnant women worldwide. More than 800,000 African children under the age of five die of malaria each year. Malaria also contributes to malnutrition in children which indirectly causes the death of half of all children under age five worldwide. (National Institute of Allergy and Infectious Disease, 2011)

According to UNICEF's Ghana Facts Sheet, 2007, 3.5 million people contract malaria every year. Approx. 20,000 children die from Malaria every year (25 per cent of the deaths of children under the age of five). Even if a child survives, the consequences from severe malaria such as convulsions or brain dysfunction can

hamper long-term development and schooling. The annual economic burden of malaria is estimated 1-2 per cent of the Gross Domestic Product in Ghana.

In Ghana, malaria and acute respiratory infections (principally pneumonia) are the leading causes of fever in children. Malaria is responsible for the majority of childhood admissions and 22% of childhood deaths while pneumonia is also responsible for about 22% of hospital admissions in tertiary health facilities. Current control strategies for both malaria and pneumonia include early diagnosis and prompt treatment, particularly among those at risk of death and severe complications. Until recently when the Ministry of Health introduced community interventions, the official interventions for both illnesses were largely geared towards the formal health sector, even though many people (especially in rural areas) have little or no physical access to health facilities - i.e. live beyond 30 minutes travel time from facilities. The result is that many caregivers of children seek treatment from outside the home (usually non-hospitals sources) or self-treat. (Nonvignon et al, 2010).

From a study conducted in 14 localities around Agogo and Konongo in The Ashanti Region of Ghana) malaria is prevalent during the entire year and accounts for about 32-42% of all outpatient admissions and for major in-patient causes of death . Socio-demographic factors such as ethnic group, parent's education and occupation, use of protective measures, and living standard of the family are suggested to be important risk factors for malaria and malaria epidemics. The impact of socio-economic factors, namely the family's financial situation, is difficult to assess due to the lack of standardized economic data of income and tax. Additional socioeconomic factors assessed in the Demographic and Health Survey 2008 are marital status and religion.

Malaria has always been the most significant public health threat to the Obuasi community of Ghana .To deal with the spread of malaria, AngloGold Ashanti and Obuasi Municipal Assembly undertook to implement an Integrated Malaria Control Programme, focusing on Indoor Residual Spraying (IRS) in the Obuasi

municipality and its surrounding villages. The programme covered the entire Municipality. The total number of dwellings in the intervention area was 35000. Malaria, still a major health concern in the Obuasi municipality recorded in 2005, an average of 12,000 cases monthly. Forty-eight percent (48%) of all Out Patient attendants were due to malaria and the disease headed the top ten killers, being responsible for 22% of all deaths.There are consistent efforts to reduce malaria episodes which include chemical spraying, use of treated mosquito bed nets, clearing bushes, cleaning drains and subsidized treatments and yet prevalence rates and malaria incidence remain high. It is probable that the efforts to reduce malaria do not specifically take into account the risks factors likely to aggravate malaria disease. The high incidence of malaria cases among the age-structured population is unknown to the district, the season which recorded the highest despite the integrated malaria control programme is also unknown in the Obuasi Municipality.

Malaria remains an area of concern for AngloGold Ashanti's operations in Ghana, Guinea, Mali and Tanzania. Not only does the disease result in death, illness and absenteeism among employees, but it is a major cause of death in young children and pregnant women, which has an obvious impact on employees' families and communities.

There has been a dramatic decline in the number of cases of malaria in the past three years. From 79,237 cases in 2005, there were 53,070 in 2006, reducing to 20,976 in 2007 (a 73% decline from 2006 to 2007). Correspondingly, the malaria incidence rate (per 1,000 people) declined from 238 in 2005, to 164 in 2006 and 69 in 2007. The Malaria Lost Time Injury Frequency Rate (MLTIFR) *was 1,016.00 in 2005,634.74 in 2006 and 229.83 in 2007 (2006: 435) a 77% decrease in three years. These statistics demonstrate the success of the Malaria Control Programme, which was introduced at Obuasi in January 2006(Country report Ghana, Obuasi, 2007).

Malaria is a disease caused by a protozoon called plasmodium. Children under

five years and pregnant women are vulnerable to this disease in Ghana and Africa as a whole. The disease is transmitted to humans through the bite of a mosquito. The effects of malaria go beyond mortality and morbidity as malaria endemic areas suffer dearly in terms of human productivity and economic loss. AngloGold Ashanti is a mining company operating in Obuasi and in some other countries in Africa. It realized that malaria posed threats to its operations in East and West Africa. This was clearly shown in increased morbidity, mortality and absenteeism in the workforce as well as decrease in productivity (AngloGold Ashanti Report to Society 2007).

The situation was not different in the company's operational area in Obuasi. In April 2004, the company realized that an estimated average of 11,000 malaria cases per month was recorded according to the municipal health authority. An additional 6,800 cases were reported by the mine medical service. At any point in time, 20% of the workforce had malaria and the average time off work due to malaria ailment was between two and three days. Extending the trend to the whole Obuasi community clearly indicated that the disease had gained ground and therefore called for actions to be taken to reverse the increasing trend of the disease (AngloGold Report to Society 2004).

For the purpose of the study Insecticide Treated Nets (ITNs) and Indoor Residual Spraying (IRS), the major interventions executed by the malaria control programme in Obuasi shall be defined.

Insecticides treated nets(ITNs) are nets that have been dipped into mosquito insecticides within the last 12 months and long lasting insecticides nets are nets that have been permanently treated with insecticides that lasts for the useful life of a mosquito net, defined as at least 20 washes and at least three (3) years of field condition (Unicef, 2007)

Indoor Residual Spraying (IRS) involves applying a long lasting insecticide to the inside of houses and other structures to kill mosquitoes on interior walls (Unicef, 2007)

Figure 1.1: A spray operator at the Obuasi Malaria Control Centre.

Here he is spraying a home in the Bongobiri suburb of Obuasi.

Source: Anglogold Ashanti Website, May 7, 2013.

## 1.4 Problem Statement

Obuasi Municipality, a mining community in the Ashanti Region of Ghana is one of the endemic environments. Due to its cosmopolitan nature, the Municipality invariably has its fair share of environmental hazards. The malaria control programme , an Indoor Residual Spraying Exercise, happens to be the Brainchild of Obuasi AngloGold Ashanti. It was initiated in 2006 as part of their cooperate responsibility to the indigenous of Obuasi. Again it was equally purported to augment Obuasi Ghana Health Service' Insecticide Treatment Nets (ITNs) distribution exercise.

The Obuasi Ghana Health Service reports waning trend in OPD malaria cases in the particular instance of under 5 children. An average decline of 5800 cases per month is recorded since 2006. The veracity of this feat is being studied for the

last 5 years. Thus whether the intervention strategies namely the use of Insecticide treated nets or Indoor residual Spraying or Both have indeed contributed to a significant reduction of malaria cases for children under 5 years.

## 1.5 Objectives

The study seeks to:

1. To establish the relationship between the intervention strategies and the incidence of malaria.

2. To test the difference in clinical visits in the past and current years, given the types of malaria interventions, by children under 5 years.

## 1.6 Methodology

The data capture is limited to the last five years (2009-2014) because of Target group concerned that is children under 5 years. Mothers of such children have been the major respondents. The data is collected from primary source through a baseline household survey with an unbiased sample between 450 and 500 children across 22 clusters in the Obuasi Municipality. For this study a cluster is one of the 22 carefully selected defined geographical areas (suburbs or localities) across which our samples are selected proportional to the population size of that locality. A major statistical tool that will be used in the data analysis and processing is SAS and the study is modeled by logistic regression where the individual intervention tools are modeled against Risk of reduction in malaria cases.

In statistics, logistic regression or logit regression is a type of probabilistic statistical classification model. It is also used to predict a binary response from a binary predictor used for predicting the outcome of a categorical dependent variable (i.e., a class label) based on one or more predictor variables (features). That is, it is used in estimating empirical values of the parameters in a qualitative

response model. The probabilities describing the possible outcomes of a single trial are modeled, as a function of the explanatory (predictor) variables, using a logistic function. Frequently "logistic regression" is used to refer specifically to the problem in which the dependent variable is binary-that is, the number of available categories is two-and problems with more than two categories are referred to as multinomial logistic regression or, if the multiple categories are ordered, as ordered logistic regression. Logistic regression measures the relationship between a categorical dependent variable and one or more independent variables, which are usually (but not necessarily) continuous by using probability scores as the predicted values of the dependent variable. As such it treats the same set of problems as does probit regression using similar techniques

In this write-up, references are made from the internet, Ghana Health Service library, The University library and the Ghana Statistical Service library among others.

## 1.7 Justification

In Ghana Up to 30% deaths in children less than 5 years are attributable to malaria (MOH,2008) Again Boateng,2010 reports that Under 5 mortality is 80/1000 live births. The National Malaria control programme a subsidiary of Anglogold Ashanti has been the major initiator of Indoor residual spraying and has been collaborating with Zoomlion company limited in spraying stands (houses) and structures (rooms). It has also on one time or the other carried out the distribution of Insecticide treated nets in conjunction with Ghana Health Service, Obuasi all in an attempt to combat the debilitating and life threatening malady. Two rounds of spraying is done each year since the inception of The programme from February to April and July to December. The malaria control programme has been expanded to cover other Anglogold operational areas particularly districts that share immediate boarders with Obuasi Municipality. There is the evidence of the people's continued reliance on available health facilities for malaria treat-

ment in Obuasi. Besides Obuasi Health Service reports waning trends in OPD malaria cases in the particular instance of children under five years old and Anglogold Ashanti reports of an average decline of 5800 cases per month. According to Steve Knowles (AngloGold's Malaria control programme Director), 2012, there is an average monthly decline in cost of medication from $ 55,000.00 (2005) to $ 510.00 (2012).Thus malaria control interventions has not only reduced the burden of malaria in the community, increased school attendance and won the gratitude of the Obuasi community, but has also reduced absenteeism at the mine, increased productivity and reduced the cost of malaria medication to employees and dependants. Consequently the need to assess the veracity of the foregoing feats and the efficacy of the various interventions will not be out of place. It is envisaged that the write-up will not only be consumed by the academics but the larger research community. The study will augment related studies carried out by the Malaria Control Department of AngloGold Ashanti and individual researchers. Again it is expected that child mortality figures in Obuasi Municipal reduces if indeed the study reveals a reduction in clinical cases. The document will also be useful to the Obuasi Municipal Assembly particularly it's environmental and sanitation Department in addressing the environmental challenges faced by the Assembly. The recommendations there- in if implemented shall go a long way to ameliorate the health, social and economic status of the indigenous of Obuasi.

## 1.8 Limitation and Scope of Study

The study was intently restricted to the under five years' group since that class is perceived to be vulnerable. Also due to budgetary constraints the sample size has been restricted to not more than five hundred households but with an unbiased spread among 22 clusters. Again the study has been confined within the boundaries of the municipality considering the extent of the IRS coverage. In the absence of existing data from a baseline survey, I had to fall on secondary data from Edwin Cade memorial Hospital (AGA) and Obuasi Government Hospital

to support my work.

## 1.9 Thesis Organisation

The Thesis has been carefully organized in a sequel of chapters as follows.
The chapters are preceded by a list of Appendices and Table of Contents. Chapter one which is the introduction comprises the background of the study, statement of the problem, purpose or objective of the study, methodology used and justification of Thesis Topic or problem statement. Chapter two which is the literature review of previous works of a number of authors focuses on the computational and analytical part of the document with about three abstract summaries on a page.
Chapter three which elaborates on the methodology employed in the study deals with formulations, models and variants, research design, procedures and instrumentation. Chapter four deals with data collection, analysis and results. Finally, chapter five deals with conclusions and recommendation for researchers and policy makers.

# Chapter 2

# LITERATURE REVIEW

## 2.1 Introduction

In this chapter, the related literature on the current study has been reviewed. The review constitutes the computational and analytical aspects of their study which is captured in their abstract documents and primarily modelled by logistic regression.

The intervention effectiveness experienced by children under the age of 5 (in Sub-Saharan Africa) exposed to both insecticide treated nets and the indoor residual spraying compared to each intervention alone based on nationally representative sample collected from 17 communities has been assesed. Living in households with both Insecticide Treated Nets (ITNs) and Indoor Residual Spraying (IRS) was associated with a significant risk reduction against Parasitaemia in medium and high transmission areas, 53% and 31% respectively. For low and high transmission areas, having both ITN and IRS was not significantly more protective against the parasitaemia.In rural and the urban settings exposure to both interventions provided significant protection against parasitaemia, 57% and 39% respectively. The effect was significantly greater than having singular intervention. According to Fullman et al,the findings suggest greater reductions in malaria morbidity and health gains for children may be achieved with ITNs and IRS combined beyond the protection offered by IRS or ITNs alone.

Again a related study asseses the effect of insecticide-treated nets and indoor residual spraying for malaria control in the rural 'Koboles' of Adam District, south central Ethiopia. According to them Indoor residual spraying (IRS) and Insecticide-treated nets has been the main tool used to control malaria. The study

was purported to examine the effect of IRS and ITS control strategies in Anono shisho kebele compared with Kamo Gerbi which was supplied with ITN only and the Jela Aluto (no IRS and ITNs) with regard to the prevalence of malaria and mosquito density. Data collected on parasitological and knowledge, attitude and practice surveys were managed and analyzed using SPSS 13.0.A $p-value < 0.05$ was considered statistically significant. According to Damtew et al the overall prevalence of malaria was 8.6% in Jela Aluto, 4.4% in Kamo Gerbi and 1.3% in Anono Shisho in two seasons surveys. The difference in overall malaria prevalence and mosquito density between the three Kelebes was significant ($p < 0.05$). The study has provided some evidence for the success of ITNs or IRS combined malaria control measures in Anono Shiho Kebele in Adam Tulu District.

Gabriel et al. (2013) analyses the effect of malaria prevalence and the indoor spraying on the probability of sleeping under an insecticide-treated bed net in nine sub-Saharan countries .Their study examines specifically the responsiveness of insecticides treated nets usage to malaria prevalence and also whether indoor residual spraying crowds out usage of net.

According to Gabriel and colleagues, bed net usage elasticity to malaria prevalence ranges from 0.42 for adult women to 0.59 for older women .Thus by their model they demonstrate that malaria prevalence has positive effect on usage of bed net but that net usage is inelastic with respect to malaria prevalence. They further suggest that indoor residual spraying(IRS) does not "crowd out" sleeping under Insecticide treated nets (ITNs) for any of the population group studied and that households consider IRS and ITN usage as complements.

Helena 2011, in her Dissertation write up towards the award of a Master of Science Degree by the university of Science and Technology(Dept of social and clinical pharmacy) outlines the use of bed nets by mothers of under 5 children in the Pediatrics Out-Patients Department (P.O.P.D.) of the 37 Military Hospital against the background of mothers. According to Helena over 3.2 billion deaths are threatened globally by malaria and more than one million malaria related deaths

are recorded annually. The disease accounts for 44% clinical (OPD) attendance and 22% of mortality among children under 5 years .Helena further indicates that insecticide treated nets are becoming increasingly available to vulnerable populations. Nonetheless the use of insecticide treated nets (ITNs) does not measure up to ownership. According to her findings from the Rollback Malaria(RBM) initiated by WHO,UNICEF,UNDP and World Bank in one of their recent researches suggests that ITNs can reduce malaria prevalence by 48-50% and can prevent nearly 7% under 5 mortality globally.

According to Unicef (2007) Rollback Malaria Program (RMP), 3 billion people (almost 50% of world's population) live in malaria endemic areas. The disease is prevalent in 107 countries as well as tropical and subtropical regions and the hardest hit is sub-Saharan Africa. The report further suggests that 350-500 million clinical malaria cases occur each year culminating in nearly 1 million deaths. Of this number 8 in 10 are children under age 5.It assess available key interventions that reduce malaria burden with particular emphasis on the progress across sub-Saharan Africa. According to UNICEF, the use of insecticide treated nets (ITN) is gaining coverage in the Sub-Saharan region and immense progress has been made. The use of ITNs among children under 5 is at least tripling coverage with between 16-20 countries. Notwithstanding these gains, global target are yet to be met.

According to Netforlife's Episcopal Relief and Development program(2011),there are an estimated 216 million cases of malaria each year, resulting in nearly 660,000 deaths, most of which are under 5 years old children. Long lasting insecticide treated nets simultaneously provide a protective covering for the body while releasing chemicals to useful repel and kill the infection carrying mosquitoes. The net in recent times has been effective for children under 5 years because latest technology has dramatically improved their efficacy infusing the insecticide in the netting material. The report further suggests that scientific evidence has validated the safety of indoor residual spraying (IRS).It defines IRS as the appli-

cation of long lasting insecticides (including DDT) on walls and roofs of houses, public buildings and domestic animal shelters in order to kill malaria-carrying mosquitoes that settle on these surfaces .The report warns that the spraying is ineffective at specific places and must not be encouraged.

Morel et al assess cost effectiveness analysis of strategies to combat malaria in the Developing countries .The objective was to determine the cost effectiveness of selected malaria control interventions in the context of attaining the millennium development goals for malaria. Two highly endemic Sub-Saharan African regions predominantly Southern and Eastern African (African-E) and Western Africa (Africa-D) which high child and adult mortality is recorded were selected for a case study. Under 5 malaria incidence measures 1436 per 1000 in Africa-D while it is 1184 per 1000 for Africa-E. Cost were assessed in year 2000 international dollars against effects as disability adjusted life years averted by 10-year implementation plan. According to Jeremy et al(2005) limited number of preventive interventions include insecticide treated nets (ITNs) and Indoor residual spraying(IRS).By their findings, the study quantifies the gains of shifting resources towards artemisinin based combination treatment besides using preventive and curative interventions in tandem.

Malaria prevalence remains high in many African countries despite massive scaling -up of insecticide treated nets (ITNs) and indoor Residual Spraying (IRS).Data collected on IRS and ITN coverage age group 0.5-14years in two separate surveys reveals plasmodium falciparum infection prevalence was 9.3% and 22.8% respectively, Philipa et al (2013).Two main transmission periods which occur after the short and long raining seasons were considered in the two surveys which covered 5,152 and 4,325 children aged 0.5-14years.

An estimated 17% global reduction in malaria incidence has been achieved between 2000 -2010, nonetheless, 174 million malaria episodes are recorded in 2010 for only Africa. In Africa 53% of household owns ITN and 11% of population at risk use IRS, (WHO world malaria report, 2012).Thirty-one (72%) endemic

countries in Africa were reported to use both IRS and ITN in at least some specific areas in 2010.

The effectiveness efficacy and cost-effectiveness of indoor residual house-spraying (IRS) and insecticide treated nets (ITNs) against plasmodium falciparum considered as part of malaria control in the highlands of western Kenya. A sample of 590 households comprising 200 with no vector control ,200 with IRS and 190 with ITNs were selected with the ages of 0.5-4years,5-15years and $>$ 15years as the target groups.(Guyatt et al,2002).Plasmodium falcum prevalence among household members not protected by either IRS or ITN was 13%.ITN usage reduced risk of infection by 63% and usage of IRS reduced infection by 75%.The economic costs for the IRS and ITN interventions were respectively US$9 and US$29.

Studies over the past two decades suggest that ITNs could substantially reduce the burden caused by plasmodium falciparium in Africa. ITNs are currently widely promoted as a means of preventing man-vector contact in malaria control and form the basis of disease control within the recently launched roll back malaria initiative. The role and cost effectiveness of indoor residual-house spraying (IRS) for malaria control has received little attention. A comparison of historical trials of IRS in Africa against contemporary evidence of ITN effects emphasized the need to re-visit the comparative advantage of the two interventions under a variety of endemic environment.(Antis & Mnzava, 2000).

Both Insecticides Treated Bed Nets (ITNs) and Indoor Residual Spraying (IRS) reduce malaria in high transmission areas. The combined effect of these interventions is unknown. Malaria continues to be a leading cause of morbidity and mortality in Africa. Over 247 million cases of malaria and nearly 1 million deaths were recorded in 2008. ITN reduced malaria morbidity and all causes of malaria mortality across a variety of transmission settings. Though IRS also reduces malaria morbidity and mortality, the intervention has relegated to seasonal transmission or epidemic prone areas due to logistical complexity and expensive nature of spray comparison (Hamel et al., 2011). A non-randomized prospective

cohort study was conducted to determine protective efficacy (P.E) of IRS with ITN (IRS+ITN) compared with ITNs alone in preventing plasmodium falciparium parasitemia. At base line, participants provided blood samples for malaria smears, were presumptively treated for malaria and received ITNs. Blood smears were made monthly and at sick visits.

Transparent evidence based on the cost and cost effectiveness of malaria control interventions is provided to inform rational resource allocation by donors and domestic health budgets and the selection of optimal packages of interventions of malaria control programs. Despite being a largely preventable and treatable disease, malaria is responsible for an estimated 800,000 deaths globally each year with a majority of morbidity and mortality occurring in children in sub-Saharan Africa. Besides its impact on health, malaria imposes heavy economic burden on individuals and entire economies. In response to calls for widespread control and elimination of malaria and the challenge of meeting the Millennium Development Goals, there has been a rapid scale-up of existing effective anti-malaria interventions, in particular ITNs. There is a wide range of malaria control interventions whose efficacy and effectiveness have been repeatedly demonstrated over many years. These include ITNs and IRS interventions (White et al., 2011).

Long term success of ongoing malaria control effort based on mosquito bed nets (long lasting insecticide nets) and Indoor Residual Spraying is dependent on continues monitoring of mosquito vectors and thus effective mosquito sample tools. With the low vector densities in coastal Kenya and across much of sub-Saharan Africa wherever malaria interventions, long lasting insecticidal nets (LLITN) and / or Indoor Residual Spraying (IRS) are in place the use of a single collection method will not be sufficient to achieve a representative sample of mosquito population structure Onyango et al. (2013).

Participatory health education intervention contributes to decreased malaria prevalence among children. It has a positive impact not only on school children but also on community adults through the improvement of knowledge and practices. An

estimated 90% of the world's malaria-attributable deaths occur in sub-Saharan Africa. In Ghana malaria account for more than 44% of reported clinical cases and those 2 in 10 children under age 5 die of malaria. Malaria Control Intervention in Ghana target pregnant women and children under 5 years as this group is most vulnerable. Major activities involved in the program are the distribution of Insecticide Treated Nets (ITNs) during antenatal care and the provision of intermittent preventive treatment (IPT) (Ayi et al., 2010).

Improving the health of school-aged children can yield substantial benefit for cognitive development and educational achievement. However, there is limited experimental evidence of the benefit of alternative school based malaria interventions or how the impacts of interventions vary according to intensity of malaria transmission. In many malaria endemic countries, successful control programs have recently reduced the level of malaria transmission, and as a consequence, immunity to malaria is acquired more slowly and the burden of clinical malaria is stiffing from the very young to older children (Haliday et al., 2014).

Insecticide-treated mosquito net (ITNs) and Indoor Residual Spraying (IRS) are recommended strategies for preventing malaria in children. While their impact on all-cause child mortality is well documented, their impact on reducing malaria-attributable mortality has not been quantified. Two systematic literature reviews in plasmodium falciparium endemic settings: one to estimate the effect of ITNs and IRS on preventing malaria-attributable mortality in children 1-59 months and another to estimate the effect of ITN and intermittent preventable therapy (IPTp) on preventing neonatal and child through improvement in birth outcomes. The protective efficacy (PE) of ITNs and IRS on reducing malaria-attributable mortality 1-59 months is estimated to be 55% with a range of 49-51% (Eisele et al., 2010).

According to Moorthy et al. (2009) in their methodological review of clinical trials to estimate the efficacy of preventive interventions against malaria in pediatric populations, recent years have seen publications of a considerable number of clin-

ical trials of preventive interventions against clinical malaria in children. There have been variability in the specification of end-point, case definitions, analysis methods and reporting and the relative lack of standardization complicates the ability to make comparative evaluation between trials. Control trials of preventive interventions against children in endemic countries were identified in which clinical malaria or death had been one of the main points. Trials were included that evaluated the impact of vaccines, Insecticide Treated Nets (ITN), Intermittent Presumptive or Preventive Therapy (IPT) in infants. These were aimed at preparing for a WHO consultation on design issues in malaria vaccine trials. 29 controlled trials of preventive malaria interventions were identified of which eight were vaccine trials. Vaccine trials that were designed to detect an effect on clinical malaria all reported the incidence rate of first episodes of clinical malaria of their end point. Only one trial of preventive intervention (of ITN) was identified that was designed to detect an effect on severe malaria. The development and deployment of new and improved intervention methods for malaria control shows promising signs of reducing significantly the global burden of malaria. Control trials still remain essential for the rigorous assessment of the potential impact of new tools and strategies to reduce morbidity and mortality caused by malaria.

In their systematic review for the Lives Saved Tools (LiST) on protective efficacy of malaria case management and intermittent preventive treatment for preventing malaria mortality in children, Thwin et al, 2011 reveals that the LiST model was developed to estimate the impact of the scale-up of child Survival interventions on child mortality. Systematic literature reviews of published studios in P. falciparum endemic settings to determine the protective efficacy (PE) of ATI treatment against malaria deaths among children with uncomplicated malaria. In lieu of randomized placebo-controlled trials of malaria treatment, multiple data sources to ascertain estimates of Protective Efficacy (PE) including a previously performed Delphi estimate for treatment of uncomplicated malaria. The Protective Efficacy of ACT treatment of uncomplicated P. falciparium malaria

with effective case management including intravenous quinine on reducing malaria mortality in children 1- 59 months is estimated to be 82% (range: 63-94%) compared to no treatment.

In their global digital fundraising comparing to help eliminate malaria deaths, Novartis through their "power of one" campaign aims to address the malaria treatment gap through direct donations and existing government Commitments (Novartis, 2013). Malaria is a preventable and treatable disease but it still kills a child every minute. It is estimated that over 300 million additional treatments will be needed to treat malaria patients across Africa between now and the end of 2015.

According to AngloGold Ashanti sustainability Review, 2009, the reduction of malaria in the community and mine made good economic sense besides benefiting employees, their families and communities. A malaria control program is the set example of a sustainable corporate social responsibility program for a company operating in a malaria endemic area, a win/win for company and community. You only have to superimpose a map of our global operations over the world malaria map to see the need for an overall group malaria strategy (Steve Knowles, AngloGold, 2009). At Obuasi the program has not only reduced the burden of malaria in the community, increased school attendance and won the gratitude of the community, but has also reduced absenteeism at the mine, increased productivity and reduced the cost malaria medication to employees and dependants.

In analyzing the effect of pyrethroid resistance on the cost effectiveness of mass distribution of long- lasting insecticidal nets, Hardy et al, 2014 indicates that the effectiveness of insecticides- treated nets in preventing malaria is threatened by developing resistance against pyrethroids and that little is known about how strongly pyrethroid resistance affects the effectiveness of vector control programs. In their analysis, data from experimental hut studies on the effect of long lasting, insecticidal nets (LLINs) on mine anopheline mosquito populations, with varying

levels of mortality in World Health Organization susceptibility tests, were used to parameterize models.

IRS is an effective strategy for reducing malaria incidence. It is about as effective as using ITNs though the later may be more effective at reducing morbidity in some situations (Pluess et al., 2010). In assessing the cost effectiveness and efficacy of IRS The review identifies that a few studies have directly compared the cost effectiveness of IRS directly with other methods of malaria control. According to Pluess et al. (2010) a study from 2008 assessed the cost effectiveness of seven African anti malaria campaigns: two IRS campaigns and five insecticides treated bed nets (ITN) distribution campaigns. On a cost-per-child-death averted basis all interventions were about the same but the ITN campaigns were slightly more Cost effective.

During the last decade, pyrethroid-treated mosquito nets have become the main method of malaria prevention in many malaria-endemic African countries. In a few notable exceptions, usually those with a more developed health infrastructure such as South Africa, a long standing practice of applying Indoor residual spraying (IRS) has been successful. The two approaches to malaria prevention, Insecticide Treated Nets (ITNs) and spraying(IRS) are not mutually exclusive and in malaria-endemic areas where ITN coverage is still limited, the feasibility of introducing IRS to reduce transmission is being considered (N'Guessan et al., 2007).

Spraying houses with insecticides (Indoor Residual Spraying) to kill mosquitoes is one of the main methods that have been used to control malaria on a larger scale: IRS has helped to eliminate malaria from great parts of Asia, Russia, Europe, Latin America and successful IRS programmes have also been run in parts of Africa. Another successful method of mosquito control relies on the use of physical barriers such as bednets or curtains that can also be sprayed with insecticides. (Insecticide Treated Nets; ITN). In this review Pluess et al, 2010 compares the health benefits and efficacy of IRS and ITN. Primarily malaria prevention on a

large scale depends on two vector control interventions: Indoor Residual Spraying (IRS) and Insecticide Treated (mosquito) Nets(ITNs). Historically IRS has reduced malaria transmission in many settings in the world but the health effects have never been properly quantified. Cluster randomized control trials (RTCs), controlled before-and-after studies (CBA) and interrupted time series (ITS) of IRS compared to no IRS or ITNs are some selection criteria considered.

Vector control is the key intervention for global malaria control and elimination efforts. It is critical for the reduction and ultimately, for the interruption of malaria transmission. Currently, the two most common vector control interventions are long-lasting insecticidal nets (LLINs) and Indoor Residual Spraying (IRS). Together these account for almost 60% of global investment in malaria control. The number of LLINs delivered by manufacturers has increased dramatically in recent years, rising from 5.6 million in 2004 to 145 million in 2010 in sub-Saharan Africa. The number of people protected by IRS in the WHO African Region increased from 10 million in 2005 to 78 million in 2010, representing 6% of the global population at risk (WHO, 2013).

Malaria control has been dramatically scaled up the past decade, mainly due to increasing international donor funding since 2003.Assessments used domestic malaria financing reported by national programmes and global fund data on donor financing for 90 endemic low and middle-income countries, WHO estimates of households owning one or more insecticide-treated mosquito net (ITN) for countries in sub-Saharan Africa and WHO-estimated malaria case incidence and deaths in countries outside Sub-Saharan Africa. ITN coverage in 2010 in Africa and declines in case and death rates per person at risk over 2004 to 2010 outside Africa were greatest in countries with highest donor funding per person at risk and smallest in countries with lowest donor malaria financing per person at risk. Associations between programme financing per person at risk and ITN coverage increases and declines in case and death rates suggest opportunity to maximize the impact of donor funding, by strategic re-allocation to countries with

highest and continued need (Korenromp et al., 2013).

As successful malaria control programmes re-orientate towards elimination, the identification of transmission foci, targeting of attack measures to high risk areas and management of importation risk become high priorities. Using the Namibia Example a method of targeting of interventions using surveillance data satellite imagery, and mobile phone call records to support elimination planning is described. One year of aggregated movement patterns for over a million people across Namibia are analyzed and linked with case- based risk maps built on satellite imagery. By combining case-data and movement, the way human population movements connect transmission risk areas is demonstrated. Communities that were strongly connected by relatively higher levels of movement were then identified and net export and import of travelers and infection risks by region were quantified (Tatem, 2014).

The review highlights malaria events, achievements and scaling up impact with scientific evidence in moving malaria from epidemic status towards sustained control and elimination from 1960-2011. The unprecedented and substantial reduction in malaria incidence and consequently mortality rates, at varied degrees across African countries and People's Republic of China (P.R. China) are very encouraging although the gains are still fragile. There has been improvement in the health situation in most African countries since 1960 and malaria in particular is decreasing over time in South Saharan Africa where the global burden of disease is significantly approximately 90% and People's Republic of China accounts for less than 10% (Tambo et al., 2012).

In an attempt to determine whether scaling up of malaria control by combining Indoor Residual Spraying (IRS) and long-lasting Insecticidal nets (LLIN), enhance protection to population, Kleinschmidt et al. (2009) presents results from a literature search and household surveys in Bioko, Equatorial Guinea, Zambezia and Mozambique. Five out of eight previous studies reported a reduced risk of infection in those protected by both interventions compared with one interven-

tion alone. Surveys in Bioko and Zambezia showed strong evidence a protective effect of IRS combined with nets relative to IRS alone. Odds ratio ($OR$) = 0.71 for Bioko and $OR = 0.63$ for Zambezia. The effect of both interventions combined compared with those who had neither $OR = 0.46$ in Bioko and 0.34 in Zambezia. Although the effect of confounding cannot be excluded, these results provide encouragement that the additional resources for combining IRS and LLIN are justified.

Insecticide Treated Nets (ITNs) and Indoor Residual Spraying of houses provide effective malaria transmission control. There is conflicting evidence about whether it is more beneficial to provide both interventions in combination. A cluster randomized control trial was conducted to investigate the combination provides additional protection compared to ITN alone. Fifty clusters in Northwest Tanzania were randomly allocated to ITNs only or ITNs and IRS. Dwellings in the $ITN+IRS$ arm were sprayed with two rounds of bendiocarb in 2012. Plasmodium falciparium prevalence rate (P/PR) in children 0.5-14 years old (primary outcome) were compared using three cross- sectional surveys in 2012. IRS coverage was 90%. ITN use ranged from 36%-50%. P/PR was 13% in the ITN+IRS arm and 26% in the ITN only arm. Odds $ratio = 0.43$ (West et al., 2014).

## 2.2 Modeling on Malaria Interventions

De Oliveira et al. (2013), examine the use of Geographic information systems (GIS) analysis and logistic regression as a tool to identify the relative likelihood and its socio environmental Determinants of malaria infection in the vale do Amanhecer rural settlement, Brazil.

The model that represents the likelihood of infection by malaria at the Vale do Amanhecer settlement was generated based on Logistic Regression with the 'backward conditional stepwise' procedure, comparing cases/non-cases with multiple explanatory variables. It was opted for a dichotomous modeling approach, as no reliable statics on the total number of inhabitants domicile could be obtained,

which would a pre-requisite of bias-free estimate of absolute cases or its probability (eg. by Poisson regression or generalized linear models). Model performance for different cut-off values was assessed by its sensitivity and specificity (ROC curve). In logistic regression, the canonical link functions (logits) for the binomial distribution, of the unknown binomial probabilities are modelled as a linear function of the risk factors ($x_i$):

$g(P_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_i x_i$

In which:

$g(P_i)$ = link function

$P_i$ = likelihood of response for the $-ith$ factor (or covariate)

$\beta_0$ = intercept

$\beta_i$ = coefficient

$x_i$ = independent variables

Logistic regression outcomes such as the Wald statistics, significance levels of variable coefficients and overall classification accuracy were used to test the importance of the environmental factors for the occurrence of malaria cases and for the development of a risk model. Using the stepwise, backward conditional method, as implemented in SPSS 15 (SPSS Inc.), only variables with significance higher than 95% ($p < 0.05$) were maintained in the final model.

In 2005, the Vale do Amanhecer settlement had 718 inhabitants, of which 394 (54.87%) were men and 324 (45.13%) were women. In the settlement, 359 cases of malaria were notified, distributed in 200 domiciles, the city of Juruena presented 720 positive smears for malaria in 2004. This corresponded to an API of 116.8 per 1.000 inhabitants, which represented an increase of 284.9% in the incidence of positive slides when compared to the 2003 API of 41.0 per 1,000 inhabitants. Five cases were excluded because the notification forms of SIVEP-malaria did not inform their dwelling places and another 18 because they belonged to domiciles outside the settlement limits. The 7,200 ha reserved for land owners and another 7,200 ha designated to permanent environmental reserve was considered

the official limit area of the settlement. Out of the total 336 cases of malaria, 133 positive slides were from dwellers at Road 08, which corresponds to 37.60% of the notifications. As for Roads 13 and 5, 124 cases were notified (35, 10%) and 58 (16, 40%) cases of malaria were notified, respectively .The logistic regression model performance for a cut-value of at least one case per domicile show a poor overall performance (65.4%), very low sensitivity (0.39) and percentage of explained variance (Nagelkerke R square = 0.22), as only VSHI and mining area distance are found to be significant predictors . A cut-value of 2 strongly improves the logistic regression model, increasing overall performance to 74.5%, sensitivity to 0.79 and Nagelkerke R square to 0.46. For higher cut-offs, both overall performance and specificity sharply decrease. This model for a cut-off value of 2 included as significant ($p < 0.05$) the variables Land Use, VSHI, and NDVI; moreover, it included as highly significant ($p < 0.01$) the variable mining area distance.

The Wald statistics underpin the high importance of this variable (26.4) as predictor. Consequently, areas with highest likelihoods of malaria infection are located in the southern part of the settlement, where mining activities are concentrated (negative variable coefficient). In the model these situations are related. The highest relative likelihoods however are only obtained if an area presents intense use and occupation, high level of wetness (positive coefficients) and low NDVIs, indicating little remaining vegetation. Additional hot-spots of elevated risks occur in the mid-western region of the settlement where secondary predictors such as VSHI, NDVI and land use account for elevated risks, but where the density of mining areas is lower. Razo et al,2012 models risk among children in Cote d'ivoire.

Binomial regression models were fitted in STATA/IC version 10.1 (StataCorp LP; College Station, TX, USA) to assess the relation between ecological predictors and Plasmodium spp. prevalence. Significant ecological factors, based on likelihood ratio test (LRT) with significance levels of 15%, were included as covariates

in further analyses. Bayesian non-spatial and geo-statistical logistic regression models were fitted in Open BUGS version 3.0.3 (Imperial College and Medical Research Council; London, UK). Spatial dependency was modeled assuming stationary (i.e. spatial correlation was modeled as a function of distance between locations only), as well as non-stationary (i.e. spatial correlation was modeled as a function of distance between locations and position within the study area) latent spatial processes.

Let $N_i$ be the number of children tested at location si ($i = 1, ..., n$) and $Y_i$ the number of those found with Plasmodium parasites in a blood sample. It was assumed that $Y_i$ arises from a binomial distribution, that is $Y_i$ $Bin(N_i, p_i)$, with $p_i$ measuring malaria risk at location $s_i$. The relation between the malaria risk and the m associated environmental covariates $X_i$ at location $s_i$, $X_i = (X_{i1}, X_{i2}...X_{im})^T$, was modeled via the logistic regression $logit(p_i) = X_i^T\beta$, where $\beta = (\beta_1, \beta_2, ..., \beta_p)^T$ are the regression coefficients. Exchangeable random effects $\epsilon_i$ were added on the logit scale, such as $logit(p_i) = X_i^T\beta + \epsilon_i$. Spatial correlation was introduced on location-specific random effect parameters $\varphi_i$, that is $logit(p_i) = X_i^T\beta + \phi_i$, assuming that $\phi_i = (\varphi_1, \varphi_2, ..., \varphi_n)^T \sim MVN(0, \xi)$ with variance-covariance matrix $\xi$. It was further assumed that spatial process is isotropic and decays exponentially with distance, i.e. $\xi_{ij} = \sigma^2 exp(-\rho d_{ij})$, where $d_{ij}$ is the Euclidean distance between villages $s_i$ and $s_j$; $\sigma^2$ is the geographic variability known as sill, and $\rho$ is a smoothing parameter that controls the rate of correlation decay with increasing distance. The spatial range is defined as the minimum distance at which spatial correlation between locations is below 5%, and is calculated as $3/\rho$ for the exponential correlation structure.

To take into account non-stationarity, the study area was partitioned into three ecological sub-regions ($K = 3$), assuming local independent stationary spatial processes $\omega_k = (\omega_{k1}\omega_{k2}, ..., \omega_{kN})^T$ in each ecological sub-region ($k = 1, ..., K$). The spatial processes were assumed to be multi-variate normally distributed, $\omega_k \sim MVN(0, \xi_k)$, with variance-covariance matrixes $\xi_k$ defined by $(\xi_k)_{ij} =$

$\sigma_k^2 exp(\rho_k dij)$. It was further considered that the spatial correlation $\varphi_i$ at location $s_i$ in the study area is a mixture of the independent spatial processes modeled as weighed average, such as $\varphi_i = \sum Kk = 1 a_{ik}\omega_{ik}$, where the weights $aik$ are decreasing functions of the distance between location si and the centroids of the sub-regions k. Under these specifications, $\varphi$ follows a multivariate normal distribution, $\varphi \sim MVN(0, \sum Kk = 1 A_{Tk}\xi_k A_k)$, where $A_k = diag(a_{1k}a_{2k}, ..., a_{nk})$.

In a Bayesian modeling framework, specification of prior distributions of all model parameters is required. Vague normal priors with large variance were assumed for the $\beta$ parameters, while inverse gamma priors were chosen for $\sigma^2$ and $\sigma_k^2$ and uniform priors for $\rho$ and $\rho_k$. Markov chain Monte Carlo (MCMC) simulation was employed to estimate the model parameters . A single chain sampler with a burn-in of 2,000 iterations was run for around 100,000 iterations. Convergence was assessed by inspection of ergodic averages of selected model parameters. The deviance information criterion (DIC) was used to assess the goodness-of-fit of the models without and with exchangeable random effects, and the stationary and non-stationary geo-statistical models . The smaller the DIC, the better the model fit. Finally, Bayesian kriging was used to generate smooth risk maps for Plasmodium infection prevalence based on the parameter estimates of the best fitting model.

Chirombo et al. (2014) attempt to model risk factors of malaria from a Malaria Indicator Survey data.

Suppose $y_i$ is the malaria status of a child such that positive malaria test is recorded as 1, or 0 otherwise. Then, binary response data is generated, which follows a Bernoulli distribution, $y_i \sim Bernoulli(p_i)$, where $p_i$ is the probability of a positive test. With an appropriate link function, the risk of malaria disease can be associated with explanatory variables using a generalized linear model (GLM) framework. GLMs are a flexible alternative to ordinary linear regression, that allow for non-normal response variables

The GLM can be specified with linear predictor $\eta_i = \omega_i^{'}\alpha$ where $\eta_i = logit\left(\frac{p_i}{1-p_i}\right)$

is the logit link function and $\omega_i^{'} = (\omega_{i1}, \omega_{i2}, ..., \omega_{ip})^{'}$ is a matrix of explanatory variables. An ordinary logistic regression is then specified as follows

$$\eta_i = logit\left(\frac{p_i}{1-p_i}\right) = \alpha_0 + \omega_i^{'}\alpha_i \tag{2.1}$$

where $\alpha_0$ is the intercept, $\omega_i^{'}$ is a vector of covariates and $\alpha_i$ is a vector of regression coefficient A limitation to standard GLMs is that they assume independent (or at least uncorrelated) observations. However, this assumption is not always met as sometimes observations exhibit spatial and/or temporal dependence. This needs to be incorporated in models in order to provide a more accurate estimation and prediction of the response variable. The linear predictor, by taking into account the spatial autocorrelation, can be expanded as follows

$$\eta_i = logit\left(\frac{p_i}{1-p_i}\right) = \alpha_0 + \omega_i^{'}\alpha_i + \sum_{k=1}^{q} f_k\left(x_{ik}\right) + \phi_i + v_i \tag{2.2}$$

where $\alpha_0$ is the intercept, $\alpha_i$ is the parameter corresponding to the categorical fixed variables, $\omega_i^{'} = (\omega_{i1}, \omega_{i2}, ..., \omega_{ip})^{'}$ (e.g. wealth index, age category, location, bed net use) and $f$ is an appropriate smoothing function of continuous covariates, $x_{ik}$ (rainfall, minimum temperature, altitude). Spatially unstructured random effects, $\phi_i$, capture the unobserved spatial heterogeneity and over dispersion at each location such as immunity to malaria while spatially structured random effects, $v_i$, allow for spatial autocorrelation and clustering, for example variation in access to interventions such as ITNs among the communities. Equation 2.2 gives rise to a class of models known as structured additive regression (STAR) models. Generalized additive models (GAM), generalized additive mixed models (GAMMs) and geoadditive models are special cases of the STAR models. All of these models make use of smooth functions to model covariate effects on the response variable. These models are increasingly being applied to model health impacts and outcomes such as spatial variation of HIV infections and effects of climate on malaria across Africa. The implementation of the model follows a

Bayesian approach. In Bayesian analysis, all the regression coefficients and the smooth functions $f_j$ are considered as random variables and are assigned prior distributions. Without any prior knowledge, the coefficients $\alpha$ of the continuous covariates are assigned diffuse priors, i.e $p(\alpha_i) \propto const.$ The unknown smooth functions $f_i(x_{ik})$ are assigned Bayesian penalized splines priors . The functions are assumed to be approximated by a polynomial of degree $l$ which is defined over a set of equally spaced knots of the form $x_{min} = \varsigma_0 < \varsigma_1 < ... < \varsigma_{m-1} < \varsigma_m = x_{max}$. The spline is expressed as a linear combination of B-spline basis functions. This approach is similar to fitting second order random walk priors of the form $\beta_k = 2\beta_{k-1} - \beta_{k-2} + \mu_k$ with Gaussian errors,$\mu_k$, assigned to the smooth terms. The spatial random terms are also fitted as splines, particularly as a two-dimensional tensor product. The unknown $\beta_j$ are assigned priors of the general form

$$p\left(\beta_j|\tau_j^2\right) \propto \frac{1}{\left(\tau_j^2\right)^{rank(\frac{k_j}{2})}} exp\left(-\frac{1}{2\tau_j^2}\beta_j' K_j \beta_j\right) \tag{2.3}$$

where $K_j$ is the penalty matrix and $\tau_j$ s the variance parameter that controls the tradeoff between flexibility and smoothness. The $\tau_j^2$ is assigned non-informative dispersed inverse Gamma priors , $p\left(\tau_j^2\right) \sim IG(a_j, b_j)$ where

$$\tau_j^2 \propto \frac{1}{(\tau_j)^{a_j+1}} exp\left(\frac{-b_j}{\tau_j}\right) \tag{2.4}$$

To capture the spatial effects we assumed stationary Gaussian process with zero mean and variance $\xi_{ij} = \sigma^2 corr(d_{ij}, \rho)$ where $\sigma^2$ is the sill, and $corr(d_{ij}, \rho)$ is the spatial correlation. The spatial correlation is considered a function of distance, $d_{ij}$ between the spatial locations $s_i$ and $s_j$ under isotropic assumptions. Usually the exponential correlation function is assumed such that $corr(d_{ij}, \rho) = exp(-d_{ij}, \rho)$. The parameter $\rho$ measures how fast the correlation decays as the distance between the locations increases. Bayesian inference was done using MCMC simulation

based on the posterior distribution

$$p\left(\beta_1, ..., \beta_p, \tau_1^2, ..., \tau_p^2, \alpha|y\right) \propto L\left(y, \beta_1, ..., \beta_p, \alpha\right) \prod_{j=i}^{p} p\left(\beta_j|\tau_j^2\right) p\left(\tau_j^2\right) \tag{2.5}$$

In order to assess factors that are associated with the probability of an under five child testing positive for malaria, different models were fitted as follows:

$$J : \eta_i = \alpha_0 + \omega_i^{'}\alpha_i$$

$$K : \eta_i = \alpha_0 + \omega_i^{'}\alpha_i + f_1(rainfall) + f_2(altitutde) + f_3(min \quad temp) + f_4(latitude)$$

$$L : \eta_i = \alpha_0 + \omega_i^{'}\alpha_i + \phi_i + v_i$$

$$K : \eta_i = \alpha_0 + \omega_i^{'}\alpha_i + f_1(rainfall) + f_2(altitutde) + f_3(min \quad temp) + f_4(latitude) + \phi_i + v_i$$

In the fixed effects model, $J$, categorical and continuous variables were fitted linearly in the usual GLM framework. In these models, $\alpha_0$ is the intercept and $\alpha_i$ is the vector of coefficients of the categorical variables, $\omega_i^{'}$. The second model $K$ includes smooth functions of the $q$ continuous $\sum_{k}^{q} f_k x_{ik}$ such as rainfall and altitude, to assess the importance of possible non-linear associations. In model $L$ random effects of location were included, together with all other covariates, fitted as fixed effects. Lastly, model $M$ included categorical variables fitted as fixed effects, continuous covariates fitted as smooth functions to account for non-linearity, and spatial random effects.

Bivariate tests were carried out in order to determine which variables to include in the models. Initial descriptive analysis was done using cross tabulations and assessed using the Chi-square test to investigate the relationship between the outcome of the malaria test and several categorical variables at the 95% confidence level (CI).

In running the MCMC algorithm, 10 000 iterations were made with a burn in of 1000 and a thinning parameter of 50. To ensure that the choice of the priors in the Bayesian analysis did not influence the results, a sensitivity analysis was performed by running the chosen model several times, changing the prior param-

eters at each run and then comparing the observed changes in the estimates. The default gamma prior with hyper-parameters equal to $(a = 0.001, b = 0.001)$ was changed and the model was run 3 times with the new priors $(a = 0.00001, b = 0.00001)$, $(a = 0.0005, b = 0.0005)$ and $(a = 1, b = 0.005)$. The Deviance Information Criterion (DIC) was used to compare the fitted models $J, K, L$ and $M$ (the smaller the DIC, the better the model). Convergence was assessed through trace plots. Analyses were performed using the free software BayesX in a full Bayesian approach using MCMC. The R statistical software and BayesX package in R were also used to analyze and visualize results.

Adekunle et al. (2013), analyses treatment decision making in the home management Of malaria among children in south western Nigeria. The variable "persons who advised caregivers on the drug to give the child $(Y_i)$" which is the outcome variable has 3 categories which are (1) Mother and Grandmother (2) Father and (3) CBD which comprised of health worker, mother, trainer and Patent Medicine Seller (PMS).The outcome variable has more than two categories hence; the multinomial logistic regression model was fitted. The dependent variable, Yi, is the choice made by the $ith$ individual among the treatment decision makers while the independent variables say $x_i = x_{i1}, x_{i2,x_{i3}}T$ are divided into a set of individual-specific decisions that are alternative-specific, say $w_i = (w_{ni1}, ..., w_{ni}Ka)T$ for $i = 1, ..., j$. The probability of individual "i" choosing alternative "j" is given by the standard multinomial logit formula:

$$P_{ij} = \frac{e^{y_{ij}}}{\sum\limits_{j=1} e^{y_{ij}}}$$

Where, $Log(Y_{ij}) = \beta_0 + \beta_1 X_{i1j} + \beta_2 X_{i2j} + \beta_3 X_{i3j}$. The models as expressed above has one important property related to the odds for two response which is the ability to determine the probability of decision making in each category for individuals I making alternative decision $j$. A base alternative is set and the remaining alternatives are treated as differences to the base.Thus, if the base

alternative is 1, $V_{i1} = 0$ for all $i$. Since each decision maker has a probability then the relative risk ratio of decision making relative to the base decision maker is given by .

$$\frac{P\left(Y_{ij} = k\left(x_i, U_{ij}, \varepsilon_{ij}\right)\right)}{P\left(Y_{ij} = l\left(x_i, U_{ij}, \varepsilon_{ij}\right)\right)} = e^{V_{ij}^k - V_j} \tag{2.6}$$

The equation above proves that the probability depends only on the difference on the linear predictor.The responses are conditionally uncorrelated when controlling for $U_j$. In this case, the introduction of the random effects allows the independence from irrelevant alternatives conditionally on the covariates to hold. The factors that were considered as independent variables are "How far is your home from where orthodox medicine was received $(x_{1j})$", "When did you start treatment afterillness started $(x_{2j})$" and "If orthodox medicine was given, where was the treatment obtained $(x_{3j})$". Like a binary logistic model, the multinomial logistic model helps to measure the contribution of each independent variable to the category of the dependent variable using their relative risk ratios.The multinomial model used for the data is given by :$\ln(p(y_1 = k_2/x, U_j,$
$varepsilon_{ij})) = b_1(When\ did\ you\ start\ treatment\ after\ illness\ started)_{i1}$
$+ b_2(How\ far\ is\ your\ home\ from\ where\ orthodox\ medicine\ was\ received)_{i2}$
$+ b_3(If\ orthodox\ medicine\ was\ given,\ where\ was\ the\ treatment\ obtained)_{i3}$
where: $(y_i = k_1) = person\ who\ makes\ decision(mother\ /grandmother)(y_i = k_2) = person\ who\ makes\ decision\ (father).$

In an attempt to analyze spatially exploit Burden estimates of malaria, Gosonia et al, 2012 models a malaria survey data in Tanzania. At each location in $S = (s_1, s_2, ..., s_n)^T$ let us consider the binary outcome $Y_{ij}$ which takes value 1 or 0 to indicate whether the child $j$ at location $s_i$ was found parasitaemia positive. A logistic regression model was used to relate the outcome to its predictors. The multivariate logistic regression model is given by

$$logit\left(p_{ij}\right) = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ij}^k \tag{2.7}$$

where $p_{ij}$ is the risk of child $j$ at location $s_i$ of having parasitaemia, $Xij = \left(X_{ij}^1, X_{ij}^2, ..., X_{ij}^p\right)^T$ are the covariates and $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^T$ is the vector of regression coefficients, including the intercept. Preliminary frequentist statistical analysis was performed in Stata version 10.0 (Stata corporation, College Station, TX) to identify the covariates significantly associated with the parasitaemia prevalence and to check for possible nonlinear trends of the relationship between explanatory variables and response variable. To account for the nonlinear relationship between malaria and demographic or environmental and climatic covariates, the continuous variables were categorized, where the cutoffs points were chosen based on outcome-covariate scatter plots. The model described above does not consider the spatial relationship among the parasitaemia survey locations. The standard way of incorporating the geographical dependence in the model is by introducing spatially correlated random effects $\phi$ at every sampled location . Using the geostatistical design described in $s_i$ the underlying spatial process was modeled by the residuals via a multivariate Normal distribution with mean 0 and the covariance matrix $\xi$, $\phi = (\phi_1, ..., \phi_n) \sim MVN(0, \xi)$. The covariance matrix is defined as a function which represents the decay in correlation between pairs of locations with distance. For this analysis, an exponential correlation function was chosen, that is $\xi_{ij} = \sigma^2 exp\left(\frac{-d_{ij}}{\rho}\right)$ where $\rho$ describes how fast the spatial correlation declines with distance between locations $i$ and $j$ and $\sigma^2$ represents the variance of the spatial process. Measurement error or microscale variations are modeled by the independent and normally distributed random effects $\varepsilon_i \sim N(0, \tau^2)$ where $\tau^2$ is interpreted as a nugget effect. The model in equation 2.7 can be written

$$logit\left(p_{ij}\right) = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ij}^k + \phi_i + \varepsilon_i \tag{2.8}$$

This type of hierarchical models are usually fitted within a Bayesian framework because it allows flexible modeling and inference and avoids the computational problems met in likelihood-based fitting. The trade-off for the flexibility of a

fully Bayesian approach is the complexity of the model fit. This step is carried out via the implementation of Markov chain Monte Carlo methods. To complete the specification of the Bayesian hierarchical model, prior distributions need to be assigned to the model parameters $\theta = (\beta, \sigma, \rho)$ . Further, Bayesian inference is based on the posterior distribution, that is the conditional distribution over parameters given observed data. Non-informative Normal prior distributions were specified for the intercept and the regression coefficients,$p(\beta) = N(0, 1000)$ . The spatial correlation parameters $\sigma$ and $\rho$ were assigned an inverse gamma and a gamma prior distribution, respectively, $p(\sigma) = IG(a_1, b_1)$ and $p(\phi) = IG(a_2, b_2)$ . Non-informative inverse gamma prior distribution was chosen for the non-spatial variance, $p(\tau^2) = IG(a_3, b_3)$ . The values of the parameters $a_1, b_1, a_2, b_2, a_3, b_3$ were chosen such that the mean of the corresponding distribution is 1 and the variance 100 . A two chain sampler of 100000 iterations was run with a burn-in of 10000 iterations and the convergence was assessed by examining the ergodic averages of selected parameters.

In an ecological study to detect possible risk factors in the Brazilian Amazon region, Achcar (2011) uses a longitudinal data analysis approach dubbed Poisson spatiotemporal regression models to analyze malaria count for the period from 1999 to 2008. Let $Y_{ij}$ be the yearly number of malaria cases for province i in the year $j$, $i = 1, ..., n$ and $j = 1, ..., T$, with a Poisson distribution, given by $Y_{ij}|\mu_{ij} \sim P(\mu_{ij})$ where $P(\mu)$ denotes a Poisson distribution with parameter $\mu$. In connection with the malaria data we consider the following covariates: $X_{1i}$, denoting the number of inhabitants per $km^2$ (in the year 2004) in province $i$; $X_{2i}$, denoting the percentage of urban population (in the year 2004) in province $i$; $X_{3i}$, denoting the number of doctors per 10,000 inhabitants (in the year 2005) in province $i$; and $X_{4i}$, denoting the human development index (HDI) (in the year 2005) in province $i$ of the Amazon forest region, $i = 1, ..., 9$. We also consider the deforestation index for province $i$ in the year $j$ as another covariate, denoted by $X_{5iji}$, $i = 1, ..., 9$, $j = 1, ..., 10$ (corresponding to the years 1999 to 2008).

Considering the covariates $X_1, X_2, X_3, X_4$, and $X_5$, let us assume the following regression model for the Poisson distribution

$$\begin{aligned}\log(\mu_{ij}) =& \beta_{0j} + b_i + W_{ij} + \beta_{1j}\left(X_{1j} - \bar{X_1}\right) + \beta_{2j}\left(X_{2j} - \bar{X_2}\right) + \beta_{3j}\left(X_{3j} - \bar{X_3}\right) \\ &+ \beta_{4j}\left(X_{4j} - \bar{X_4}\right) + \beta_{5j}\left(X_{5j} - \bar{X_5}\right)\end{aligned} \tag{2.9}$$

for $i = 1, ..., 9$ (provinces) and $j = 1, ..., 10$ (years); , $l = 1, ..., 4$ denotes the sample average for the covariate $X_l$, that is, $\bar{X_l} = \sum_{i=l}^{n} X_{li}$. In equation 2.9, we observe that $\beta_{0j}, \beta_{1j}, \beta_{2j}, \beta_{3j}, \beta_{4j}$, and $\beta_{5j}$ are fixed effect regression parameters associated with the covariates $X_1, X_2, X_3, X_4$, and $X_5$; $i = 1, ..., 9$ and $j = 1, ...., 10$. bi is a random effect that captures the possible correlations among the malaria counts, taking into account the region effects of neighboring provinces assumed to have a normal CAR structure model, that is,

$$b_i|\{b_{i^*}, i^* \in A^*(i)\}, \varsigma_b \sim \left(\bar{\eta_i}, \frac{\varsigma_b^2}{n(i)}\right) \tag{2.10}$$

where $A^*(i)$ denotes the set of neighbours corresponding to province i, $n(i)$ denotes the number elements in $A^*(i)$, $\bar{\eta_i}$ denotes the mean of the neighbouring random effects for province i, and $\varsigma_b$ is an unknown parameter.

In equation 2.8 we also assume a random effect $W_{ij}$ for the longitudinal trend specified as a Gaussian process with a multivariate normal distribution with a mean vector $10 \times 1$, with all components equal to zero and a $10 \times 10$ variance-covariance matrix $\xi = [Cov(W_{ij}, W_{ij^*})]$ , with elements given by:

$$Cov(W_{ij}, W_{ij^*}) = \theta_1 exp\left(\sum_{k=2}^{K} \theta_k \left|j - j^*\right|^{k-1}\right) \tag{2.11}$$

assuming a fixed value for $K$. Different fixed values for $K$ can be considered, which gives a great flexibility of fit for the data, and $\theta_k$, $k = 1, 2, ..., K$, is an unknown parameter. Observe that model (2.11) generalizes the longitudinal trend specification introduced by Branscum et al

For the first stage of the hierar-chical Bayesian analysis, let us assume normal prior distributions $\beta_{1j} \sim N(0, a_{lj}^2)$, $l = 0, 1, 2, ..., 5$, $j = 1, ..., 10$, where $a_{lj}$ are known hyperparameters. Taking large values for $a_{lj}$, we have highly dispersed prior distributions for the regression parameters. For the second stage of the hierarchical Bayesian analysis, let us assume uniform prior distributions $\varsigma_b \sim U(0, c)$ and $\theta_k \sim U(0, d_k)$ for $k = 1, 2, ..., K$; $c$ and $d_k$ are known hyperparameters.

For a hierarchical Bayesian analysis of the model, we consider the use of Markov Chain Monte Carlo (MCMC) methods. For the model choice, assuming different values for K in (2.11), we use some existing Bayesian model discrimination criteria, one of which is given by the Deviance Information Criterion (DIC). The smaller the DIC, the better the fit of the model for the data. We also consider some goodness-of-fit techniques in choosing the best model. In this way, we compare the observed data with the fitted or estimated posterior means for $\mu_{ij}$ using the simulated Gibbs samples for each parameter of the model, given by:

$$C(l) = \sum_{i=1}^{9} \sum_{j=1}^{10} |y_{ij} - \mu_{ij}| \tag{2.12}$$

where $l$ indexes the model ($l = 1$ for model 1 and $l = 2$ for model 2), and $\mu_{ij}$ is the MCMC estimate for $E(\mu_{ij}|y, x)$, $i = 1, ..., 9$, $j = 1, ..., 10$.

To verify if the covariate deforestation rate and the covariate number of inhabitants per $km^2$, percentage of urban population, number of doctors per 10,000 inhabitants, and HDI index have some significative effects in the yearly counts of malaria in the Brazilian provinces of the Amazon forest region, we assume the Poisson regression considering two special cases. First, a model defined by (2.7) and (2.8), not considering the temporal random effect $W_{ij}$ and assuming that the random effect $b_i$ has a normal distribution $N(0, \sigma_b^2)$, where $\sigma_b^2$ is an unknown variance; let us denote this as model 1. Second, a model defined by (2.7) and (2.8), with a normal CAR structure (2.9) for the random effects $b_i$ (spatial structure) and a multivariate normal distribution with covariance structure (2.11) for the random effect $W_{ij}$ (temporal structure) with $K = 2$; let us denote this as

model 2. For a hierarchical Bayesian analysis of model 1 and model 2, let us assume a uniform $U(0,1)$ prior distribution for $\varsigma_h$, where $\varsigma_h = \frac{1}{\sigma_b^2}$, the normal prior distributions for the regression parameters, $\beta_{lj} \sim N(0, \alpha_{lj}^2)$, $l = 0, 1, ..., 5$; $j = 1, 2, ..., 10$; and the uniform prior distributions for $\varsigma_b^2$ and $\theta_k$, $k = 1, 2$ with $\alpha_{lj}^2 = 10$; $b = 0$, $c = 1$, $d_1 = 3$, and $d_2 = 1$. Using the WinBUGS software, we simulate the two models: 10,000 initial Gibbs samples as a burn-in sample to eliminate the effect of the initial values on the Gibbs sampling algorithm. After this burn-in sample period, we simulate another 1,600 samples, taking every $50^{th}$ sample to have approximately uncorrelated samples for the joint posterior distribution of all parameters of the model. Convergence of the Gibbs sampling algorithm was observed using plots of the simulated time-series samples.

In a conscious effort to measure malaria transmission intensity and to observe transmission patterns, Bosomprah,2012 formulates a 'Mathematical model of seropositivity to malaria antigen; allowing seropositivity to be prolonged by exposure'. He models from samples from two separate data sources and time points in Tanzania. The concept of superinfection has been used to describe periods of infection prolonged by repeated exposure to infection. Persons in endemic areas often have pre-existing partial immunity. But when these persons are removed from exposure the immunity can be lost gradually. When the person is re-exposed while seropositive the level of antibody response can be boosted. A simple way to allow for the antibody response to be boosted by exposure is to consider that each exposure gives rise to an antibody response. This can be thought of as a set of antibody-producing cells that are triggered by the exposure. Suppose the random variable $v$ represents the number of such sets of cells, and $x_i$ is the probability $P(v = i)$. For every exposure the value of $v$ has a one-unit increase, and when any of the sets of cells dies the value of $v$ has a one-unit decrease. If the average duration of a set of cells is $\frac{1}{r}$, the rate for the transition from $v = i$ to $v = i - 1$ is the product of $i$ and $r$ (i.e. $i \times r$). Because there are $i$ sets of cells the value of $v$ reverts from $i$ to $i - 1$ when any of the $i$ sets of cells dies off .

The compartmental model can be represented in differential equation as shown in equation (2.13) below.

$$\frac{dx_i}{da} = \lambda x_i + \lambda x_{i-1} - 1 - irx_i + (i+1)rx_i + 1,\ i = 0, ..., \infty; x_i = 0 \ for\ i < 0 \quad (2.13)$$

where $x_i$ = Probability (number of exposures, $v = i$), so that $\sum \infty, i = 0, x_i = 1$. These equations can be solved by a standard method, using generating functions . The basic superinfection model with seroconversion rate, $\lambda$, assumed constant over time has been derived as:

$$P(a) = 1 - exp\left(-\lambda r\left(1 - exp(-ra)\right)\right) \quad (2.14)$$

where $r$ is the annual rate of reversion from seropositive to seronegative state per exposure. It follows that the number of exposures in a person aged, $a$, follows a Poisson distribution with mean:

$$P(a) = \lambda r\left(1 - exp(-ra)\right) \quad (2.15)$$

Suppose that $\lambda$ has changed abruptly from $\lambda_1$ to $\lambda_2$ at a certain point in calendar time, $\mu$, but is otherwise constant for different ages (i.e. $\lambda(t) = \lambda_1(t < \mu)$, $\lambda(t) = \lambda_2(t > \mu)$) , and people were observed at time $t = \mu + c$, where $c > 0$. The seroprevalence at age a at time $t$ is therefore given by:

$$P(t,a) = 1 - exp\left[-\left\{\left(\lambda_1 r\left(exp\left[-r(t-u)\right] - exp[-ra]\right)\right) f(t-a) + \lambda_2 r(1 - exp[-r(t-\mu)])\right\}\right] \quad (2.16)$$

This is the superinfection model with an abrupt change in seroconversion rate, $\lambda$. This model can be used to investigate abrupt changes in malaria transmission in the recent time past. In the specification of the basic superinfection model there are two main parameters $(\lambda, r)$. But in the specification of the superinfection model, which allowed an abrupt change in seroconversion rate there are three

main parameters $(\lambda_1, \lambda_2, r)$. The model was fitted to age-stratified serological data using the method of maximum likelihood. In this model, the dependent variable $Y$ is an indicator variable, meaning that it takes on only the values 0 (= seronegative state) or 1 (= seropositive state). The probability mass function from the Bernoulli distribution - the distribution for the random indicator variable is: $f(y_j; \pi_j) = \pi_j \ \textit{if}\ y_j = 1, 1 - \pi_j \ \textit{if}\ y_j = 0$

where $0 \leq \pi_j \leq 1$ and $\pi_j$ was identified as the probability for a success (arbitrarily $y_j = 1$ is called a success). The log-likelihood function for the $j^{th}$ observation is:

$$\ln \ell_j = \ln(P(\alpha_j)) \ \textit{if}\ y_j = 1, \ln(1 - P(\alpha_j)) \ \textit{if}\ y_j = 0 \tag{2.17}$$

where $y_j$ is the indicator variable $y_j = 1$ if person $j$ is seropositive and $y_j = 0$ if they are seronegative, and $a_j$ is their age and $P(a_j)$ is the proportion seropositive at age $a_j$. Cator variable, meaning that it takes on only the values 0 (= seronegative state) or 1 (= seropositive state). The probability mass function from the Bernoulli distribution - the distribution for the random indicator variable is: $f(Y_j, \pi_j) = \begin{cases} \pi_j \ \textit{if}\ Y_j = 1 \\ 1 - \pi_j \ \textit{if}\ Y_j = 0 \end{cases}$ where $0 \leq \pi_j \leq 1$ and $\pi_j$ was identified as the prob. For a success (arbitrary $Y_j = 1$ is called success).The log-likelihood function for the jth observation is: $\ln \ell = \begin{cases} \ln(P(\alpha_j)) \ \textit{if}\ Y_j = 1 \\ \ln(1 - P(\alpha_j)) \ \textit{if}\ Y_j = 0 \end{cases}$ Where $y_j$ is the indicator variable $y_j = 1$ if person $j$ is seropositive and $y_j = 0$ if they are seronegative, and $a_j$ is their age and $P(a_j)$ is the proportion seropositive at age $a_j$.

Mideo et al. (2008), attempts to model malaria pathogenesis. To illustrate what such models typically look like and how, through the process of model development, the irrelevant biological details are uncovered, we present a simplified generic example based on Mideo et al. (2008). The model is in discrete time to account for the distinctly discrete life cycle of malaria parasites, and the densities are evaluated every day, corresponding to the 24 h cycle of the rodent malaria

system on which this model was based. The model predicts how the density of three quantities, merozoites (M), gametocytes (G), and red blood cells (R), change from one day to the next;

$$M(t+1) = f_M[M(t), R(t)]$$
$$G(t+1) = f_M[M(t), R(t), G(t)]$$
$$R(t+1) = f_R[M(t), R(t)]$$

The above equations capture the idea that the density of each quantity in the next day (time $t+1$) is some function of their densities on the present day (time $t$). Notice that two of these functions do not depend on the gametocyte density, $G(t)$, reflecting an assumption that gametocytes play no role in determining the merozoite or RBC counts on the next day. Other assumptions about how various biological processes work (e.g. erythropoeisis, RBC infection, gametocytogenesis, etc.) are captured by the specific forms of the functions $fM[M, R]$, $fM[M, R, G]$ and $fR[M, R]$ (TabModels like that above can also be further refined as necessary, by including things such as RBC age structure and time-lags in erythropoeisis. They can also be extended to include other regulatory factors related to innate and adaptive immune responses. For example, one might introduce other variables that represent the densities of different immune effectors molecules and cells. If we use $T(t)$ to denote the density of specific $T$ cells on day t then the model might be extended as

$$M(t+1) = f_M[M(t), R(t), T(t)]$$
$$G(t+1) = f_M[M(t), R(t), G(t), T(t)]$$
$$R(t+1) = f_R[M(t), R(t), T(t)]$$
$$T(t+1) = f_T[M(t), R(t), T(t)]$$

where the functions $fM[M, R, T]$, $fG[M, R, G, T]$, $fR[M, R, T]$ and $fT[M, R, T]$ are specified to account for the relevant assumptions about how these processes

work .The predictions of each model obtained by employing a different set of assumptions can then be tested against data to determine its ability to explain known patterns of pathogenesis (Mideo et al., 2008).

Key words:

Indoor Residual Spraying (IRS)

Insecticide Treated Nets (ITNs)

Long Lasting Insecticides Nets (LLINs)

# Chapter 3

# METHODOLOGY

## 3.1 Data Collection

The data obtained for this study was mainly from a primary source due to the group and scanty nature of existing secondary data at the various health posts and malaria control outfit.

The data collection involved a 3 week household cross sectional survey across 22 clusters within The Malaria Control Programme (MCP) Coverage area in the Obuasi Municipality. The target group was children under 5 years and the respondents were mostly mothers aged between 15 and 55 years old. A sample of 508 children was selected by Quota and Simple Random Sampling (SRS) from 22 clusters.

Again, the mothers of some of these children misreported the ages of their children by either underestimating or rounding them off to 1 year. These mothers could not produce the weighing cards or relevant documents to confirm the ages of their children. The instrument used for the data collection (the questionnaire) went through some validation and scrutiny. It was also pre-tested before the main data collection.

The Survey was meant to capture information on whether children under 5 years in the Obuasi Municipality benefited from the two major malaria interventions strategies namely Indoor Residual Spraying (IRS) and Insecticide Treated Nets (ITNs) distribution and to use this data to obtain a logistic regression model. It was also purported to exact information on the status of clinical visits within two extreme time points namely the first one year and the last one year of the child and subsequently use the data as a basis for paired samples test.

The main objectives of the study as the foregoing discussion suggests and stated earlier in chapter one are

1. To model the risk of reduction in clinical visits, given the types interventions, in the under 5 years

2. To model frequency of clinical visits in the 1st one year (past) and last 1 year (present) by the Under 5 years, given the types of malaria Control interventions. Consequently for the logistic regression model, we derive the following outcomes.

Reduction (success or 'case') = 1

Increase in visits (Failure or 'No case') = 0

Net usage (ITN) = 1, otherwise = 0

Indoor residual spraying (IRS) = 1, otherwise = 0

Both ITN and IRS = 1, otherwise = 0

Neither = 1, otherwise = 0

In considering several predictors which were factored into the outcomes (Increase or Reduction), their aggregated ODDS was obtained as

$$ODD = \frac{P}{1-P} = e^{(\beta_0+\beta_1(ITN)+\beta_2(IRS)+\beta_3(Both)+\beta_4(Neither))} \tag{3.1}$$

Taking the logit

$P(x=1) = logit\left(\frac{P(x=1,y=0)}{1-P(x=1,y=0)}\right)$

This converts the original discrete nature of the data to its continuous form and to enable the determination of individual Contribution of each Covariate $(x_1, x_2, ..., x_p)$ to the probability of success or to make the estimation of the Odds ratio more feasible.

In this case, the parameters $\beta_0, \beta_1, \beta_2, ...$ are clearly defined. Thus

$\beta_0$... Risk of clinical visits (reduction or increase) in totality.

$\beta_1$...Marginal change in ITN usage

$\beta_2$...Marginal change in IRS usage.

## 3.2 Modeling Logistic Regression

The multivariate problem that usually crops up in studies that involve the use of logistic model is discussed briefly. Here the relationship established between at least one explanatory (exposure) variable and a dependent (response) variable is being described. Thus a binary risk of reduction in clinical attendance outcome with one (1) denoting a reduction in clinical attendance (success or case) and zero (0) representing an increase in clinical attendance (failure or no case) is being examined in this study. Apparently, we seek to use the variables $X_1, X_2, X_3, ..., X_k$ (where $k$ is the number of explanatory variables or number of exposures available to the subjects) to determine (predict) the dependent binary variable $Y$.

## 3.3 Brief Historical Background

Logistic regression was put forth in the 1940's as an alternative to Fisher's 1936 classification method linear discriminant analysis. It is extensively used in many disciplines, including the medical and social science fields. For example, the Trauma and Injury Severity Score (TRISS), which is widely used to predict mortality in injured patient, was originally developed by Boyd et al using logistic regression. In recent times, a number of researchers have extensively used logistic regression to model their study. For example, Chirombo et al. (2014) attempt to model risk factors of malaria from a malaria Indicator data. They tried to assess factors associated with the probability of an under five child testing positive for malaria by fitting different logistics models. Again, Adekunle et al. (2013), analysis treatment decision making in the home management of malaria among children in south western Nigeria. A multinomial logistic regression was fitted for the outcome variable $Y_1$ which is the choice made by the individual among the treatment decision makers and the explanatory variables categorized into (i) Mother and grandmother (ii) Father and (iii) Health worker, mother, trainer and patent medicine seller and unpresented by $X_1 = X_{1_1}, X_{1_2}, X_{1_3}$

## 3.4 The Logistic Function

An explanation of the logistic regression begins with an explanation of the logistic function which always takes on values between zero (0) and one (1). That is a logistic function; $f(g) : 0 \leq f(g) \leq 1$.Whenever we are analyzing a multivariate problem involving a set of $k$ explanatory variable $X$ and an outcome (dependent) variable $Y$, a typical mathematical model is fitted to cater for the absolute relationship among several variables. Logistic regression is a type of probabilistic statistical classification model used in establishing or describing the relationship between many of $X's$ (explanatory variables) and a binary outcome (dependent) variable $Y$.

Logistic regression is the most ubiquitously used modeling approaches compared to other approaches because it is suited for analyzing survey data whenever the outcome variable is binary. The logistic function which describes the mathematical nature of the logistic model is

$$f(g) = \frac{1}{1 + e^{-g}} \tag{3.2}$$

$$f(-\infty) = \frac{1}{1 + e^{-(-\infty)}} \tag{3.3}$$

$$f(+\infty) = \frac{1}{1 + e^{+\infty}} \tag{3.4}$$

The logistic function $f(g)$ necessarily takes values in the range 0 to 1.Thus; it presupposes that the model is designed to assume a risk estimate between the 0 and 1 inclusive. Consequently, the logistic function has the practical effect of converting the probability (which is bounded to be between 0 and 1) to a variable that ranges over $(-\infty, +\infty)$.

The characteristic S-shaped logistic function is believed by researchers to be suited for numerous research conditions. It explains the extreme 0 and 1 cut-off point conditions where $f(g)$ becomes extremely close to 0, low for minimal values of $g$ and becomes extremely close to 1, high for high values of $g$ respectively and also

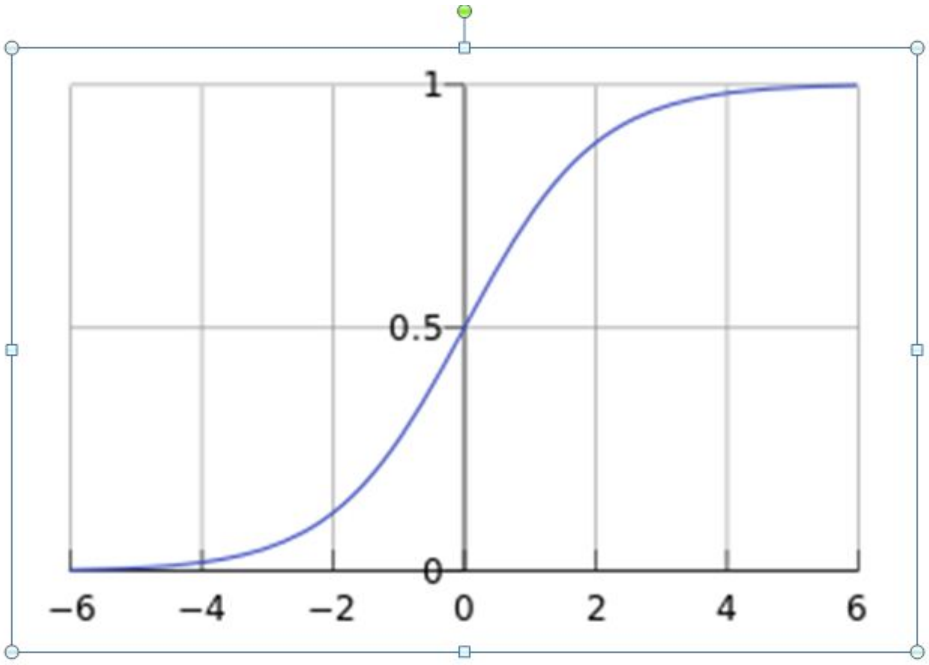


Figure 3.1: The logistic Model

increases sharply for some ranges of $g$ values. This epitomizes the S-shape of $f(g)$ which represent the risk estimate of the index $g$ which aggregates the inputs of many risk determinants.

## 3.5 The Logistic Model

The logistic model is derived by writing $g$ as the linear sum;

$$g = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...\beta_k X_k, \tag{3.5}$$

modeling the risk of reduction in clinical attendance against some explanatory variables defined in this context as;

$X_1$; The use of Insecticide Treated Nets (ITNs)

$X_2$; The use of Indoor Residual Spray (IRS)

$X_3$; The use of both Interventions

$X_4$; The use of neither of the Interventions

and $\beta_0, \beta_1, \beta_2, ..., \beta_k$ denote unknown but determinable parameters. Essentially the $X$ variables are aggregated by the index $g$. Again we present our logistic

model in (3.5) as function of $X$ as

$$\pi(x) = \frac{e^{\beta_0 + \varepsilon\beta_i X_i}}{1 + e^{-(\beta_0 + \varepsilon\beta_i X_i)}} \tag{3.6}$$

The model takes into account the occurrence chances of some observed explanatory variables $X_1, X_2, ..., X_k$. Thus, if success (or case) is assigned 1 and failure (non -case) is assigned 0, then within a time span $T_k$, $X_k \to Y(0,1)$.Consequently, the probability statement

$$P(Y_i|X_1, X_2, ..., X_k) = \frac{1}{1 + e^{-(\beta_0 + \varepsilon\beta_i X_i)}} \tag{3.7}$$

defines the success of an event $X$, hence the logistic model in (3.6), where $\beta_0$ and $\beta_i$ are unknown but determinable parameter is used to predict the probability of an event $X$ being a success within a time span $T_k$. We write are logistics model as

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \varepsilon\beta_i X_i)}} \tag{3.8}$$

## 3.6 Application To Risk Ratio (RR)

The ratio between probabilities $\pi_1$ and $\pi_2$ of success for two subjects usually in a follow up study is the Risk Ratio .Thus $RR = \frac{\pi_1}{\pi_2}$.
Supposing we fit a model and it yields the following estimated parameters; $\beta_0 = 3.910$, $\beta_1 = 0.650$, $\beta_2 = 0.028$ and $\beta_0 = 0.34$, substituting, our model now is $\pi(X) = \frac{1}{1+e-(-3.910+0.650X_1+0.028X_2+0.340X_3)}$ Now we consider the following data sets for two individual subjects ;

SET A: $X_1 = 0, X_1 = 40, X_1 = 0$

SET B: $X_1 = 1, X_1 = 40, X_1 = 0$

For data SET A, $\pi_1(X) = 0.7000$ , for data SET B, $\pi_2(X) = 1100$.The risk ratio is estimated to be RR = 0.1100 / 0.0700 = 0.6364,thus the logistic model can be used to estimate a Risk Ratio which makes two separate individual subsets in a

follow-up study comparable with each other. The direct estimation of Risk Ratio (RR) is guaranteed by the following conditions:

1. There must be a follow-up study to conveniently determine individual Risk estimate.

2. The numerical values for the entire set of explanatory variable need to be stated in our fitted model in order to obtain individual subject's Risk estimates.

One these conditions are met, Risk Ratio (RR) can be estimated directly.

## 3.7 The Odds Ratio

To overcome aforementioned challenges association between two subjects irrespective of whether a particular design is a follow-up, cross sectional or case control was formulated as an indirect means of computing the Risk Ratio.

Research paper from a couple of researchers including Pike and Prentice, 1979 and Bay and Breslow, 1981 have examined and come out with situations in which logistic model can apply to case-control and cross sectional data. In case control for instance, the converse is what obtains in the modelling of a follow-up study. In the latter the response variable is the outcome ($Y_i$) and explanatory (exposure) variables. However, in a follow-up study the situation is the other way round. Essentially, the data from a case-control study is manipulated to conform to a follow-up design, and then a logistic model can be fitted.

Non-the less, there is a unique drawback in a case-control or cross sectional study with respect to logistic modelling. Though the follow-up studies individual subjects' risk is determinable given specific explanatory variables by use of a fitted logistic model, same model cannot be employed in predicting individual risk for case control and cross sectional designs. To overcome this limitation, ODD Ratio estimates are determined for case- control or cross sectional studies. This presupposes that it is the ODD Ratio (OR) and not individual Risk which is

determinable for case-control and cross sectional studies. The ODD Ratio can be extended to the analysis of some simple data, specifically two by 2 tabular data construct. We demonstrate an example as follows:

For such a simple matrixed data the ODD Ratio is suitable so long as the set

Table 3.1: 2 by 2 tabular data analysis

| | | |
|---|---|---|
| $Y = 1$ | p | q |
| $Y = 0$ | r | s |

is generated from a case-control or cross sectional data. However, if the data is obtained from a follow-up data, then the Risk Ratio (RR) applies.

From the table 1, the cell frequencies are denoted by $p, q, r$, and $s$. The Odd Ratio estimate is given by;

$$OR = \frac{ps}{qr} \tag{3.9}$$

Alternatively, the foregoing expression can be expressed as ratio of probabilities for exposure status as far as case control and cross sectional studies are concerned. We have;

$$OR = \frac{P(X = 1|Y = 1)/P(X = 0|Y = 1)}{P(X = 1|Y = 0)/P(X = 0|Y = 0)} = \frac{ps}{qr} \tag{3.10}$$

and from follow-up studies

$$OR = \frac{P(X = 1|Y = 1)}{P(X = 1|Y = 0)} = \frac{p}{q} \tag{3.11}$$

It is observed from (3.11) that the ratio assumes the form Pr(Y/X) and Pr(X/Y) from (3.10). This presupposes that it is extremely difficult to obtain risk estimates for case-control and cross sectional studies because such estimation necessarily need to have conditional probabilities of the form $P(Y/X)$,meanwhile the conditional probability that is feasible for the estimation of risk estimates assumes the form $P(X/Y)$.

It can be justified mathematically why risk estimates is unobtainable by the use of logistic regression in case control or cross sectional studies. In demonstrating this, valid estimates are to be made for the parameters $\beta_0$ and $\beta_i$ in the logistic

model in order to obtain an estimated risk $P(X)$ from the fitted model.

$$\hat{P}(X) = \frac{1}{1 + e^{-(\hat{\beta}_0 + \sum \hat{\beta}_i X_i)}} \tag{3.12}$$

Unfortunately, in the use of the logistic model for case control or cross sectional data $\beta_0$ is not validly determinable without obtaining the sample proportion of the population. Consequently a good estimate of the predicted risk $\hat{P}(X)$ is unobtainable without a reliable estimate of $\beta_0$ since it is determinant of $\hat{P}(X)$. Nonetheless the $\beta_i$ can be conveniently estimated from case control or cross sectional studies and this can lead to the estimation of measures of association by means of the Odd Ratio. In computing the Odd Ratio directly all control variables are assumed to be constant and not specified. When these conditions are met the logistic model can be used to estimate Odd Ratio directly. In the event that the logistic regression is to be employed in obtaining indirectly, a risk estimate, some assumptions need to be made. The major assumption is that, the Odd Ratio presents a reliable approximation to the Risk Ratio.

## 3.8 Logit Transformation

From the foregoing discussions, the Odds are defined as the probability that a particular outcome is a 'case' or 'success' divided by the probability that it is a 'non case' or 'failure'. Logistic regression makes use of at least one predictor variable that may be either continuous or categorical data. However unlike ordinary linear regression, logistic regression is used for predicting binary outcomes of the explanatory variable rather than continuous outcomes. Given this distinction, it is necessary to take the logarithm of the Odd of the explanatory variable being a 'case' or 'success'. This is referred to as the logit or log-odds and it creates a continuous criterion upon which the linear regression is conducted. Consequently the logit transformation is referred to as the link function in logistic regression.

We present the logit transformation as;

$$Logit\pi(X) = \ln\left(\frac{\pi(X)}{1-\pi(X)}\right) \tag{3.13}$$

$$\pi(X) = \frac{e^{x_i} + \sum \beta_i x_i}{1 + e^{-(\beta_0 + \sum \beta_i x_i)}} \tag{3.14}$$

For an individual explanatory variable $X$, the transformation in (3.13) enable us evaluate the $logit\pi(X)$ or the log-odds. For an illustration, given that $\pi(X) = 0.109$, $1 - \pi(X) = 0.891$

$\frac{\pi(X)}{1-\pi(X)} = \frac{0.109}{0.891} = 0.122$

Substituting in (3.13),

$logit\pi(X) = \ln(0.122) = -2.095$. Thus $logit(0.109) = -2.095$.

Now substituting the logistic model $\pi(X) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i x_i)}}$ into the logit transformation logit $\pi(X) = \ln\left(\frac{\pi(X)}{1-\pi(X)}\right)$ and simplifying, we have logit $\pi(X) = \ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \sum \beta_i x_i$ reducing the logit of $\pi(x)$ to a linear sum.

In a nutshell, the ratio $\frac{\pi(X)}{1-\pi(X)}$, defines the Odds for modeling a problem with explanatory variables defined by $X$.

Illustration: Given that $\pi(x) = 0.20$, $1 - \pi(x) = 0.80$

$Odds = \frac{\pi(X)}{1-\pi(X)} = \frac{0.2}{0.8} = \frac{1}{4}$

INTERPRETATION: The result of one-fourth means that the event has 1 in 4 chance that it will occur or the event is 4 times likely that it will not happen than it will occur.

## 3.9 Establishing The Odd Ratio

Ideally, the odds ratio (OR) establishes the relationship between two odds by determining their ratio. Let the first and second odd denoted by $D_1$ and $D_2$ respectively. Then the Odds Ratio denoted by OR is given by $\frac{D_1}{D_2}$.

Supposing in a clinical trial the Treatment and the Placebo groups were observed

to assess their chances of recovering from the following data:

We want to find the Odds Ratio of recovery for Treatment group against the

Table 3.2: Two-way contingency for treatment and placebo group

| | Recovery | Non-Recovery | Total |
|---|---|---|---|
| Treatment | 46 | 74 | 120 |
| Placebo | 66 | 84 | 150 |
| TOTAL | 140 | 130 | 270 |

Placebo group. For the treatment group the Odds of recovery will be $D_1 = \frac{\pi_1(X)}{1-\pi_1(X)} = \frac{46/120}{1-46/120} = \frac{46/120}{74/120}$

$D_1 = \frac{46}{74} = 0.622$

For the placebo group the odds of recovery will be $D_2 = \frac{\pi_2(X)}{1-\pi_2(X)} \frac{66/120}{84/120}$

$D_2 = \frac{66}{84} = 0.786$

Odd Ratio (OR) Treatment, Placebo $= \frac{D_1}{D_2} = \frac{0.622}{0.786} = 0.791$ .This means that the Treatment group has 0.8 times chance of recovery as the placebo group, in fact almost equal chances of recovery.

## 3.10 Derivation of a General Odd Ratio Formula

It is already established that two given data set $S_1$ and $S_2$ we can be compared by means of their Odd Ratio, $OR = \frac{D_1}{D_2}$

where $D_1$ is the Odd of data set $S_1$ and $D_2$ is the Odd of data set $S_2$. Again, $D_1 = \frac{\pi_1(X)}{1-\pi_1(X)}$ and $D_2 = \frac{\pi_2(X)}{1-\pi_2(X)}$.

The Odd Ratio is therefore;

$$OR = \frac{D_1}{D_2} = \frac{\frac{\pi_1(X)}{1-\pi_1(X)}}{\frac{\pi_2(X)}{1-\pi_2(X)}} \tag{3.15}$$

It is also true to say that the probabilities $D_1$ and $D_1$ in OR (Odd Ratio) are both

Table 3.3: Two-way table for the logistic analysis

| | X | X |
|---|---|---|
| Y | p | q |
| Y | r | s |

risk estimates and therefore OR is indeed a risk Odds Ratio. Now substituting our logistic model from (3.8) i.e $\pi(x) = \dfrac{1}{1+e^{-(\beta_0+\sum \beta_i X_i)}}$ and simplifying

$$\begin{aligned} OR_{s_1,s_2} &= \frac{\frac{\pi_1(s_1)}{1-\pi_1(s_1)}}{\frac{\pi_2(s_2)}{1-\pi_2(s_2)}} = \frac{e^{(\beta_0+\sum \beta_i X_{1i})}}{e^{(\beta_0+\sum \beta_i X_{2i})}} = \frac{e^{\sum \beta_i X_{1i}}}{e^{\sum \beta_i X_{2i}}} \\ &= e^{\sum \beta_i (X_{1i}-X_{2i})} = \prod_{i=1}^{k} e^{\beta_i (X_{1i}-X_{2i})} \end{aligned} \tag{3.16}$$

Thus the product Odd Ratio (OR) formula is derived and by the logistic model it is the most suitable for the purpose of estimation. Supposing for $i = 3$ and $i = 5$ equation (15) yields 6 and 9 respectively, then we obtain $6 \times 9 = 54$ as OR.

## 3.11 Analysis Of Logistic Model Involving One Binary Exposure Variable

We discuss a simple scenario that takes into account one binary exposure variable represented by $X_1 = X(0, 1)$. We also denote the outcome variable by $Y$ which is equally binary. To explain this analysis condition, we use a four celled matrix table.

An analysis scenario of this nature will have logistic model $\dfrac{1}{1+e^{-(\beta_0+\beta_1 X)}}$ with $X = (0, 1)$ variable. The corresponding logit model assumes the form

$$Logit \quad \pi(X) = \beta_0 + \beta_1 X \tag{3.17}$$

The expression $\pi(X)$ in this very analysis represents the probability that the outcome variable $Y$ is assigned the value 1, for any given explanatory variable $X$. For a given research design $\pi(X)$ denote the Likelihood of the outcome being a

'success' or 'case' for any given exposure status. A situation in which the exposure variable is 1, is referred to as risk ($R_1$). This is interpreted as the conditional probability of $Y$ being equal to 1 when $X = 1$.When $X$ takes on the value 0, $R_0$ represents the risk and this situation is the conditional probability that $Y$ is 1 given that $X$ is 0.Thus ;

$$\pi(X) = P(Y = 1)$$
$$X = 1 : R_1 = P(Y = 1|X = 1)$$
$$X = 0 : R_0 = P(Y = 1|X = 0)$$

We can now state our Odds Ratio as

$$OR_{X=1,X=0} = \frac{\frac{R_1}{1-R_1}}{\frac{R_0}{1-R_0}}$$

we substitute the logistic model expression into the above formula for Odd Ratio to be able to estimate the Odd Ratio in terms of the parameters $\beta_0$ and $\beta_1$.

$$\pi(X) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1)}}$$
$$X = 1 : R_1 = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1)}} = \frac{1}{1 + e^{-(\beta_0+\beta_1)}}$$
$$X = 0 : R_0 = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1)}} = \frac{1}{1 + e^{-\beta_0}}$$
$$OR_{1,0} = \frac{\frac{R_1}{1-R_1}}{\frac{R_0}{1-R_0}}$$

Substituting

$$OR_{1,0} = \frac{1}{1 + e^{-(\beta_0+\beta_1)}} / \frac{1}{1 + e^{-\beta_0}} \Rightarrow \frac{1 + e^{-\beta_0}}{1 + e^{-(\beta_0+\beta_1)}} = \frac{1}{e^{-\beta_1}}$$

Therefore

$$OR_{1,0} = e^{\beta_1} \tag{3.18}$$

## 3.12 Modelling For Matched Variables

In epidemiological case-control design people with infection and are being investigated are compared with a control group (people without the infection).Here cases and control are compared for a possible risk factor in a conscious effort to examine the nature of the infection. In medical and epidemiological research alike, cases and controls sometimes happen to be matches. In order to ensure that cases and controls are alike in terms of variables, they are matched. Such variables might be related to those ones being studies but are not necessarily variables of interest.

We discuss here a unique situation of logistic model and it's Odd Ratio for a matched data analysis. An essential feature concerning matched data in is its stratified nature .In this case; the strata are the matched which invariably display homogeneity among the pairs. Examples are the pairs in a matched case-control study of association between Use of Oral Conjugated Estrogens and Cervical Cancer.

In defining the strata of matched sets, we use dummy variables, that is, if logistic regression is employed in doing the matched analysis. Again in trying to define a model for a matched analysis, we examine the case of a single (0,1) exposure variable of paramount interest in tandem with a set of control variable $T_1, T_2, T_2, ...T_q$, subject to adjustment due to likely confounding and interaction effects. The assumption here is that some of the C variables have already been matched design by use of one of frequency matching, individual matching or pair matching. The unmatched C variables, as a matter of interest, are treated as control variables.

Let S= (0, 1) exposure variables, $T_1, T_2, T_2, ...T_q$ denote control variables. Also assume some T's matched by design, other T's unmatched. We define some set of variables, given the foregoing background, and factor them in a logistic model for matched data. We define an outcome variable $Y$ by (0,1) and an exposure

variable $X$ by (0,1).Again $V_{1i}$ are (matched) dummy variables, $V_{2i}$ are variables for prospective confounders and there are a set of $X's$ resulting from the product $YW_i$, in which case the $W's$ usually assumes the form of $V'_{2i}s$. Consequently, we state the logistic model for a matched variable as

$$Logit \quad \pi(X) = \beta_0 + \beta X + \Sigma Y_{1i} V_{1i} + \Sigma Y_{2i} V_{2i} + X \Sigma \delta_i W_i \tag{3.19}$$

where $\Sigma Y_{1i} V_{1i}$ =matching, $\Sigma Y_{2i} V_{2i}$ =confounder, $\Sigma \delta_i W_i$ =interaction and $\beta_0, \beta, Y_{1i}, Y_{2i}, \delta_i$ are coefficients.

## 3.13 Odds Ratio For Coding 'X' Arbitrary

Here we discuss an expression for the Odd Ratio for non-specific single exposure variable X, irrespective of whether binary, interval or ordinal which for controls for a set a set of T variables. Invariably X is defined as an arbitrary variable of interest.

$$Logit \quad \pi(X) = \beta_0 + \beta X + \Sigma Y_{1i} V_{1i} + X \Sigma \delta_i W_i \tag{3.20}$$

For a generally defined X of this nature, it is important to have two specific comparable X values in order to determine an Odd Ratio. Let the two specific values of interest be represented by $X^*$ and $X^{**}$. We thus emphasize that even for more than two values of X in the particular instance of X being interval or ordinal, two levels of X is a prerequisite. Again it takes the comparison of two data sets to determine the Odds Ratio.The Odds Ratio for our two values of interest $X^*$ and $X^{**}$ is given by:

$$OR_{X^*, X^{**}} = exp[(X^* - X^{**})\beta + (X^* - X^{**}) \times \Sigma \delta_i W_i]$$

**Illustration:**

Suppose X denotes Health Insurance status (HI) being an index spanning from 1 to 6 where 1 represents someone without Health Insurance and 6 represent a

person with absolute Health Insurance Cover. We seek to determine an Odd Ratio concerning health Insurance status (HI) in the context of $X, V, W$ model. We state two specific values of X, namely $HI^* = 6$ and $HI^{**} = 1$, which conveniently compares the odds of persons who have absolute (maximum) Health Insurance cover with the Odd of persons who have the least Health Insurance Cover. With the foregoing background we compute the Odds Ratio expression:
$HI^* = 6$ vrs $HI^{**} = 1$

$$\begin{aligned} OR_{6,1} &= exp[(HI^* - HI^{**})\beta + (HI^* - HI^{**}) \times \Sigma\delta_i W_i] \\ &= exp[(6-1)\beta + (6-1) \times \Sigma\delta_i W_i] \\ &= exp[5\beta + 5\Sigma\delta_i W_i] \end{aligned} \tag{3.21}$$

## 3.14 A Case of Many Exposure Variables With No Interaction

Besides a scenario of having a single exposure variable with several categories, we discuss the Odd Ratio for a situation of many different exposure variables in our model. In this particular instance, the variables which don't have to be dummy may be represented by $X_1, X_2, ..., X_q$. Thus the different exposure variable can be any one binary, interval or ordinal. Consequently, we model the case of many Exposure Variables with no interaction as

$$Logit \quad \pi(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_q X_q + \Sigma\gamma_i V_i \tag{3.22}$$

Again here, we need to specify the exposure variables for two separate but comparable groups to be able to obtain the Odd Ratio for many variables. These variables shall be denoted by emboldened $X^*$ and $X^{**}$. The $X^*$ category is defined by $X_1^*, X_2^*, ....X_q^*$ and $X^*$ category is defined by $X_1^{**}, X_2^{**}, ....X_q^{**}$.

Thus for $X^*$ vrs $X^{**}$

$$X^* = (X_1^*, X_2^*, ....X_q^*)$$

$$X^{**} = (X_1^{**}, X_2^{**}, ....X_q^{**})$$

The Odd Ratio expression which compares $X^*$ and $X^{**}$ is

$$OR_{X^*,X^{**}} = exp[(x_1^* - x_1^{**})\beta_1 + (x_2^* - x_2^{**})\beta_2 + ... + (x_q^* - x_q^{**})\beta_q] \tag{3.23}$$

## 3.15 A Case of Several Exposure Variables With Confounders And Interaction

Ultimately, we look at a scenario of many exposure variables with confounders defined by the V's and interaction variables defined by W's. Here the W's are absorbed by the model as a coefficient of one of the X's (.i.e. both form a product term). Once again for the Odd Ratio, we need to identify two exposure variables and compare and are still denoted by $X^*$ and $X^{**}$. We further define $X^*$ as $SMK^*$, $PAL^*$ and $SBP^*$ and $X^{**}$ as $SMK^{**}$, $PAL^{**}$, and $SBP^{**}$.

**Illustration:**

From the previous scenario suppose for $X^*$ as $SMK^* = 0$, $PAL^* = 25$ and $SBP^* = 160$ and $X^{**}$ as $SMK^{**} = 1$, $PAL^{**} = 10$, and $SBP^{**} = 120$.

$$\begin{aligned}\hat{OR} = exp[&-\hat{\beta}_1 + 15\hat{\beta}_2 + 40\hat{\beta}_3 - 35\hat{\delta}_{11} + 525\hat{\delta}_{21} + 1400\hat{\delta}_{31} - \\ &\hat{\delta}_{12} + 150\hat{\delta}_{22} + 40\hat{\delta}_{32}]\end{aligned} \tag{3.24}$$

We now state the model in a more general form to suit a generalized Odd Ratio expression:

$$\begin{aligned}Logit \quad \pi(X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_q x_q + \sum^{\pi_1} YV_i + x_1 \sum^{\pi_2} \delta_{1i} W_i \\ + x_2 \sum^{\pi_2} \delta_{2i} W_i + ... + x_q \sum^{\pi_2} \delta_{qi} W_i, \ \ where \ i = 1\end{aligned} \tag{3.25}$$

Our illustration so far highlights on Odd Ratio formula that models many exposure variables and which controls for both confounders and effect modifiers. The discussion on Odd Ratios could have been extended to capture situation for varied modifiers that corresponds to different exposure variables. None the less, this will be restricted to a model which has each exposure variable sharing common modifiers. Consequently we revert to usual notation by identifying the embolden expressions X* and X** as two comparable terms of the exposure and as it were shall form the basis of determining the Odd Ratio of variables for a case for several exposure variables with Confounders and Interaction.

Let set $X^*$ be defined by $X_1^*, X_2^*, ....X_q^*$ and set $X^{**}$ be defined by $X_1^{**}, X_2^{**}, ....X_q^{**}$. We thus present the general Odds Ratio expression for two groups as:

$$\begin{aligned} OR_{x^*x^{**}} = (x_1^* - x_1^{**})\beta_1 + (x_2^* - x_2^{**})\beta_2 + ... + (x_q^* - x_q^{**})\beta_q + (x_1^* - x_1^{**})\sum^{\pi_2}\delta_{1i}W_i \\ + (x_2^* - x_2^{**})\sum^{\pi_2}\delta_{2i}W_i + ... + (x_q^* - x_q^{**})\sum^{\pi_2}\delta_{qi}W_i, \ \ where\ i = 1 \end{aligned} \tag{3.26}$$

## 3.16 Maximum Likelihood Estimation Methods

Parameters in statistical model are usually estimated by researchers by the use of maximum likelihood (ML) estimation. This is one of many alternative procedures that have been formulated by statisticians. Besides, maximum likelihood (ML) estimation is a popular method known as least square (LS) estimation used in linear and multiple regression models. Though ML and LS differ by their approach, they invariably produce same results on condition that the outcome variable is normally distributed. In the estimation of complicated non-linear models MLE is more desirable as far as logistic regression is concerned, most especially when the logistic model is a non-linear model.

In using MLE, two fundamental approaches are considered namely conditional and unconditional methods. The choice of either by the researcher largely depends

on the parameter available in the model in relation to the number of subjects under study. Generally, if the number of parameters in the model is large in relation to the number of subjects conditional MLE is desirable while unconditional MLE is suited for a situation of small parameters in the model relative to the number of subjects.

The criteria for determining small or large parameter s has not been explicitly prescribed by researchers. However, some guiding principles have been made available in the choice of the estimation approach. Once matching is done, the model naturally becomes large as a result of the quantum of dummy variables needed to satisfy the matching stratification. This condition permits the convenient use of conditional MLE.

Unconditional MLE is employed when matching is not been carried out which invariably presupposes that the amount of variables in the model is small with respect to the number of subjects.

From the forgoing conditions and assumptions, it is clear that once the total number of Confounders and Interaction terms in our model are large the number of parameters becomes too large for the adequate use of Unconditional method. Consequently, the Conditional (MLE) is recommended to address the above setbacks. The method has also been proven by researchers to have been producing unbiased and reliable estimates. Again while the Unconditional approach, if not used suitably can produce biased results in the particular instance of Odd Ratio estimation, the Conditional approach has proven otherwise.

## 3.17 The Use of Likelihood Function

Supposing the parameters $X_1, X_2, ..., X_q$ are unknown parameters of a given model. The likelihood function $L$, is a function defined as $L(X)$ where $X$ represents the set of unknown parameters. In an analogous presentation using $X, V, W$ logistic model, we write

$$logit \quad \pi(X) = \beta_0 + \beta X + X\sum^{\pi_1} \gamma_i V_i + X\sum^{\pi_2} \delta_i W_i$$
$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

where

$$X(\beta_0, \beta, \gamma_1, \gamma_2, ..., \delta_1, \delta_2) = \beta_0, \beta_1, \beta_2, ..., \beta_k$$

The likelihood function L(x) essentially combines the inputs or contributions of all the subjects under study. It is therefore the combined likelihood (probability) of observing the data collected.

## 3.18 Application To The Binomial Based Model

Suppose in a research which involves 200 trials of a vaccine, the probability of a successful trial is presented by $\pi$ which happens to be the parameter of interest. Again, of the $n = 200$ trials being observed there are $x = 150$ successes and $n - x(200 - 150) = 50$ failures. We seek to express the joint probability of observing 150 successes out of 200 trials as a binomial based model (a binomial distribution) which is a departure from and quite simpler than the logistic model.

$n = 200$ trials

$\pi =$ probability of success

$x = 150$ successes

$n - x = 50$ failures

The probability of 150 successes, given 200 trials and written as

$$P(X = 150|n = 200) = \binom{200}{150}\alpha^{150}(1-\alpha)^{200-150} = L(\alpha)$$

Generally, we present the likelihood function of this illustration as

$$L(\alpha, X) = \binom{n}{x}\alpha^{x}(1-\alpha)^{n-x}$$

which defines the probability of observing the outcome of the study as a function of the single parameter $\alpha$ and $X$ is the number of successful trials. Now with the likelihood function $L(\alpha)$ determined, we need an estimator $\hat{\alpha}$ which maximizes $L(\alpha)$ and this is chosen by means of maximum likelihood method. The method equates the partial derivative of L, with respect to $\alpha$, and equates it to zero. Thus $L(\alpha)_{max} = \dfrac{\delta L}{\delta \alpha} = 0$. Thus the estimator $\hat{\alpha}$ which is the solution to the above equation maximizes the function $L(\alpha) = (\alpha_1, \alpha_2, ..., \alpha_q)$.

## 3.19 Statistical Inferences For Logistic Regression

Ideally, MLE is used in making statistical inferences once it is obtained. Some of these inferences borders on test of hypothesis and confidence interval estimation, for a given set of parameters in our model.

Interferences are made by the use of MAXIMIZED LIKELIHOOD VALUE $L(\hat{\alpha})$ and ESTIMATED VARIANCE-COVARIANCE MATRIX. The latter contains information used in the required calculations for hypothesis testing and the estimation of confidence interval. This makes Variance-Covariance matrix unique. The two quantities described besides other information are part of the output which standard ML estimation programs provide. The other information that the programs output besides maximized likelihood value and Variance-Covariance matrix, include listing of each variable, its corresponding ML estimate and standard error in that particular order. This is displayed by

Table 3.4: Listing of Variance, ML coefficients and Standard error.

| Variance | $X_1$.....$X_k$ |
|---|---|
| ML coefficient | $\hat{\beta}_0$.....$\hat{\beta}_k$ |
| S.E | $\hat{S}_{\beta_0}$.....$\hat{S}_{\beta_k}$ |

## 3.20 Presenting Models For Inference Making

We discuss the use of ML Techniques in making statistical interferences. To demonstrate this three models are presented using Table 3 in their logit state.

$$\begin{aligned} Logit \quad & \pi_1(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \\ Logit \quad & \pi_2(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \\ Logit \quad & \pi_2(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \end{aligned}$$

Let's denote the maximized likelihood values for 1, 2 and 3 by $L_1, L_2$ and $L_3$ respectively. We further state that a model fits a data better with more parameters.

Then $L_1 \leq L_2 \leq L_3$ and again $\ln L_1 \leq \ln L_2 \leq \ln L_3$

Multiplying through by -2

$-2 \ln L_3 \leq -2 \ln L_2 \leq -2 \ln L_3$

The log likelihood for a given model is defined by $-2 \ln L$. Similarly the individual statistics in the least inequality are the respective log likelihood for models 1-3. Their usefulness cannot be over emphasized because they serve as a means testing hypothesis about parameters in our model using LIKELIHOOD RATIO TEST.

## 3.21 Likelihood Ratio Test

In linear regression, the significance of a regression coefficient is assessed by computing a t-test. In logistic regression there are several different test designed to assess the significance of an individual explanatory variable most notably the likelihood ratio test and the Wald statistic. The likelihood ratio test is the recommended procedure to assess the contribution of individual "predictors" to a given model. In the case of a single predictor model, one simply compares the deviance of the predictor model with that of the null model on a Chi-square distribution

with a single degree of freedom. Using the difference in degrees of freedom of the two models one can conclude that there is a significant association between the explanatory and the outcome given that the predictor model has a significantly smaller deviance. Ideally a likelihood ratio statistic (LR) seeks to find the difference in the likelihood statistic for two given models, one of which has a unique extraction from the other, with an approximate Chi-square distribution in copious samples. The disparity in the number of parameters between the two models reflects the degrees of freedom (df) for this Chi-square test.

Again, we illustrate with our foregoing 3 models:

Model 1: logit $\pi_1(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Model 2: logit $\pi_2(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

Model 3: logit $\pi_3(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_5 + \beta_5 X_5$

Model 1 is a unique extraction from Model 2 and similarly Model 2 a special case of Model 3.

Ideally, in the likelihood ratio test two models need to be identified with a special extraction for the other. The bigger is referred to as full model and the smaller which is obtained by reducing certain parameters of the bigger model to zero (0) is often called the reduced model. The set of parameters put to zero in the full model form the basis of testing the null hypothesis. Again, the number of parameters in the model set to zero in order to obtain the reduced model equal the degree of freedom associated with the likelihood ratio.

Example : We compare Model 2 with Model 3

Model 2: logit $\pi_2(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

Model 3: logit $\pi_3(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_5 + \beta_5 X_5$

Model 3 is larger than Model 2 and therefore are full model is Model 3 and our reduced model shall be Model 2. We realize that two parameters $\beta_4$ and $\beta_5$ from the full model are not contained in the reduced Model. These are coefficient of the variables $X_4$ and $X_5$ respectively and fsorms the basis of the null hypothesis which compares the two models. We emphasize here that:

1. There are no product terms in our full model.

2. For that matter there are no Confounder variables such as $X_1$ and $X_1$ which multiplies $X_4$ and $X_5$ respectively to give product terms.

3. Hence there are no interactive terms such as $X_1X_4$ and $X_2X_5$.

Consequently, we state our

Null hypothesis as: $H_0 : \beta_4 = \beta_5 = 0$

Alternatively hypothesis: $H_1 : \beta_4 \pm \beta_5 \pm 0$.

The likelihood ratio associated with and comparing Model 2 and 3 is

The likelihood ratio associated with and comparing Model 2 and 3 is

$LR = -2 \ln L_2 - (-2 \ln L_3) = -2[\ln L_2 - \ln L_3]$

$LR = -2 \ln \left( \frac{L_2}{L_3} \right)$

This expression is approximately Chi-square with two degrees of freedom.

## 3.22 The Wald Statistic

Alternatively, when assessing the contribution of individual predictor in a given model, one may examine the significance of the Wald Statistic. The Wald Statistic analogous to the t-test in linear regression is used to assess the significant of the coefficients. It is the ratio of the square of regression coefficient to the square of the standard error of the coefficient and is asymptotically distributed as a Chi-square distribution. The Wald Statistics is given by

$$W_i = \frac{\beta_i^2}{SE^2(\beta_i)} \sim \chi^2 \textit{ approximately Chi-square with one degree of freedom}$$

$$Z = \frac{\beta_i}{SE(\beta_i)} \sim N(0,1) \textit{ approximately normally distributed}$$

The wild test is usually applicable when only one parameter is being tested. Typical is the comparison between Model 1 and Model 2 where the former is a special case of or extraction from the latter. Large samples are indifferent to the use

of either Likelihood Ratio ($LR$) or Wald Statistic unlike what obtains in small samples where the two tests differ in results. Thus $LR = W_i$ (for large samples) but $LR \neq W_i$ (for small samples).

EXAMPLE : In this illustration we consider models 1 and 2 already discussed.

Model 1: logit $\pi(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Model 2: logit $\pi(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

$H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

$$W_i = \frac{\beta_3^2}{SE^2(\beta_3)} \tag{3.27}$$

$$Z = \frac{\beta_3}{SE(\beta_3)} \tag{3.28}$$

It is also possible to have a scenario where the Null hypothesis $H_0$ involves more than one parameter. In the particular instance of Model 2 and 3 we realize that we the parameters $\beta_4$ and $\beta_5$ are not part of Model 2 and so from the basis of the Null hypothesis.

## 3.23 Interval Estimation(Single Coefficient)

We now discuss confidence interval estimation. Firstly, a situation where just single coefficient is our interest is being considered. To obtain a large sample confidence interval for a given parameter, we require the following:

1. An estimation of the parameter $\beta_i$
2. A percentage point of the normal distribution defined by $(1-\alpha)100\%$ where $\alpha$ is a critical value determined by a percentage confidence level.
3. The estimated standard error Consequently a $(1-\alpha)100\%$ C.I for $\beta_i$ is

$$C.I = \hat{\beta}_i \pm Z_{(1-\frac{\alpha}{2})} Se(\hat{\beta}_i)$$

For 95% confidence Interval

$$\begin{aligned} C.I &= \hat{\beta}_i \pm Z_{(1-\frac{\alpha}{2})} Se(\hat{\beta}_i) \\ &= \hat{\beta}_i \pm Z_{(1-\frac{0.05}{2})} Se(\hat{\beta}_i) \\ &= \hat{\beta}_i \pm Z_{0.975} Se(\hat{\beta}_i) \\ &= \hat{\beta}_i \pm 1.96 Se(\hat{\beta}_i) \end{aligned}$$

## 3.24 Application to Odd Ratio

It is also possible to estimate confidence intervals for Odd Ratios. For an illustration let us consider an exposure variable $X_2$ which is binary $(0, 1)$ from Model 1. Then

$$logit(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Given $X_2$ is binary (0,1), then $X_1 = X_3 = 0 \Longrightarrow OR(X_2) = e^{-\beta_2}$
Now since $X_2$ represents our binary (0,1) exposure variable of interest, it presumes that $X_1$ and $X_3$ are confounders. Consequently at 95% significant level
$C.I. = e^{C.I \ for \ \beta_2}$

$$\begin{aligned} C.I &= e^{C.I \ for \ \beta_i} \\ &= e^{\hat{\beta}_i \pm 1.96 Se(\hat{\beta}_i)} \end{aligned}$$

## 3.25 Interval Estimation (Interaction Effects)

Once again we assume that our exposure variable of interest is binary (0, 1). Here is an interaction effect as far as our Model is concerned. The expression for Odd Ratio is accordingly adjusted to reflect the effect of our variable of interest controlling for (interacting with) other variables.
Using Model 3 as an illustration
$Logit \quad \pi_3(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \beta_5 X_2 X_3$
$\hat{OR}_{x_3} = e^{\hat{\beta_3} + \hat{\beta_4} X_1 + \hat{\beta_5} X_1}$

putting $\hat{\beta}_3 + \hat{\beta}_4 X_1 + \hat{\beta}_5 X_1 = L$,

a $(1-\alpha)100\%$,C.I for $e^l$ shall be of the form $exp\left(\hat{l} \pm Z_{(1-\frac{\alpha}{2})}\sqrt{\hat{Var}\hat{l}}\right)$

# Chapter 4

# DATA COLLECTION, ORGANIZATION AND ANALYSIS

## 4.1 Introduction

The chapter explains the mode of data collection and its organization for the application of multiple logistic regression model and summary of results.
A review of major statistical tools and procedures employed in this analysis is also discussed.

## 4.2 Data Collection

The data obtained for this study was mainly from a primary source due to the group and scanty nature of existing secondary data at the various health posts and malaria control outfit. The data collection involved a 3 week household cross sectional survey across 22 clusters within The Malaria Control Programme (MCP) Coverage area in the Obuasi Municipality (from 2nd to 23rd April 2014). The target group was children under 5 years and the respondents were mostly mothers aged between 15 and 55 years old. Of the 508 respondents (in this case mothers of the under-five) sampled for the study, 307 had complete information with 201 incomplete cases. These were observations which could satisfy the requirement of research information and without missingness. Thus because we seek to compare two time points to assess risk of reduction in malaria reporting, our complete observation are without;

i Children under 1 year

ii Neutral (zero) outcomes i.e. neither reduction nor increase in malaria reporting.

The instrument used for the data collection (the questionnaire) went through some validation and scrutiny through discussion. It was also pre-tested before the main data collection. The study was meant to capture information on whether children between 1-5 years in the Obuasi Municipality benefited from the two major malaria interventions strategies namely Indoor Residual Spraying (IRS) and Insecticide Treated Nets (ITNs) distribution and to use this data to model the risk of reduction in clinical visits by means of logistic regression. Some important predictors were factored into the outcomes. The predictors were age of child and type of treatment, with four levels namely, IRS only, ITN only, both IRS and ITN and Neither.

## 4.3 Data Analysis

Table 4.1: Paired Samples Statistics

| | Clinical Visits | Mean | N | Std. deviation | Std Error mean |
|---|---|---|---|---|---|
| Pair | Past number of time | 2.3909 | 307 | 1.2355 | 0.0705 |
| | Current number of time | 1.0358 | 307 | 1.0519 | 0.0600 |

Having established that there is a significant difference, we are equally interested in finding which set of scores is higher (time 1 or time 2). To do this the paired sample statistics table is implored. From table 4.1, the mean number of clinical visits to hospitals because of malaria during past year of child (time 1) was about 3 visits and the mean number of visits in the current year (time 2) was about 1 visit. Consequently we observe that there was a significant reduction in malaria reporting for the under 5 given the vis-à-vis the interventions. It is nonetheless crucial to indicate that obtaining a significant difference in the scores for the two time points does not necessarily guarantee that the interventions used in this study namely, IRS and ITN caused the reduction. Other factors such as the use of insecticide spray, mosquito coils, repellents and attitudinal change towards

sanitation, could have contributed to this result.

Table 4.2: Paired Samples Statistics

| | | Paired Differences | | | | | t | df | Sig.(2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| Pair 1 | Past time number of clinical visits-Current time number of clinical visits | 1.35505 | 1.53212 | 0.08744 | 1.18298 | 1.52711 | 15.496 | 306 | 0.000 |

The paired sample t-test is shown in 4.2. Here we ascertain whether there is a significant change in malaria reporting between the first year and last year of the under five following the interventions (IRS and ITN) designed to reduce malaria cases. One set of matched pair of frequency of visits served as input variables for this test. The expected outcome of this test is whether there is statistically significant difference in the mean number of clinical visits for the past 1 year (time 1) and the current year (time 2). The probability value (p-value) of 0.000 in the last column of table 4.2 is less than 0.05 which explains the significant difference between the two sets of score. Since our p-value of 0.000 is substantially smaller than our specified $\alpha$ value of 0.05, we conclude that there is a significant difference in clinical visits to report malaria. The test further gives a t-value of 15.50 with 306 degree of freedom. Again the mean reduction in malaria reporting was 1.36 within 1.53 standard deviation. Our 95% Confidence Interval stretches from a lower bound of 1.18 to an upper bound 1.53.

Table 4.3: Summary Analysis by Cluster

| Cluster | n | Mean(years) | Std Dev(years) | Minimum (yrs) | Maximum (yrs) |
|---|---|---|---|---|---|
| 1 | 13 | 2.762 | 0.982 | 1.580 | 5.000 |
| 2 | 28 | 3.340 | 1.148 | 1.170 | 5.000 |
| 3 | 32 | 2.906 | 1.122 | 1.170 | 5.000 |
| 4 | 22 | 3.076 | 1.124 | 1.250 | 5.000 |
| 5 | 10 | 3.133 | 1.045 | 1.500 | 4.830 |
| 6 | 21 | 2.584 | 1.111 | 1.170 | 4.750 |
| 7 | 17 | 3.108 | 1.217 | 1.080 | 5.000 |
| 8 | 11 | 3.462 | 1.377 | 1.250 | 5.000 |
| 9 | 7 | 3.203 | 1.060 | 2.000 | 5.000 |
| 10 | 11 | 3.182 | 1.036 | 1.500 | 5.000 |
| 11 | 13 | 3.442 | 1.096 | 1.500 | 5.000 |
| 12 | 8 | 4.188 | 1.068 | 2.170 | 5.000 |
| 13 | 3 | 4.307 | 0.636 | 3.750 | 5.000 |
| 14 | 7 | 3.821 | 1.081 | 2.500 | 5.000 |
| 15 | 5 | 3.350 | 1.829 | 1.170 | 5.000 |
| 16 | 12 | 2.945 | 1.204 | 1.170 | 5.000 |
| 17 | 7 | 3.071 | 1.238 | 1.750 | 4.920 |
| 18 | 7 | 3.107 | 1.122 | 1.500 | 4.500 |
| 19 | 10 | 3.250 | 1.267 | 1.500 | 5.420 |
| 20 | 5 | 2.882 | 1.280 | 1.080 | 4.000 |
| 21 | 23 | 3.025 | 1.253 | 1.080 | 5.420 |
| 22 | 35 | 3.324 | 1.260 | 1.250 | 5.000 |

The average age of the 307 children was about 3 years two months $\pm$ 1.18 years standard deviation. The minimum age was about 1 year and a month, and a maximum of about 5years and 5 months. Table 4.3 presents cluster 22 (Kunka junction) as the cluster with the highest respondents and cluster 13 (Wawase) shows the lowest respondents. Cluster 2, 8, 9,10,11,12,13,14,15 and 19 recorded average age exceeding the overall average age.

Table 4.4: Test For Independence (Treatment by response)

| treatment | resp | | Total |
|---|---|---|---|
| | 0 | 1 | |
| Both (IRS and ITN) | 21 | 158 | 179 |
| | 6.84 | 51.47 | 58.31 |
| IRS only | 13 | 64 | 77 |
| | 4.23 | 20.85 | 25.08 |
| ITN only | 2 | 16 | 18 |
| | 0.65 | 5.21 | 5.86 |
| Neither | 11 | 22 | 33 |
| | 3.58 | 7.17 | 10.75 |
| Total | 47 | 260 | 307 |
| | 15.31 | 84.69 | 100.00 |

From table 4.4, 58.31% (179 children) experience both Indoor Residual Spraying (IRS) and Insecticide Treated Nets (ITN) intervention. Of these, we observed that 51.47% (158 children) showed reduction in malaria reporting as with 6.84% (21 children) showed an increase in malaria reporting during their last one(1) year. From the table 4.3, 2.08% (77 children) who experienced IRS intervention only, 28.85% of our sample (64 children) showed a reduction in malaria related clinical attendance with the remaining 13 children (4.23 percent) a drop in clinical attendance for malaria treatment. Again only 5.86 percent (18 children) of the 307 children received only ITN treatment with about 5% (16) showed a reduction in malaria related cases and 13 children (4.23 percent) report an increase in malaria cases within the last 1 year. Finally 11% of our sample (33) received neither IRS nor ITN interventions with approximately 7 percent reporting a drop in malaria related clinical visits in the last one (1) year while the remaining close to 4 percent had their visits to a health facility for malaria treatment increased. The chi-square test for independence (chi-square value 5.68, p-value = 0.0218) showed that treatment is associate with the reduction in malaria reporting.

Table 4.5: Omnibus Tests of Model Coefficients (Chi-square Test)

| | Chi-Square | df | Sig. |
|---|---|---|---|
| Step 1 step | 10.003 | 3 | .019 |
| Block | 10.003 | 3 | .019 |
| Model | 10.003 | 3 | .019 |

The Omnibus Test of model co efficient in table 4.5 gives us an overall indication of how well the model performs. Thus, we seek to conduct a goodness of Fit test for our model, where we want for this results a highly significant value (p-value $< 0.05$). In this particular instance our p-value of 0.019 which is less than 0.05 suggests that each child reported a reduction in clinical visits in his or her last 1 year. The chi-square value of 10.003 with 5 degrees of freedom measures the contribution of individual predictor in the model and the results presupposes that our model which contains all the predictors was statistically significant. Consequently the model was able to distinguish between children who showed or did not show a reduction in malaria cases.

Table 4.6: Analysis of Maximum Likelihood Estimates

| | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | | DF | Estimate | Standard Error | Wald Chi-Square | Pr>ChiSq |
| Intercept | | 1 | 0.0323 | 0.5336 | 0.0037 | 0.9517 |
| age | | 1 | 0.2392 | 0.1411 | 2.8734 | 0.0901 |
| treatment | Both | 1 | 1.2330 | 0.4414 | 7.8016 | 0.0052 |
| treatment | IRS only | 1 | 0.8570 | 0.4821 | 3.1600 | 0.0755 |
| treatment | ITN only | 1 | 1.3281 | 0.8398 | 2.5011 | 0.1138 |

Table 4.6 displays parameter estimates, obtained by the maximum likelihood method, for the model. The parameters being tested for using Wald statistics are The Intercept, age, Both IRS and ITN only with 1 degree of freedom. Here the group of children that benefited from neither IRS nor ITN treatment is being used as a reference point and therefore forms the basis of comparism with other treatments. A cursory look at the respective p-value estimates shows that the parameter estimate of 1.2330 for the use of both IRS and ITN treatments or interventions has the least P-value of 0.0052 (compared to the rest) which is less than a P-value of less than 0.005 and therefore significant, contributing significantly

in estimating the model. Consequently our study is modelled by the equation

$$logit\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 0.0323 + 1.2330(Both) + 0.85790(IRS) + 1.3281(ITN)$$

Table 4.7: Odds Ratio Estimates

| Effect | Point Estimate | 95% Wald Confidence Limits Lower | Upper |
|---|---|---|---|
| age | 1.270 | 0.963 | 1.675 |
| Both vs Neither | 3.432 | 1.445 | 8.152 |
| IRS only vs Neither | 2.356 | 0.916 | 6.062 |
| ITN only vs Neither | 3.774 | 0.728 | 19.570 |

Our results for contrast test shows that the contrast or variations among our explanatory variables controlling for "Neither" is statistically not significant from table 4.7. Individual corresponding P-values are less than 0.05. It means there is not much significant difference among the explanatory variables in contributing to the predictability of the model. Here the group of children that benefited from neither IRS nor ITN treatments is being used as a reference point and therefore forms the basis of comparism with other treatments for the estimation of Odds Ratio. This is the change in the Odds of experiencing either a reduction or increase in malaria reporting with a unit increase in one of the variables (predictors). In this analysis the Odds of a child experiencing a reduction in malaria reporting is 3.432 higher for a child benefiting from both IRS and ITN interventions than for a child benefiting from neither treatment. For all the Odds Ratios shown in table 4.7 there is 95% confidence interval with a range of values for which we can be 95% certain contains the actual value of our Odds Ratio. Since the Confidence Interval for our variable "both vears neither" interventions spans between 1.445 and 8.152, we can as 95 percent confident that the actual value of OR in the population is within that range. Judging from the individual Odd Ratio estimates we conclude that children reporting both IRS and ITN treatment risk reduction in malaria reporting than the other variables with respect to

children who reports neither treatment. Again the possibility of a split between the outcome variables "reduction" or "increase" in malaria reporting is shelved in this analysis since our confidence interval of 1.445 to 8.152 does not contain 1. This would have connoted equal odds.

Table 4.8: Hosmer and Lemeshow Goodness-of-fit

| $\chi^2$ | df | $p - value$ |
|---|---|---|
| 7.9162 | 8 | 0.4417 |

The results shown in this table 4.8 is the Hosmer and Lemeshow Test which supports the model as being adequate. This test apparently one of the most reliable of model fit available is interpreted differently from some other tests being discussed. Given that the threshold measure for the Hosmer- Lemeshow Goodness of Fit test is 0.05, we want to obtain a significant value greater than this. The results of our output in this model show a chi-square value for this test is 7.9162 with a significant level of 0.4417. This value is larger than 0.05, therefore indicating support for the model.

## 4.4 Predictability of Variables

(See appendix table 5.1) The first ten observations from table 5.1 in appendix, being children aged between 1.58 and 5.00 years show high predictability for children who reports both IRS and ITN treatments. Our study models "risk of reduction" and 8 of the 10 observations risk a reduction in reporting for malaria in the last 1 year.

# Chapter 5

# SUMMARY OF RESULTS, CONCLUSIONS AND RECOMMENDATIONS.

## 5.1 Introduction

We present summary of results, conclusions and some recommendations from this study.

## 5.2 Summary Of Results

The study was designed to model the risk of reduction in clinical visits by the under 5, given the types of malaria control interventions using Logistic regression. Primary data from 22 clusters in the Obuasi Municipality was mainly utilized in this study with children 5 years or less as target group.

The instrument (questionnaire) used had undergone some validation before administering on the field. Simple random sampling by quota was used to select the sample from various clusters. Direct logistic regression was performed to assess the impact of a number of factors on the likelihood that respondents would report they had a reduction in clinical visits to hospitals because of malaria. The p-value of 0.000 from the 2-tailed sample t-test (table 4.2) connotes a significant difference between the two sets of data on clinical visits (past 1 year and current year). Also from the paired sample statistics table (4.1) the mean score on clinical visits for the first year of child drops from about 3 to1 for same child in the current year. This attests to the fact that there was a significant reduction in malaria reporting for the under 5 given the interventions.

In analysis the results by cluster the average age of children across the entire 22

clusters was three (3) years with the ages ranging between one (1) and six (6) years. Cluster 22 (Kunka Junction) and cluster 13 (wawase) respectively present the highest and the lowest number of children covered.

The test for independence (Treatment by response table) reveals that nearly 52% of the 307 children used for the sample case analysis showed a reduction in malaria reporting and these are children who experienced both IRS and ITNs interventions, about the highest compared to the other children who experienced other forms of malaria interventions.

The model contained four explanatory variables (Neither ITN nor IRS, ITN only, Both ITN and IRS and only IRS). The full model containing all predictors was statistically significant, $X^2(3, N = 307) = 10.003$, $P = 0.019(< 0.05)$, indicating that the model was able to distinguish between respondents who reported and did not report a reduction in clinical visits in their last one year. As shown in Table 4.6 (Odds ratio estimates) only one explanatory variable made a unique statistically significant contribution to the model (Both ITN and IRS) and invariably happens to be the strongest predictor of reporting risk of reduction recording an Odds Ratio of 3.432. This indicated that children who benefited from both ITN and IRS interventions were about three and a half times more likely to report a reduction in clinical visits within their last one year than those children who benefited from only one Intervention or neither, controlling for all other factors in the model. Table 4.3 also shows that children who benefited from both IRS and ITN interventions risk higher reduction rate (about 52% of the children) than the others. The least number of children who reported a reduction in malaria cases are those who benefited from neither of the interventions (about 7%).

## 5.3 Conclusion

1. There is a significant difference in the frequency of clinical visits in the past and in the present.

2. The reduction in malaria cases does not depend on age of child.

3. There was malaria intervention effect on the reduction of cases.

4. Both malaria intervention strategies (ITN and IRS) showed the significant effect on the reduction of malaria cases.

5. ITNs only, IRS only and neither intervention showed no significant effect on the reduction of malaria cases.

## 5.4 Recommendation

AngloGold Ashanti, Ghana Health Service, The Municipal Assembly, NGO's and other Stakeholders should consider intensifying the use of both interventions (IRS and ITNs) than the use of individual IRS or ITN strategy.

# REFERENCES


Achcar, J. (2011). Use of poisson spatio temporal regression models for the brazilian amazon forest: Malaria count data. *Tropical Medical Journal*, vol. 44(6):pp749–754.

Adekunle, A., Oyindamola, B. Y., Ikeoluwapo, O. A., and Falade, O. C. (2013). Treatment decision making in the home management of malaria among children in south west nigeria. *Discourse Journal of Malaria Research and Reviews.*, Vol1(5):pp 54–62.

Ayi, I., Nonaka, D., Adjovu, J. K., AHanafusa, S., Jimba, M., Bosompem, K. M., Mizoue, T., Takeuchi, T., Boakye, D. A., and Kobayashi, J. (2010). School-based participatory health education for malaria control in ghana: engaging children as health messengers. *Malaria Journal*, vol. 9:98.

Briet, J. T. O., Penny, A. M., Hardy, D., Awolola, S. T., Bortel, V. W., Corbel, V., Dabire, K. R., Etang, ., Koudou, G. B., Tungu, K. P., and Chitnis, N. (2014). Effects of pyrethroid resistance on the cost effectiveness of a mass distribution of long-lasting insecticidal nets: A modeling study. *Malaria journal*, vol. 12:pp 12–89.

Chantal, M. M., Jeremy, A. L., and David, B. E. (2006). Cost effectiveness analysis of strategies to combact malaria in developing countries; achieving the millennium development goals for health. *British Medical Journal*, Vol. 331. (7528).

Chima, R. I., Goodman, C. A., and Mills, A. (2003). Economic impact of malaria in africa a critical review of the evidence. *Health policy*, vol 63:pp 17–36.

Chirombo, J., Lowe, R., and Kazembe, L. (2014). Using structured additive regression models to estimate risk factors of malaria:. *Analysis of 2010 Malawi malaria Indicators Survey Dat.*, Vol.10:pp1–20.

Coleman, P. G., Morel, C., Shillcutt, S., Goodman, C., and Mills, A. J. (2004). A threshold analysis of the cost effectiveness of artemisnim based combination therapies in sub-saharan africa. *Journal of Tropical Medicine Hygiene.*, Vol. 71:pp 196–204 (pub. Med).

De Oliveira, C. E., Dos Santos, S. E., Zeilhofer, P., Souza-Santos, R., and Atanaka-Santos, M. (2013). Geographic information system and logistic regression for high-resolution malaria risk mapping in a rural settlement of the southern brazilian amazon. *Malaria Journal vol.*, vol. 12:pp420.

Eisele, T. P., Larseon, D., and Stekete, R. W. (2010). Protective efficacy of interventions for preventing malaria mortality in children in plasmodium falciparium endemic areas. *International journal of epidemiology*, vol. 39, Issue 1:pp88–101.

Gabriel, P., Robyn, K., and Benedicte, H. A. (2013). Malaria prevalence, indoor residual spraying and insecticide-treated net usage in sub-saharan africa. *PSE working papers*, n 2013-40.:Pp.1–29.

Guyatt, H. L., Corlett, S. K., Robinson, T. P., Ochola, S. A., and Snow, R. W. (2002). Malaria prevention in highland kenya: Indoor residual house spraying vrs insecticide-treated bednets. *Tropical Medicine and International Health*, vol. 7, issue 4:pp298–303.

Haliday, E. K., Okello, G., Turner, L. E., Turner, K., Mcharo, ., Kengo, J., E., A., Dubeck, M. M., Jukes, C. H. M., and Brooker, J. S. a. (2014). Impact of intermittent screening and treatment for malaria among school children in kenya: A cluster randomised trial. *PLOS Medicine*, vol.10.

Hamel, J. M., Otieno, P., Bayoh, N., Kariuki, S., Were, V., Marwanga, D., Laserson, F. K., Williamson, J., Slutsker, L., and Gimnig, J. (2011). The combination of indoor residual spraying and insecticide-treated nets provides added protection against malaria compared with insecticide-treated nets alone.

*The American Journal of Tropical Medicine and Hygiene*, vol. 85(6):pp 1080–1086.

Hardy, D., Briet, J. T. O., Penny, A. M., Awolola, S. T., Bortel, V. W., Corbel, V., Dabire, K. R., Etang, ., Koudou, G. B., Tungu, K. P., and Chitnis, N. (2014). Effects of pyrethroid resistance on the cost effectiveness of a mass distribution of long-lasting insecticidal nets: A modeling study. *Malaria journal*, vol. 12:pp 12–89.

Kinneax, T., Burini, M., and Suoar, R. W. (2002). A comparative analysis of insecticide-treated nets (itns) and indoor residual spraying in highland kenya. *Health policy Plan.*, Vol. 17.:pp 144–53.

Kleinschmidt, I., Schwabe, C., Shiva, M., Segura, J. L., Sima, V., Mabunda, S. J., and Coleman, M. (2009). Combining indor residual spraying and insecticide treated net interventions. *AMI Trop. med. Hyg.*, Vol. 81 (3).:P 519–524.

Korenromp, L. E., Hosseini, M., Newman, D. R., and Cibulskis, E. R. (2013). Progress towards malaria control targets in relation to national malaria programme funding. *Malaria Journal*, vol. 12:pp1–18.

Mideo, N., Day, T., and Read, A. F. (2008). Modeling malaria pathogenesis. *PMC Journal*, vol. 10.:Pp1947–1955.

Moorthy, V. S., Zarifah, R., and Smith, P. G. (2009). Clinical trials to estimate the efficacy of preventive interventions against malaria in paediatric populations: a methodological review. *Malaria journal*, vol 8:pp186–475.

N'Guessan, R., Corbel, V., Akogbeto, M., and Rowland, M. (2007). Reduced efficacy of insecticide-treated nets and indoor residual spraying for malaria control in pyrethoid resistance area, bening. *PMC journal*, vol.13(2):pp199–206.

Onyango, S. A., Kitron, U., Mungai, P., Muchiri, E. M., Kokwaro, E., King, C. H., and Mutuku, F. M. (2013). Monitoring malaria vector control interventions:

Effectiveness of five different adult mosquito sampling methods. *Journal of Medical Entomology*, Vol. 50(5):pp 1140–1151.

Pallant, J. (2007). *Survival manual. A step by step Guide to Data Analysis Using SPSS for windows.* Open University Press, third edition edition.

Philippa, A. W., Protopopoff, N., Rowland, M., Cumming, ., Rand, A., Drakeley, C., Wright, A., Kivaju, Z., Kirby, M., Mosha, F. W., Kisinza, W., and Kleinschmidt, I. (2013). Malaria risk factors in north west tanzania: The effect of spraying, nets and wealth. *PLoS ONE*, Vol. 8:e65787.

Pluess, B., Janser, F., Lenger, C., and Sharp, B. (2010). Indoor residual spraying for preventing malaria (cochrane review). *Cochrane journal*, vol. 4:pp1–48.

Richard-Fabian, S. and Elena, S. (2012). Malaria in children. *Mediterr Journal of Hematology and Infectious Diseases.*, 4(1):e2012073.

Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, Vol. 415,:pp 608–685, pub. Med.

Sarpei-Nunoo, H. (2011). The use of bednets by mothers/careers of children under five years old in the pediatrics out patient department (p.o.p.d) of the 37 military hospital in the prevention of malaria. A thesis submitted to the department of social and clinical pharmacy, Kwame Nkrumah University of Science and Technology (KNUST).:Pp18–42.

Tambo, E., Adedeji, A. A., Huang, F., Chen, J., Zhou, S., and Tang, L. (2012). Scaling up impact of malaria control programme. a table of events in subsaharan africa and people republic of china. *PMC Journal*, vol.1:pp1–7.

Tatem, A. (2014). Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria Journal*, vol. 13:pp1–2.

Thwing, J., Elsie, T. P., and Stekete, R. W. (2011). Protective efficacy of malaria case management for preventing malaria mortality in children: a systematic

review for the lives saved tool. *BMC Public Health journal*, vol. 11, issue 3:pp 186–471.

Unicef (2007). Malaria and children. progress in intervention coverage. *American Journal of Epidemiology*, pages Pp.5–32.

West, P. A., Protopopoff, N., Wright, A., Kivaju, Z., Tigererwa, R., Mosha, F. W., Kisinza, W., Rowland, M., and Kleinschmidt, I. (2014). Indoor residual spraying (irs) in combination with insecticide treated nets (itns) compared to itns alone for protection against malaria. a cluster randomised trial in tanzania. *Pmed journal*, vol 10.

White, N. J. (2002). The assessment of antimalaria drug efficacy. trends parasitol. Vol.18(10):pp458–64.

White, T. M., Conteh, L., Cibulskis, R., and Ghani, C. A. (2011). Cost and cost effectiveness of malaria control interventions-a systemic review. *Malaria Journal*, vol. 10:337:pp1–14.

WHO (2013). Indoor residual spraying. an operational manual for indoor residual spraying for malaria transmission control and elimination. pages P3–13.

# Appendix

Table 5.1: Predictability of variables

| Obs | age | resp | treatment | Predicted | residuals |
|---|---|---|---|---|---|
| 1 | 2.33 | 1 | Both | 0.86087 | 0.54737 |
| 2 | 3.58 | 1 | Both | 0.89298 | 0.47580 |
| 3 | 2.25 | 1 | Both | 0.85857 | 0.55225 |
| 4 | 2.92 | 1 | Both | 0.87693 | 0.51250 |
| 5 | 5.00 | 1 | Both | 0.92137 | 0.40470 |
| 6 | 2.58 | 1 | Both | 0.86788 | 0.53235 |
| 7 | 3.92 | 1 | Both | 0.90050 | 0.45782 |
| 8 | 1.75 | 0 | Both | 0.84341 | -1.92569 |
| 9 | 1.58 | 1 | Both | 0.83797 | 0.59460 |
| 10 | 2.25 | 0 | Both | 0.85857 | -1.97784 |